Towards a Conceptual Framework for Exploring and Modelling Expressive Musical Gestures

Nicolas Rasamimanana

IRCAM, CNRS - UMR STMS,

1 Place Igor Stravinsky,

75004 PARIS, France

{Nicolas.Rasamimanana}@ircam.fr

October 21, 2011

Abstract

This paper reports on a conceptual framework for the analysis and modelling of expressive musical gestures. This framework, called space of possibilities (SoP), is based on crossing different constraints a player must face when playing his instrument. These constraints can be acoustical, biomechanical, and are considered along with musicians' acquired skills. Such approach can provide a unified ground for modelling performer - instrument relationships by encompassing different perspectives within a single view. We propose to define this SoP through the review of various computational techniques for expressive gestures: feature-based approaches, dynamic modelling, and segmental modelling. More precisely, we see how the notions of gesture invariance and variability, gesture qualities, and gesture primitives can help structure SoP.

1 Introduction

Expressiveness with digital sounds and audio processes has been the focus of many researches in computer music over the last decade. Musicians and researchers have indeed been particularly interested in designing real time interaction with synthetic sounds, trying to play and create music with a computer as musicians would do with their instruments. In this trend of developing novel ways to perform music, a keen interest was shown for gesture controlled systems and tangible interfaces. Such technologies indeed afforded the use of physical input over virtual disembodied sounds. Gesture and gesture sensing technologies were hence found promising to engage a performer into a corporeal experience with a computer (Miranda and Wanderley, 2006). In addition, these technologies also provided for novel approaches to investigate the relationships between musicians and their instruments. This second point was also very important because the study and modelling of musicians' gestures in live situations could reveal basic components of an embodied interaction (Leman, 2007).

This paper deals with the latter aspect. We indeed believe that modelling gestures can help identifying principles that are promising for the design of expressive musical interaction. Music gestures however encompass various components. Gestures can be related to sound production, music entrainment, communication between performers, and musicians' mental representations (Cadoz, 1988; Wanderley et al., 2005; Dahl and Friberg, 2004; Jensenius, 2007). We here propose to focus on those directly related to sound production because these generally are in first line when defining gesture interactions with electronic sounds. These gestures were found to embrace several expressive cues. For instance, expressivity can lie in both discrete and continuous aspects (Ortmann, 1929; Dahl et al., 2009; Dahl, 2000). Expressivity can also be the result of gestures extrema, gestures variances and invariances, as well as gesture qualities, i.e. the ways gestures are performed (Fdili Alaoui et al., 2011; Bouënard et al., 2010; Rasamimanana, 2008). The question hence arises to how to capture and use these aspects to create richer interactions with digital media.

Computational methods actually represent a promising way to deal with the manyfold aspects of expressive gestures. They indeed enable to automatically or semi-automatically extract information about a performance and are useful to analyse and model gestures. However, there can be several methods depending on the characteristics one intend to grasp from a gesture. This can eventually lead to numerous independent descriptions. In this paper, we aim to define a conceptual framework that could bring together different modelling of expressive gestures. First, we propose to circumscribe performers' expressive gestures to more specific domains of definition considering constraints on the instrument acoustics and the performer's physiological limits. We call these domains *spaces of possibilities* (SoP) (section 3). Second, we propose to review various types of computational methods used in previous works, and link them to the exploration of these spaces of possibilities (section 4). In the next section, we first investigate on pioneer concepts for modelling expressive gestures.

2 Expressive Gestures: the MIDI Standard and Beyond

One of the most emblematic endeavour in making use of performers' gestures to control synthetic sounds may well be the 80s synthesizers with the MIDI standard. This system indeed enabled musicians to use a keyboard to control a wide range of virtual sounds. Later, the advent of gesture sensing technologies, and notably affordable inertial measurement units (IMU), favored the expansion of new musical interfaces (Miranda and Wanderley, 2006), as can be witnessed for example in the NIME international conference. However, few if none of these latest interfaces knew a success comparable to MIDI and the synthesizers. We propose to investigate about possible reasons for such difference.

2.1 The MIDI standard: a Historical Separation between Gesture and Sound?

The MIDI standard is acknowledged in the computer music community for its versatility and is widely used. Two aspects responsible for that popularity can notably be brought forward. First, the specification of MIDI standard allows to separate a musical instrument into a control unit and a sound unit. Synthesizers, and thereafter MIDI controllers, have benefited from this flexibility that permitted to control sounds of different types through a single interface, namely a keyboard: synthetic sounds, sampled sounds from other keyboard instruments (e.g. Steinway, Bösendorfer) or from other instruments (e.g. guitar, glockenspiel, trumpet, violin). This separation actually made the MIDI standard one of the first solutions enabling the creation of *Digital Music Instruments* (or *DMI*) (Wanderley and Depalle, 2004).

The second aspect lies within the definition of MIDI controls. The MIDI standard is indeed based on a description of the actions related to a keyboard. The event NOTE-ON, resp. NOTE-OFF, actually describes the instant a key is pressed, resp. released. The VELOCITY parameter reflects the impulse given by the pianist on the key. This basic description eventually appears to be one of the first models of a musician's gestures, in this case those of a keyboard player (Letz, 1998). Thanks to this modelling, the MIDI standard actually sets constraints on possible mappings between gestures and sounds that allow for (1) flexible associations and re-associations between controls and sounds and (2) compatibility and relevance to keyboard playing. This aspect, although less frequently noted than the first, probably equally contributed to the success of the MIDI standard. It indeed allowed the creation of many novel musical interactions based on the keyboard while keeping the system consistent for the player as well as for the audience.

2.2 Towards Modelling Expressive Continuous Gestures

However, this standard inherited from the 80s and from the incredible development of synthesizers at that time suffered much criticism (Moore, 1988; McMillen, 1994; Wessel and Wright, 2002). In particular, if the control is persuasive for percussive sounds like a piano or a glockenspiel, it is much less convincing for sustained, continuous sounds like bowed strings. Although widely used in current synthesizers, the association between a keyboard control and a sustained sound, i.e. between discrete MIDI events and continuous control gestures, was found particularly inadequate (McMillen, 1994; Wessel and Wright, 2002; Rank, 1999).

Several alternatives were proposed (McMillen, 1994; Wessel and Wright, 2002) among which the OSC protocol. OSC notably allowed the use of continuous gesture input, which lead to the creation of numerous DMI and mapping tools (Sinclair et al., 2010). However, these solutions all targeted the same objectives: increasing bandwidth and improving time resolution, which were judged too limited in MIDI. Although they brought real improvements, in particular continuous gesture input, these solutions did not directly address another important feature: modelling instrumental gestures. This had immediate consequences. It allowed much freedom to the performer who had infinite possibilities to map gestures and sounds. However, this total freedom simultaneously contributed to the loss of mediation between the performer, the instrument and the audience (Leman, 2007).

Works by Chafe (1989) and Rank (1999) were among the first trials to define and model continuous instrumental gestures. Based on bowed string case studies, these works attempted to model players' control parameters. Nevertheless, they both pointed the difficulty to define relevant temporal evolutions. In this paper, we carry on such endeavour. We believe this kind of models are key to the design of interactive musical systems that keep a sense of mediation for performers and for the audience (Leman, 2007). In this perspective, the analysis of musicians' gestures in live situation is central (Maestre, 2009; Bouënard, 2009; Demoucron, 2008; Schoonderwaldt, 2008; Rasamimanana, 2008).

3 Crossing Constraints: Spaces of Possibilities

We propose to define a conceptual framework considering performer - instrument relationships that can provide ground to model expressive gestures. This framework is combining acoustical and physiological constraints along with musicians' skills.

We illustrate this framework with violin gestures, which offer a privileged ground to study continuous expressive gestures. Several works indeed report methods to capture violinists' gestures while minimally disturbing playing (Schoonderwaldt and Demoucron, 2009; Maestre, 2009; Young, 2007; Grosshauser, 2008).

3.1 Definition of a Space of Possibilities

The notion of space of possibilities arises from the consideration of the various constraints, whether acoustical or biomechanical, that a performer must manage to play his instrument. On the one hand, an acoustic instrument emits a sound under certain specific conditions. To play a stringed instrument for instance, a musician must perform the necessary movements to initiate and maintain the vibration of the strings of the instrument. He needs for example to adjust various parameters such as bow force, bow velocity and bow contact point on the strings (Askenfelt, 1989).

On the other hand, a musician is also submitted to physiological possibilities, meaning that some movements may not be humanly possible. In particular, players actually have *expert movements*, developed through training and regular practice to master their instruments. Such movements imply a notion of efficiency that can be seen as biomechanical constraints (Nelson, 1983). Many models were designed to account for these constraints, mostly through optimizing various effort criteria, for instance minimum jerk (Hogan, 1984), the equilibrium point control (Feldman, 1986), minimum torque change (Uno et al., 1989) or minimum muscle-tension change (Uno et al., 1989b).

We further add considerations on the musician's acquired skills. These encompass notions related to the player's experience and learning. These constraints hence permit to consider possibilities at a player's scale, and as such to explicitly deal with idiosyncrasies. They allow to account for different levels of mastery, gesture control, and eventually expressivity. Crossing these different constraints, related to instrument sound production, musician's biomechanics, and musician's skills defines what we call a *space of possibilities*. Figure 1 illustrates such a conceptual space, where we represented acoustic and biomechanical constraints along with musician's acquired skills, each symbolized with a circle.

=== FIG 1 ===

Let us consider the crossing of all constraints (zone labeled 4). This is conceptually the space that musicians exploit when playing their instruments. Developing skills and using this space to achieve different expressive gestures is actually considered part of a musician's training (Bril, 2002; Goasdoué, 2002). This space of possibilities therefore allows for the existence of different strategies of actions that we usually explore while studying performance gestures. Interestingly, the crossing of these circles also defines alternative zones that have musical interpretations. For example, Zone 1 corresponds to theoretically possible but unexploited combinations. This can typically be found for unmastered instrument possibilities, for instance with less experienced musicians. One may assume that learning students aim to reduce this area while it is minimum-sized for experienced instrumentalists. Zone 2 corresponds to combinations that would be instrumentally possible but humanly impossible. That would correspond to an instrument that would be unplayable. In traditional acoustic instruments, this area may be very small if existent at all: through their crafting history, instruments were indeed built based on both sound possibilities and playability. However, this Zone 2 can be important to consider in the design of novel instruments, as it is electronically possible to create sounds that do not match biomechanical constraints (e.g. a very long sustained sound) therefore possibly disturbing the forementioned sense of mediation (Leman, 2007). Zone 3 finally lies beyond the instrument acoustic possibilities but within the musician's skills. This corresponds to gestures not related to instrument playing. Recently, musicians were interested in using these kind of gestures for their instrument, which they eventually did thanks to electronic means therefore leading to the creation of augmented instruments (Lähdeoja, 2008; Machover and Chung, 1989; Burtner, 2002).

This conceptual framework permits to account for different constraints that a musician must face when performing. Although it may not be an exhaustive model, we believe it can already provide ground for the study of expressive gestures as it encompasses different perspectives within a single view.

3.2 SoP Exploration Through Playing Techniques

Within their SoP (Zone 4), skilled musicians can develop different strategies of actions (Bril, 2002; Goasdoué, 2002). These can be related to expressive gestures and especially to *playing techniques*. Playing techniques were developed by several generations of instrumentalists and indeed reflect the expressive capabilities of the instrument. Through specific actions musicians are actually able to produce numerous techniques that change timbre and have specific dynamic evolutions. In the case of bowed strings for example, performers can play on the fingerboard to produce a fluted sound (*flautendo*) or play with the bow stick to make the sound noisy (*con legno tratto*). By controlling the dynamics of the bow, they can produce different bowing techniques such as *Détaché*, *Martelé* and *Spiccato*. In *Martelé*, the bow has to move fast and must be stopped abruptly, while in *Détaché* the bow is driven continuously "like moving a weight up and down a hill" (Menuhin, 1973). In *Spiccato*, the bow bounces on the string with short strokes. These techniques have distinct sound results that violinists use to build their interpretations.

Playing techniques hence offer a privileged access to SoP. As a matter of fact, we here propose to focus on their study. However the idea is not to find a model for every playing techniques, which saves us from inevitable cataloging, but rather to identify underlying principles that could be the basis for structuring SoP. The works reviewed in this paper deal with a set of three violin playing techniques, i.e. *Détaché*, *Martelé* and *Spiccato*. These were chosen for their differences and complementarities: they are indeed acknowledged as three main poles in violin playing from which musicians can further build skills for their bowing technique (Galamian, 1999). In the following sections, we review studies on these violin playing techniques and consider them under a SoP framework.

4 Exploring and Modelling: Case Study with Violin Expressive Gestures

We review three different approaches to modelling playing technique gestures. The first approach consists in extracting features to describe and analyse gestures. The two other approaches deal with gestures as a continuous phenomenon, although the last one adds a higher-level structuration of gestures. These approaches offer three different angles in modelling and hence complementary viewpoints on SoP.

4.1 Feature-Based Approach

Feature-based approach consists in extracting one measure out of a set of data, which can be discrete values or continuous signals. For instance, one can compute the mean of a signal or its standard deviation over different periods of time, which permits to gather information on the data trend. Such approach is very popular in music information retrieval for audio signal description and several audio features can be found (Peeters, 2004). However, contrary to this field, there is still no consensus for a set of features to describe and analyse gesture signals as there can be for audio. Ad hoc informative features are most often computed for each study case. Two extreme positions can be envisaged. On the one hand, we have some a priori knowledge on the data and we are able to design specific features that can account for this knowledge. This is rarely the case as we most often aim to gather knowledge. On the other hand, no a priori is taken on the data, and we are to perform a "blind" feature selection using techniques of dimension reduction (Young, 2008). Such approaches are more popular because they permit to highlight some structural aspects of data sets, e.g. variance with principal component analysis (PCA) or discrimination power with linear discriminant analysis (LDA) without prior knowledge. However, contrary to the first option, they also usually produce features with no clear understandable interpretation which makes them hard to reuse for other analyses.

Between these two extremes, mixed approaches that combine feature selection and interpretability of the data can be found (Bouënard et al., 2010; Rasamimanana et al., 2006). As an illustration, we here performed analyses similar to (Rasamimanana et al., 2006). While this study focused on two players, we extended it to 12 violinists with the aim to compare performances. Figure 2 plots the results of this feature description for the violin techniques *Détaché*, *Martelé* and *Spiccato*. The twelve violinists played scales in the three bowing techniques, and we used a description based on maximum and minimum acceleration within a stroke (see Rasamimanana et al. (2006) for more details on feature computation and selection). All violinists were students with at least ten years of training in a music conservatory.

=== FIG 2 ===

As in (Rasamimanana et al., 2006), the three playing techniques appear in clusters for all players. We can notice that some players show better defined clusters (3, 4 and 7) while others have more mixed ones (8, 9, 11). Besides, idiosyncrasies can be observed: *Martelé* from player 1 appears in a different region than *Martelé* from player 2 for instance, therefore indicating that the two players performed this bowing style in a different manner. These results can be directly related to SoP as they actually tell us how the violinists used SoP. Interestingly, some performances occupy smaller areas (e.g. 3, 9) while others have bigger (e.g. 8 or 12). This could be related to the players' skills, bigger areas corresponding to more developed SoP.

More generally, the results bring forward the notions of variance and invariance of the playing techniques. For each bowing style, bow strokes are consistent together (intra class invariance), they are not exactly identical (intra class variance), and they are different from bow strokes of other bowing styles (inter class variance). Parameter variances actually delimitate playing techniques, allowing potential variations around one pattern, and directly account for the playing techniques expressive power. These concepts are central as they structure SoP: it can be seen as an ensemble of distinct gesture classes, with variations that actually account for gesture expressivity. This can be directly used for the design of an electronic music system where each playing technique can be recognized and can control an electronic effect: bow strokes variations can then be used to vary each effect (Bevilacqua et al., 2006).

Moreover, when considering all the violinists' performances together, one can see that some regions in the graphs are filled with points, while others remain empty. This puts forward an important aspect. As shown in (Rasamimanana et al., 2006), there can be continuity between playing techniques, i.e. one can go continuously from one bowing style to another. This indicates that some intermediate points may correspond to existing bow strokes and as a matter of fact that feature-based approaches can account for such property. However, there may also be discontinuities as shown in (Rasamimanana et al., 2008): gesture discontinuity can typically be found between slow and fast gestures. These results are also important for the structure of SoP.

They reveal that SoP may not always be a smooth continuous space but can be more complex and changing. This last point can be challenging for feature based approaches to grasp.

Feature-based modelling can give synthetic information on an ongoing gesture. However, such modelling is also known for mixing down data temporal sequencing. For example, two curves may have the same mean value, while having completely different shapes. This weakness, known as the "bag-of-frame" effect in audio description (Casey and Slaney, 2006), leads to examine approaches that can better encompass temporal sequencing of gesture signals.

4.2 Dynamic Modelling: a Continuous Approach

Dynamic modelling permits to address gesture as a process over time, generally based on a parametric modelling. This approach can be remarkably useful to deal with temporal shapes: (1) it can be used to describe gestures as continuous functions and (2) based on this description it can be used to re-synthesize data. This last point is particularly promising for the design of electronic music systems. For example, synthesized gestures could be used as a rich control over sound synthesis.

Several works can be reported on modelling violin gestures with similar approaches. Works by Chafe (1989) and Rank (1999) used ad hoc functions but pointed the difficulty to define their temporal evolutions. Demoucron (2008) used captured data to abstract gesture functions. Based on this parameterization, he could control the shapes of several violin parameters (e.g. bow velocity, bow force on string), which he used as input for physical modelling sound synthesis. Using polynomials and spline functions as in (Maestre et al., 2010) is a common way of dealing with function definitions. Nevertheless, controlling these models can be difficult especially because users have to a priori choose a polynomial order. If not adapted, the model may not fit data: some details can be skipped (underfitting) or the fitting can show additional peaks (overshooting). An alternate way of defining gesture functions is to use a probabilistic framework. Captured gestures can be given as templates of a generative model (Bevilacqua et al., 2010). The model variances can be set through statistical learning such as described in (Rabiner, 1990), or can be explicitly controlled to account for specific gesture variations. In these presented works the question of parameter control is essential: changing parameters can actually lead to synthesize humanly impossible gestures. We therefore fall within Zone 2 in Figure 1, where designed electronic music system may eventually become hard to understand or even play. In these presented works, it can be hard to know when one actually falls in such cases. Other approaches propose to explicitly deal with this aspect.

The biomechanical level of SoP can provide ground for such modelling by adding constraints to possible models. Gesture can for instance be modeled as a dynamic biomechanical system (Nelson, 1983). For example, violin expressive gestures can be related to the minimization of effort costs as in (Rasamimanana and Bevilacqua, 2009). In this case, a single model is used for all bowings: a mass M moving a distance D in time T. However, depending on the choice of a cost minimization, e.g. jerk or impulse, this model can synthesize various gestures with specific temporal shapes (see Figure 3): *Détaché* can be modeled with a trapezoidal model (minimum impulse) and *Martelé* can be modeled with minimum jerk. Such modelling is actually promising for expressive gestures because not only does it focus on a physical action, i.e. moving a certain distance in a certain amount of time, but it also, and maybe most importantly, focuses on the way gesture is performed. This biomechanical level hence permits to address the notion of *gesture quality* (Fdili Alaoui et al., 2011). Similarly, other works based on dynamic formulations relied directly on the simulation of motion equations, while proposing high-level control policies coming from feature-based analysis methods. This has been proved efficient for building an electronic music system dedicated to the synthesis of percussion performances (Bouënard et al., 2011a,b).

=== FIG 3 ===

Dynamic modelling with biomechanical considerations permits to describe expressive gestures with their intrinsic temporal natures, shapes and durations, while taking into account motion constraints. However, it may be necessary to use several action strategies along time to better describe a player's performance: for example on Figure 3, using only one model for all four strokes instead of four models as done in the study would lead to erroneous analysis and hence synthesis. To solve this problem, we have to consider smaller gesture segments. In the following section, we propose a framework that can support gesture units in SoP.

4.3 Segmental Approach and Gesture Primitives

Segmental approach consists in modelling complex gesture performances based on smaller gesture segments. These segments, also called *gesture primitives*, represent gesture units with specific meaning. More complex gestures are hence seen as a concatenation of these primitives. This formalism actually resonates with previous works on chunking and coarticulating gestures in musicology (Godøy et al., 2010), psychology (Palmer, 1997), and computer music (Bianco et al., 2010; Rasamimanana and Bevilacqua, 2009; Rasamimanana et al., 2009). With this approach, SoP acquires a hierarchical structure with a first level addressing gesture primitives and a second dealing with more complex concatenated gestures.

This formalism can be modeled using probabilistic frameworks. For example, segmental models can account for this two level hierarchy. They can be viewed as simplified Hierarchical Hidden Markov Model (HHMM) (Ostendorf et al., 1996) where the first layer represent the segments. A higher layer, also called model topology, then defines structures between these segments by specifying privileged transitions from one segment to another. Such models proved to be efficient to model writing gestures from a graphic tablet (Artières et al., 2007), or temporal patterns of audio features (Bloit et al., 2010).

We present here an illustration of such modelling carried out on performance gestures (see Figure 4). We defined a set of four primitives $S1, S2$, $S3$ and $S4$. The three first primitives $S1, S2$, $S3$ were sampled from a recorded gesture, while $S4$ was defined from scratch as a flat horizontal primitive. The primitives were additionally parameterized so as to allow for minor stretch and shrink while keeping their overall shapes. The model topology describes the possible primitive sequences. There can be two options: decoding the model either gives $S1$ followed by $S2$ followed by $S3$, or alternatively $S4$ alone. Figure 4 (bottom) shows the decoding results for this succession of repeated gestures.

=== FIG 4 ===

In this example, we chose sampled primitives. However, as previously discussed, it could be interesting in the future to consider primitives defined from motion equations. This could have two consequences. First, it would lead to a more powerful model where primitives can be defined a priori and independently from the data. Second, it could contribute to the definition of a set of standard primitives from which to model more complex gestures. Besides, being based on a generative model, this formalization of SoP could also be used to synthesize new gestures according to the model topology. Such approach would guarantee high level consistency such as required by biomechanical or even musical constraints, while having enough flexibility to generate expressive gestures.

5 Conclusion and Perspectives

In this paper, we presented the conceptual framework of space of possibilities (SoP) for expressive gestures. It is based on the acoustical, physiological and skill constraints that a musician has to manage when playing his instrument. We investigated how this framework can provide ground for modelling gestures, whether for performance analysis or for gesture synthesis.

We identified some underlying principles of SoP through the review of three different and complementary approaches to model expressive gestures. SoP can account for gesture invariance and variances, which are essential for gesture expressivity. SoP explicitly deals with humanly possible gestures, while also allowing for more artificial, transformed gestures. SoP can be structured with two hierarchical levels to account for more complex gestures: first level is composed of gesture primitives, while second is made from concatenating these elementary gestures.

We feel SoP can be used for the design of electronic music systems that keep a sense of mediation with the audience and with performers. For instance, systems presented in (Schnell et al., 2011) and (Rasamimanana et al., 2011) were partly designed based on these concepts. These systems were relatively intuitive to design and feedback from expert users as well as from novices was quite encouraging. These systems were based on probabilistic approaches such as described in (Bevilacqua et al., 2011). We aim to draw more on motion equations and dynamic models in future implementations. This could encompass promising notions of gesture qualities.

In this paper, SoP is actually defined considering relatively low-level constraints between the performer and his instrument. Considering additional higher-level constraints such as aesthetics or cognition could bring further fruitful insights. Finally, we believe approaches based on gesture primitives are especially promising for expressive gesture modelling. It can be a step towards defining a set of meaningful data-independent analysis tools, which is still missing so far. We believe investigating SoP will help defining such a set.

6 Acknowledgments

We acknowledge support of ANR and CapDigital (Interlude project ANR-08-CORD-010), and la Région Ile de France. We also thank the whole Interlude consortium.

References

Artières, T., Marukatat, S., and Gallinari, P. (2007). Online Handwritten Shape Recognition Using Segmental Hidden Markov Models. *IEEE transactions on pattern analysis and machine intelligence*, 29:205–217.

Askenfelt, A. (1989). Measurement of the bowing parameters in violin playing. II: bow-bridge distance, dynamic range, and limits of bow force. *The Journal of the Acoustical Society of America*, 86(2):503–516.

Bevilacqua, F., Rasamimanana, N. H., Fléty, E., Lemouton, S., and Baschet, F. (2006). The augmented violin project: research, composition and performance report. In *NIME*, pages 402 – 406.

Bevilacqua, F., Schnell, N., Rasamimanana, N., Zamborlin, B., and Guédy, F. (2011). *Online Gesture Analysis and Control of Audio Processing*, volume 74 of *Springer Tracts in Advanced Robotics*, pages 127–142. Springer Berlin Heidelberg, Berlin, Heidelberg.

Bevilacqua, F., Zamborlin, B., Sypniewski, A., Schnell, N., Guédy, F., and Rasamimanana, N. (2010). Continuous realtime gesture following and recognition. *Gesture in Embodied Communication and Human-Computer Interaction (LNCS)*, 5934:73–84.

Bianco, T., Freour, V., Rasamimanana, N., Bevilaqua, F., and Caussé, R. (2010). On gestural variation and coarticulation effects in sound control. *Gesture in Embodied Communication and Human-Computer Interaction (LNCS)*, 5934:134–145.

Bloit, J., Rasamimanana, N., and Bevilacqua, F. (2010). Modelling and segmentation of audio descriptor profiles with Segmental Models. *Pattern Recognition Letters, Special Issue on Future Trends in Pattern Recognition of Non-Speech Audio Signal*, 31(12):1507–1513.

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

Bouënard, A., Gibet, S., and Wanderley, M. M. (2011a). Hybrid inverse motion control for virtual characters interacting with sound synthesis. *The Visual Computer*, pages 1–14.

Bouënard, A., Wanderley, M. M., and Gibet, S. (2010). Gestural control of sound synthesis: Analysis and Classification of Percussion Gestures. *Acta Acustica united with Acustica*, 96(4):668–677.

Bouënard, A., Wanderley, M. M., Gibet, S., and Marandola, F. (2011b). Virtual Gesture Control and Synthesis of Music Performances: Qualitative Evaluation of Synthesized Timpani Exercises. *Computer Music Journal*, 35(3):57–72.

Bril, B. (2002). Apprentissage et contexte. *Intellectica*, 2(35):251–268.

Burtner, M. (2002). The Metasaxophone: concept, implementation, and mapping strategies for a new computer music instrument. *Organised Sound*, 7(2):201–213.

Cadoz, C. (1988). Instrumental gesture and musical composition. In *Proceedings of the International Computer Music Conference*, pages 1 – 12.

Casey, M. and Slaney, M. (2006). The importance of sequences in musical similarity. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, ICASSP*, pages V–5–8.

Chafe, C. (1989). Simulating performance on a bowed instrument. In *Current directions in computer music research*, pages 185–198. MIT Press.

Dahl, S. (2000). The playing of an accent - Preliminary observations from temporal and kinematic analysis of percussionists. *Journal of New Music Research*, 29(3):225–233.

Dahl, S., Bevilacqua, F., Bresin, R., Clayton, M., Leante, L., Poggi, I., and Rasamimanana, N. (2009). Gesture in performance. In Godø y, R. I. and Leman, M., editors, *Musical Gestures: Sound, Movement, and Meaning*.

Dahl, S. and Friberg, A. (2004). *Expressiveness of Musician's Body Movements in Performances on Marimba*, volume 2915, pages 361–362. Springer Verlag.

Demoucron, M. (2008). *On the control of virtual violins: physical modelling and control of bowed string instruments*. PhD thesis, Université Paris 6 - IRCAM UMR STMS.

Fdili Alaoui, S., Caramiaux, B., and Serrano, M. (2011). From dance to touch: movement qualities for interaction design. In *CHI*, pages 1465–1470. ACM.

Feldman, A. (1986). Once more on the equilibrium-point hypothesis (lambda model) for motor control. *Journal of motor behavior*, 18(1):17–54.

Galamian, I. (1999). *Principles of Violin Playing and Teaching*. Shar Products Co, 3rd edition.

Goasdoué, R. (2002). Recherches sur l'apprentissage musical instrumental. *Journal de Recherche en Education Musicale*, 1(1):34–54.

Godøy, R. I., Jensenius, A. R., and Nymoen, K. (2010). Chunking in Music by Coarticulation. *Acta Acustica united with Acustica*, 96(4):690–700.

Grosshauser, T. (2008). Low Force Pressure Measurement: Pressure Sensor Matrices for Gesture Analysis, Stiffness Recognition and Augmented Instruments. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, pages 97–102.

Hogan, N. (1984). An organizing principle for a class of voluntary movements. *The Journal of Neuroscience*, 4(11):2745–2754.

Jensenius, A. R. (2007). *Action – Sound: Developing Methods and Tools to Study Music-related Body Movement*. PhD thesis, Department of Musicology, University of Oslo.

Lähdeoja, O. (2008). An approach to instrument augmentation: the electric guitar. In *Proceedings of the 2008 Conference on New Interfaces for Musical Expression (NIME08)*.

Leman, M. (2007). *Embodied music cognition and mediation technology*. The MIT Press, Cambridge, Massassuchetts.

Letz, S. (1998). *Recherches et applications en informatiques musicales: Les normes MIDI et MIDIFiles*, chapter 13. Hermès.

Machover, T. and Chung, J. (1989). Hyperinstruments: Musically intelligent and interactive performance and creativity systems. In *ICMC*, pages 186 – 190.

Maestre, E. (2009). *Modelling Instrumental Gestures: An Analysis/Synthesis Framework for Violin Bowing*. PhD thesis, Universitat Pompeu Fabra.

Maestre, E., Blaauw, M., Bonada, J., Guaus, E., and Perez, A. (2010). Statistical Modelling of Bowing Control Applied to Violin Sound Synthesis. *IEEE Transactions on Audio, Speech and Language Processing*, 18(4):855–871.

McMillen, K. (1994). ZIPI - origins and motivations. *The Computer Music Journal*, 18(4):47–51.

Menuhin, Y. (1973). *L'Art de jouer du violon*. Buchet / Chastel.

Miranda, E. and Wanderley, M. (2006). *New Digital Musical Instruments: Control and Interaction beyond the Keyboard*. A-R.

Moore, R. (1988). The dysfunctions of MIDI. *The Computer Music Journal*, 12(1):19–28.

Nelson, W. L. (1983). Physical principles for economies of skilled movements. *Journal Biological Cybernetics*, 46(2):135–147.

Ortmann, O. (1929). *The physiological mechanics of piano technique.* Dutton, New York.

Ostendorf, M., Digalakis, V., and Kimball, O. (1996). From HMM's to segment models: a unified view of stochastic modelling for speech recognition. *IEEE Transactions on Speech and Audio Processing*, 4(5):360–378.

Palmer, C. (1997). Music performance. *Annual Reviews in Psychology*, 48:115–138.

Peeters, G. (2004). A large set of audio features for sound description (similarity and classification) in the CUIDADO project. *CUIDADO IST Project Report*, (version 1.0):1–25.

Rabiner, L. R. (1990). A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. In Waibel, A. and Lee, K.-F., editors, *Readings in Speech Recognition*, pages 267–296. Kaufmann, San Mateo, CA.

Rank, E. (1999). A player model for MIDI control of synthetic bowed strings. In Feichtinger, H. and Dörfler, M., editors, *Diderot Forum on Mathematics and Music, Computational and Mathematical Methods in Music*, pages 297–300.

Rasamimanana, N., Bevilacqua, F., Schnell, N., Guedy, F., Flety, E., Maestracci, C., Zamborlin, B., Frechin, J.-L., and Petreski, U. (2011). Modular Musical Objects Towards Embodied Control Of Digital Music. In *TEI*, pages 9–12.

Rasamimanana, N. (2008). *Geste instrumental du violoniste en situation de jeu : analyse et modélisation.* PhD thesis, Université Paris 6 - IRCAM UMR STMS.

Rasamimanana, N., Bernardin, D., Wanderley, M., and Bevilacqua, F. (2008). String Bowing Gestures at Varying Bow Stroke Frequencies: A Case Study. In *Advances in Gesture-Based Human-Computer Interaction and Simulation*, volume 5085 of *Lecture Notes in Computer Science*, pages 216–226. Springer Verlag.

Rasamimanana, N. and Bevilacqua, F. (2009). Effort-based analysis of bowing movements: evidence of anticipation effects. *The Journal of New Music Research*, 37(4):339–351.

Rasamimanana, N., Fléty, E., and Bevilacqua, F. (2006). Gesture Analysis of Violin Bow Strokes. In *Gesture in Human-Computer Interaction and Simulation*, volume 3881 of *Lecture Notes in Computer Science / Artificial Intelligence (LNAI)*, pages 145–155. Springer Verlag.

Rasamimanana, N., Kaiser, F., and Bevilacqua, F. (2009). Perspectives on gesture-sound relationships informed from acoustic instrument studies. *Organised Sound*, 14(2):208–216.

Schnell, N., Bevilacqua, F., Rasamimanana, N., Bloit, J., Guedy, F., and Flety, E. (2011). Playing the "MO": Gestural Control and Re-Embodiment of Recorded Sound and Music. In *NIME*, pages 535-536

Schoonderwaldt, E. (2008). *Mechanics and acoustics of violin bowing : Freedom, constraints and control in performance*. PhD thesis, KTH, Speech, Music and Hearing.

Schoonderwaldt, E. and Demoucron, M. (2009). Extraction of bowing parameters from violin performance combining motion capture and sensors. *The Journal of the Acoustical Society of America*, 126(5):2695–708.

Sinclair, S., Malloch, J., Rudraraju, V., Nika, J., and Wanderley, M. M. (2010). LibMapper (http://www.idmil.org/software/libmapper).

Uno, Y., Suzuki, Y., and Kawato, M. (1989b). Minimum muscle-tension-change model which reproduces human arm movement. *Proceedings of the 4th Symposium on Biological and Physiological Engineering*, pages 299–302.

Uno, Y., Kawato, M., and Suzuki, R. (1989). Formation and control of optimal trajectory in human multijoint arm movement. *Biological cybernetics*, 61(2):89–101.

Wanderley, M. and Depalle, P. (2004). Gestural control of sound synthesis. In *Proceedings of the IEEE*, volume 92, pages 632–644.

Wanderley, M., Vines, B., Middleton, N., McKay, C., and Hatch, W. (2005). The musical significance of clarinetists' ancillary gestures: An exploration of the field. *The Journal of New Music Research*, 34(1):97–113.

Wessel, D. and Wright, M. (2002). Problems and prospects for intimate musical control of computers. *The Computer Music Journal*, 26(3):11–22.

Young, D. (2007). *A Methodology for Investigation of Bowed String Performance Through Measurement of Violin Bowing Technique*. PhD thesis, M.I.T.

Young, D. (2008). Classification of Common Violin Bowing Techniques Using Gesture Data from a Playable Measurement System. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, pages 44–48.

List of Figures

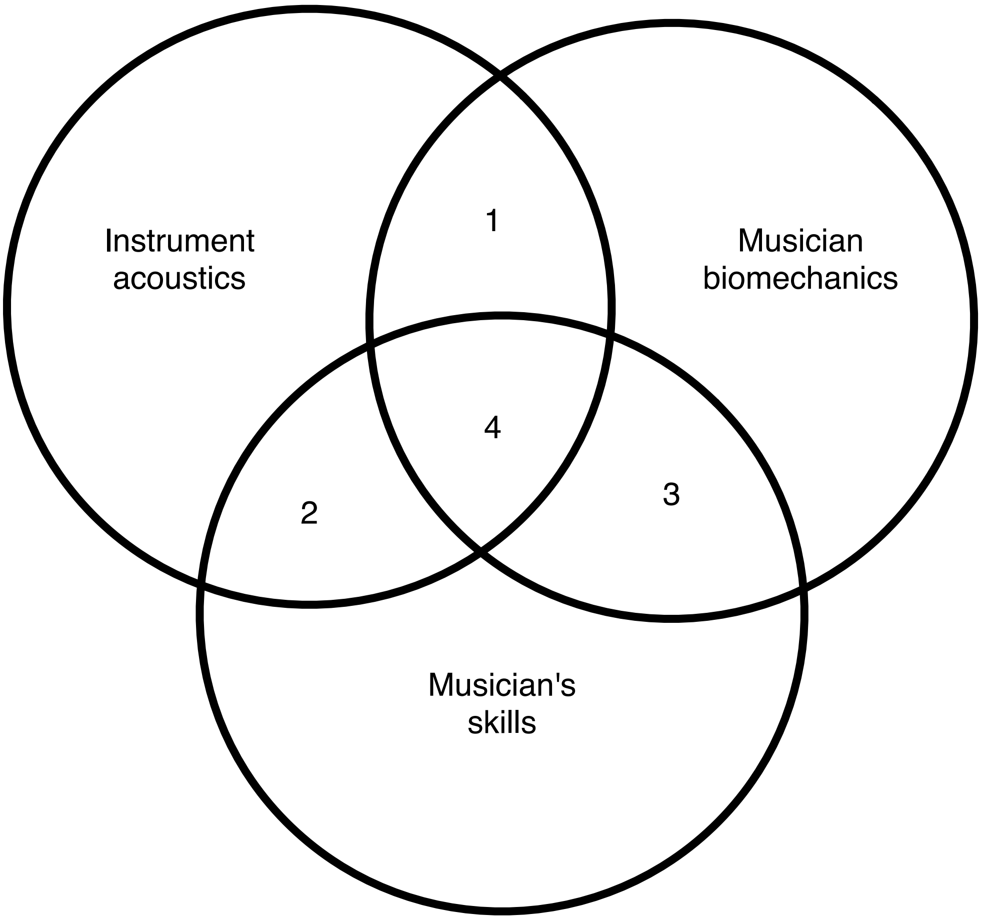


Figure 1: Definition of a gesture space of possibilities: crossing between acoustic constraints (e.g. bow velocity, bow force), biomechanical constraints (e.g. minimum jerk, equilibrium point, minimum torque change) with musician's skills (e.g. experience, level of gesture control). Zone 1: unexploited combinations due to lack of skills. Zone 2: unplayable, non biomechanical combinations. Zone 3: beyond instrument possibilities, room for electronic augmentation. Zone 4: exploited combinations where musicians develop various strategies of actions.

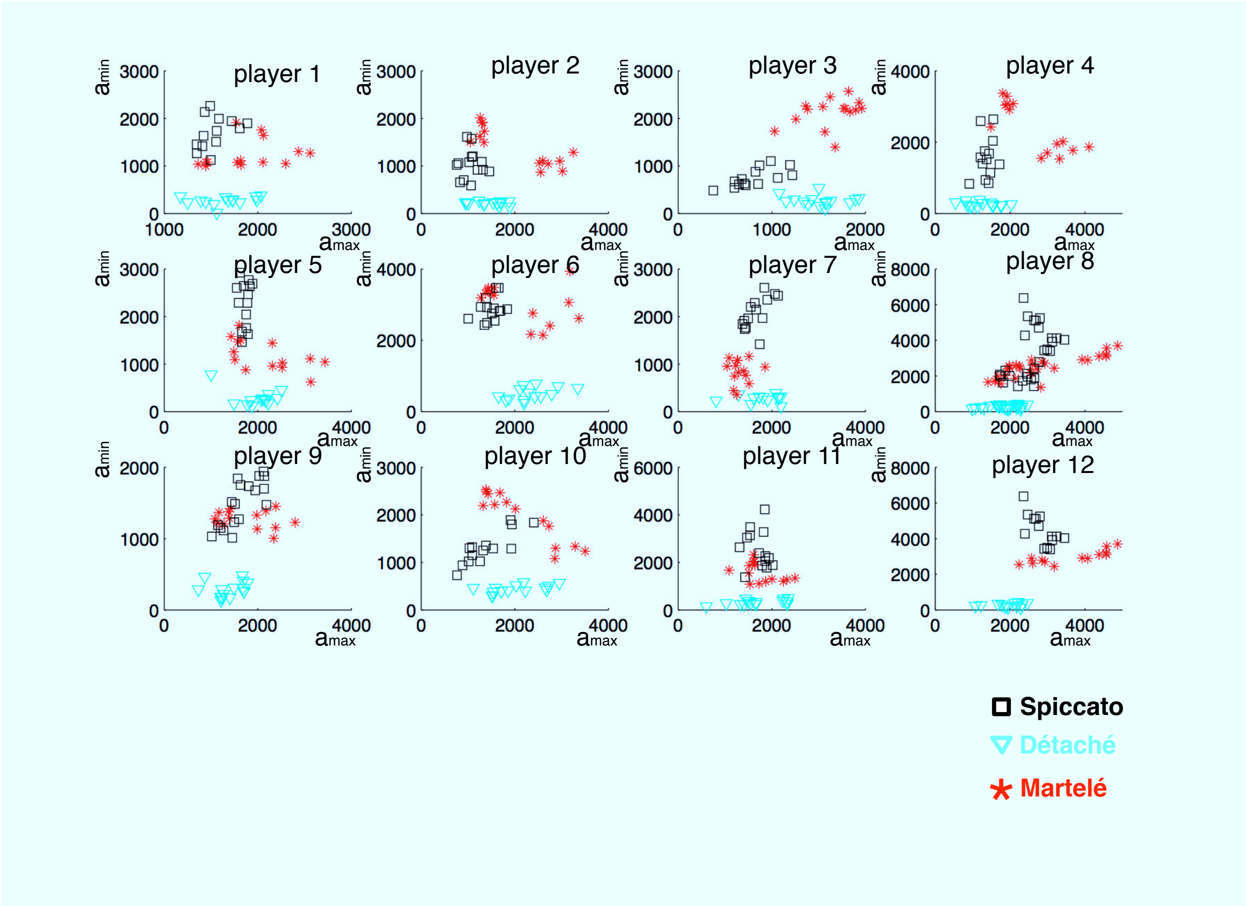
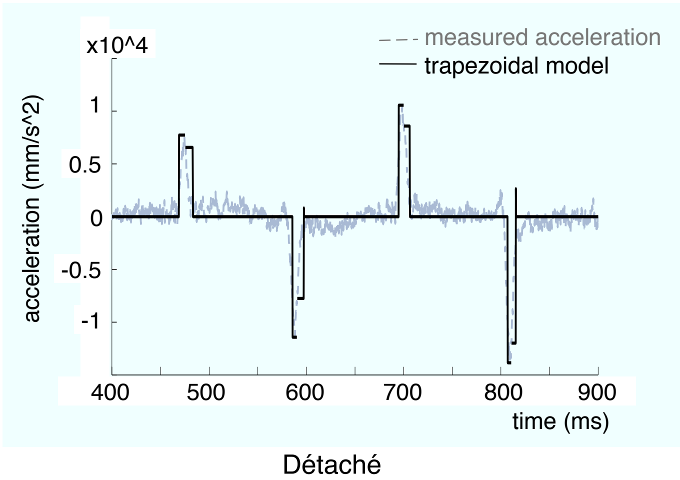


Figure 2: Projection of violin expressive gesture space of possibilities on the parameter space defined by maximum and minimum accelerations within a stroke. Twelve violinists represented playing scales in *Détaché*, *Martelé* and *Spiccato*. (arbitrary units)



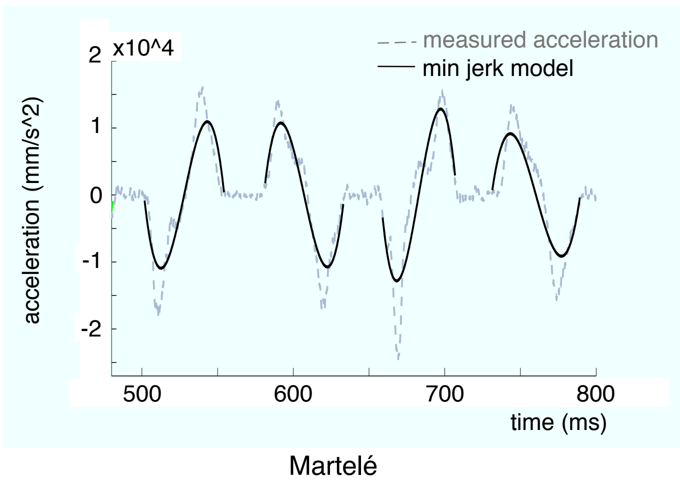


Figure 3: Continuous modelling of expressive gestures. Gestures are modeled using two components: a model for the action of moving the bow and some additional contraints to express how the bow is moved. Here are plotted accelerations for a set of four strokes in *Détaché* (top) and in *Martelé* (bottom), with the corresponding continuous models, resp. trapezoidal (minimum impulse) model and minimum jerk model. Extracted from (Rasamimanana and Bevilacqua, 2009)

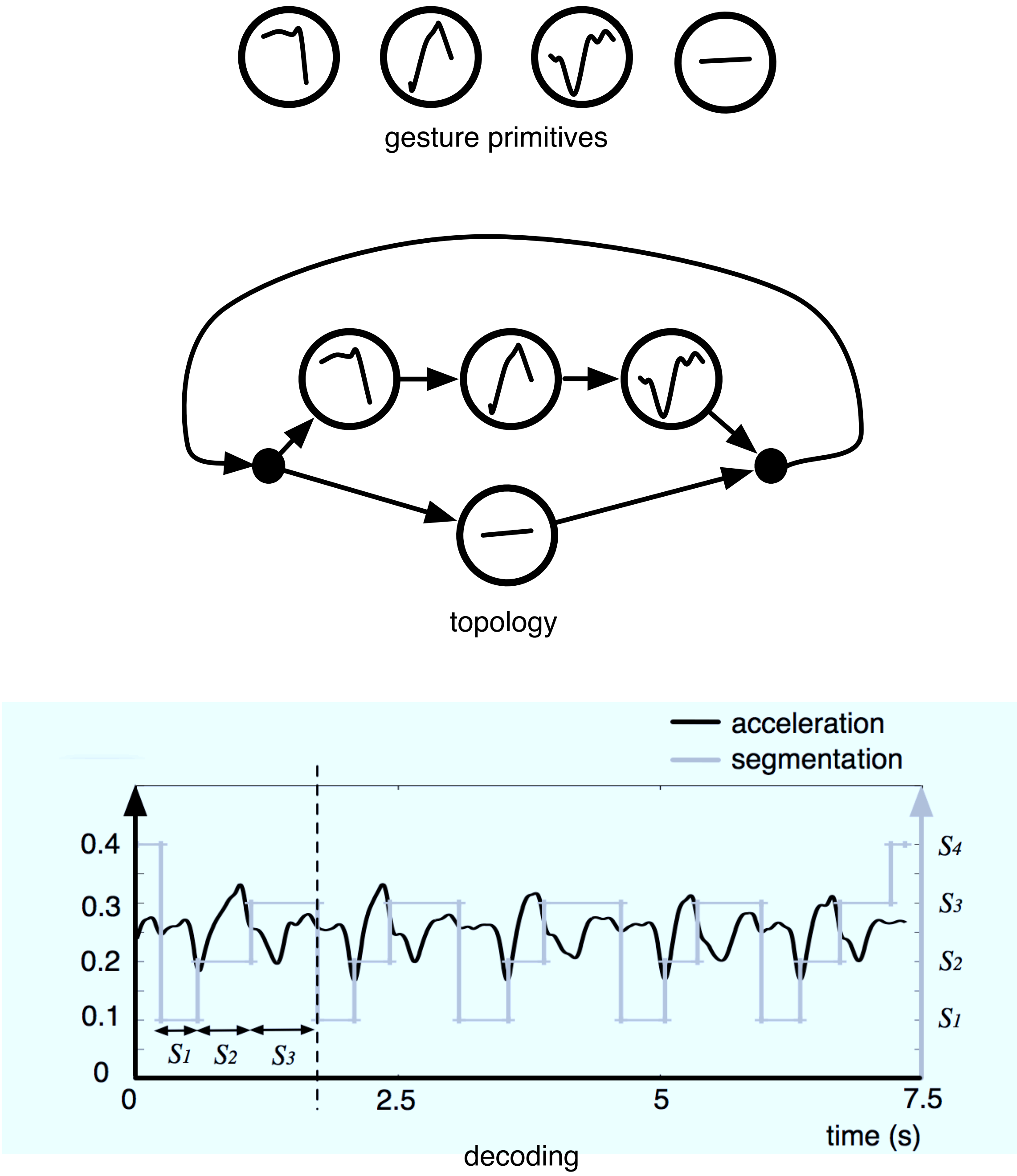


Figure 4: Segmental modelling of gestures. Top: set of 3 sampled gesture primitives and one parametric primitive. Middle: Model topology that defines possible primitives sequences. Bottom: decoding result on an acceleration signal.