Chapter 8
Online Gesture Analysis and Control of Audio Processing

Frédéric Bevilacqua, Norbert Schnell, Nicolas Rasamimanana,
Bruno Zamborlin, and Fabrice Guédy

Abstract. This chapter presents a general framework for gesture-controlled audio processing. The gesture parameters are assumed to be multi-dimensional temporal profiles obtained from movement or sound capture systems. The analysis is based on machine learning techniques, comparing the incoming dataflow with stored templates. The mapping procedures between the gesture and the audio processing include a specific method we called temporal mapping. In this case, the temporal evolution of the gesture input is taken into account in the mapping process. We describe an example of a possible use of the framework that we experimented with in various contexts, including music and dance performances, music pedagogy and installations.

8.1 Introduction

The role of gesture control in musical interactive systems has constantly increased over the last ten years as a direct consequence of both new conceptual and new technological advances. First, the fundamental role of physical gesture in human-machine interaction has been fully recognized, influenced by theories such as enaction or embodied cognition [24]. For example, Leman laid out an embodied cognition approach to music [38], and, like several other authors, insisted on the role of action in music [27, 28 and 35]. These concepts resonate with the increased availability of cost-effective sensors and interfaces, which also drive the development of new musical digital instruments where the role of physical gesture is central. From a research perspective, important works have been reported in the

Frederic Bevilacqua · Norbert Schnell · Nicolas Rasamimanana ·
Bruno Zamborlin · Fabrice Guedy
IRCAM - Real Time Musical Interactions Team, STMS IRCAM-CNRS-UPMC,
Place Igor Stravinsky, Paris, 75004, France
e-mail: {Frederic.Bevilacqua,Norbert.Schnell,Nicolas.Rasamimanana,
Bruno.Zamborlin,Fabrice.Guedy}@ircam.fr

J. Solis and K.C. Ng (Eds.): Musical Robots and Interactive Multimodal Systems, STAR 74, pp. 127–142.
springerlink.com

community at the NIME conferences (New Interfaces for Musical Expressions) since 2001.

Using gesture-controlled audio processing naturally raises the question about the relationship between the gesture input data and the sound control parameters. This relationship has been conceptualized as a ”mapping” procedure between different input-output parameters. Several approaches of gesture-sound mapping have been proposed [1, 33, 34, 39, 63, 66 and 67]. The mapping can be described for low-level parameters, or for high-level descriptors that can be comprehended from a cognitive standpoint (perceived or semantic levels) [15, 32, 33 and 65]. Some authors also proposed to use emotional characteristics that could be applied to both gesture and sound [16 and 19].

Most often in mapping procedures, the relationship is established as instantaneous, i.e. the input values at any given time are linked to output parameters [1, 23, 34, 62 and 60]. The user must dynamically modify the control parameters in order to imprint any corresponding time behavior of the sound evolution. Similarly to traditional instruments, this might involve important practice to effectively master such a control. Nevertheless, mapping strategies that directly address the evolution of gesture and sound parameters, e.g. including advanced techniques of temporal modeling, are still rarely used. Modeling generally includes statistical measures over buffered data. While such methods, inspired by the “bag of words” approach using gesture features or “bag of frames” using sound descriptors [3], can be powerful for a classification task, they might be unfit for real-time audio control where the continuous time sequence of descriptors is crucial.

We present here a framework that we developed for gesture-controlled interactive audio processing. By refining several prototypes in different contexts, music pedagogy, music and dance performances, we developed a generic approach for gesture analysis and mapping [6, 7, 10, 29, 30 and 52]. From a technical point of view, this approach is based on mapping strategies using gesture recognition techniques, as also proposed by others [4, 5, 25 and 22]. Our approach is based on a general principle: the gestures are assumed to be temporal processes, characterized by temporal profiles. The gesture analysis is thus based on a tool that we specifically developed for the analysis of temporal data in real-time, called the *gesture follower*. Furthermore, we introduced the notion of *temporal mapping*, as opposed to *spatial mapping*, to insist on the temporal aspects of the relationship between gesture, sound and musical structures [10 and 54]. This distinction between spatial and temporal views of control was also pointed out by Van Nort [63]. *Temporal mapping* brings to focus the time evolution of the data rather than their absolute values for the design of musical interaction systems. This approach relies on the modeling of specific temporal profiles or behavior obtained through either a training process or set manually [11, 20, 23, 40, 51 and 53].

This chapter is structured as follows: first, we present the general architecture, followed by a description of the gesture recognition system and temporal mapping procedure; finally, we describe one possible interaction paradigm with this framework that was used in several different applications.

8.2 Gesture Capture and Processing

The general architecture of our system is illustrated in Fig. 8.1. Similar to the description of digital musical instruments given by [66], we considered four distinct parts: gesture capture, gesture processing, mapping, and audio processing. The specific elements of the gesture analysis and mapping are described in detail in the next sections. Note that this architecture is supported by a set of software libraries that greatly facilitates its design and rapid prototyping [9, 56, 57 and 58], in providing tools for the storage and processing of time-based data, such as gesture and sound data.

8.2.1 Gesture Capture System

The system was designed for a broad range of input data, obtained from various heterogeneous capture systems (see [42] for a review). We cite below the different types of systems that we have involved in our projects, clearly illustrating the variety of data types that can be taken into account.

- 3D spatial data obtained using motion capture data
- Image analysis parameters obtained from video capture
- Gesture data obtained from inertial measurement units, including any combination of accelerometers, gyroscopes and magnetometers (obtained through prototypes or commercial devices including game and mobile phone interface)
- Gesture data obtained from sensors such as FSR (Force Sensing Resistor), bend, strain gauge, piezoelectric, Hall, optical, ultrasound sensors
- Tablet and multitouch interfaces
- Sliders and potentiometers
- MIDI interfaces
- Sound descriptors derived from sound capture
- Any combination of the above

Taking into account the heterogeneity of possible input data, the term "gesture" can refer to completely different types of physical quantities. Nevertheless, any of these inputs can still be considered as multidimensional temporal data. In order to work efficiently with such different input, we separate the data stream analysis into two procedures: 1) the *preprocessing* that is specific to each input system and 2) the *online temporal profile analysis* that is generic (as illustrated in Figure 8.1).

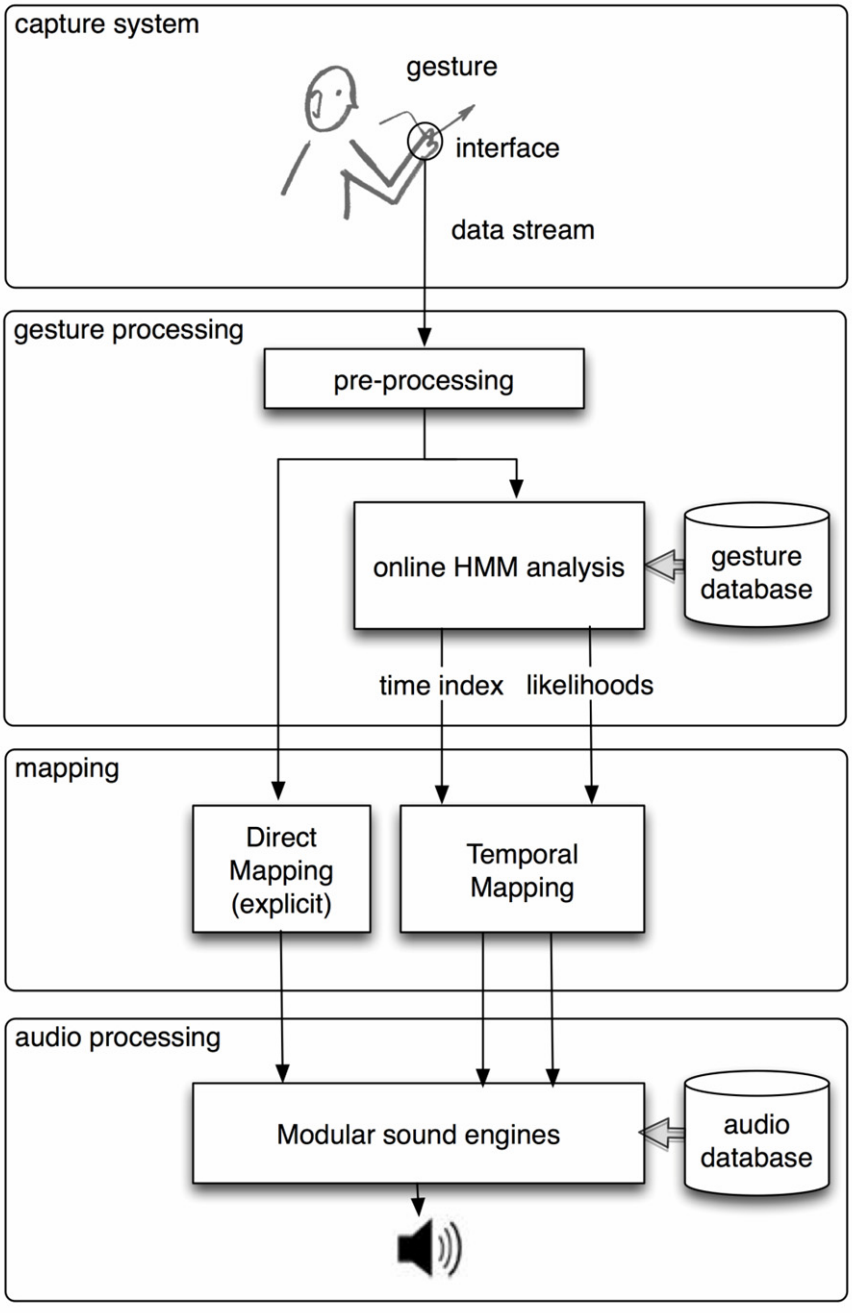

Fig. 8.1 General architecture for gesture-controlled audio processing

8.2.2 Pre-processing

The pre-processing procedure includes:

- Filtering (e.g. low/high/band pass)
- Re-sampling to ensure a constant sampling rate
- Data fusion and dimension reduction (e.g. using Principal Component Analysis)
- Normalization
- Segmentation (optional)

The pre-processing is designed specifically for a given gesture capture system, taking into account its particularities. The preprocessing should ensure that the data is formatted as a temporal stream of vectors x(t) of dimension M, regularly sampled over a time interval Δt. Thus, the preprocessed data can be seen as series $\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_n$. A recorded gesture from $t = 0$ to $t = (N-1)\Delta t$ is stored in a matrix G of dimension $N \times M$.

8.2.3 Temporal Profiles Processing

The processing is based on machine learning techniques, and thus requires proceeding in two steps. The first step corresponds to the training of the system using a database: this is called learning procedure. During this step, the system computes model parameters based on the database (described in the next section). The second step is the online processing, which corresponds to the actual use of the system during performance. During this step, the system outputs parameters used for the audio control.

8.2.3.1 Modeling and Learning Procedure

Let us first summarize our requirements. First, we need a very fine-grained time modeling system so that we can consider "gesture data" profiles at several time scales. This is desirable when dealing with musical gestures [39]. Second, for practical reasons, we wish to be able to use a single example in the learning process. This is necessary to ensure that the gesture vocabulary can be set very efficiently or easily adaptable to idiosyncrasies of a particular performer. We found that this requirement is of crucial importance when working in pedagogical or artistic contexts.

These requirements led us to develop a hybrid approach between methods such as Hidden Markov Models (HMM), Dynamic Time Warping (DTW) and Linear Dynamic Systems (LDS) [41, 43, 46, 47, 61 and 68]. Interestingly, Rijko and coworkers also working in a performing arts context, proposed similar approaches [48, 49 and 50].

DTW is a common method for gesture recognition consisting in temporally aligning a reference and a test gesture, using dynamic programming techniques.

This allows for the computation of a similarity measure between different gestures that is invariant to speed variation. HMM have also been widely used for gesture recognition. They can also account for speed variations, while benefiting for a more general formalism. HMM allows for data modeling with a reduced number of hidden states, which can be characterized through a training phase with a large number of examples. While more effective, the standard HMM implementation might suffer from coarse time modeling. This might be improved using Semi-Markov Model [21] or Segmental HMM [2, 12 and 13].

We proposed a method that is close to DTW, in the sense that we keep the usage simplicity of comparing the test gesture with a single reference example, as well as the advantage of applying fine-grained time warping procedure. Nevertheless, we present our method using an HMM formalism [47], for the convenience provided by a probabilistic approach. As such, our method is based on the forward procedure in order to satisfy our constraint of real-time computation, while standard DTW techniques require operating on completed gesture [5]. Nevertheless, to guarantee our requirements, i.e. a fine-grained time modeling and a simplified training procedure, we adopted a non-standard HMM implementation, previously described in [7, 8 and 10].

We recall here only the main features of our implementations. Similar to example-based methods, we associate each gesture template to a state-based structure: each data sample represents a "state" [14]. Furthermore, a probability density function is associated to each state, setting the observation probability of the data. This structure can then be associated to an HMM (Figure 8.2).

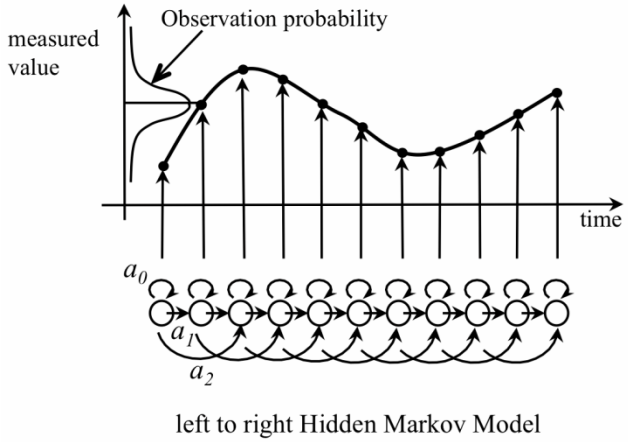

Fig. 8.2 HMM structure associated to a gesture template

The fact that the sampling rate is regular simplifies the learning procedure. In this case, all transition coefficients of the same type between states (a_0=stay, a_1=next, a_2=skip, etc) must share identical values, which can be manually set using

prior knowledge. See [10] for choice examples in setting these transition probability values.

The learning procedure consists simply of recording at least one gesture, stored in a matrix G ($N \times M$). Each matrix element is associated to the mean μ_i of a normal probability function b_i, corresponding to the observation probability function. Using several gesture templates corresponds to recording and storing of an array of G_k matrices.

The value of the variance can also be set using prior knowledge. For example, prior experiments can establish typical variance values for a given type of gestures and capture systems. A global scaling factor, which operates on all the variance values, can be manually adjusted by the user.

8.2.3.2 Online Processing: Time Warping and Likelihood Estimation

The gesture follower operates in real-time on the input dataflow. The live input is continuously compared to the templates stored in the system using an HMM online decoding algorithm. The gesture follower provides in real-time two types of parameters that are used in the mapping procedure.

First, during the performance, it reports the time progression index of the live gesture given a pre-recorded template. This corresponds to computing the time warping during the performance as shown in Fig. 8.3. We call this procedure "following", since it is similar to the paradigm of score following [21 and 59]. Note that as explained above, in the case of the gesture follower, the Markov chains are built based on recorded templates, while in the case of score following the Markov chains are built using a symbolic score.

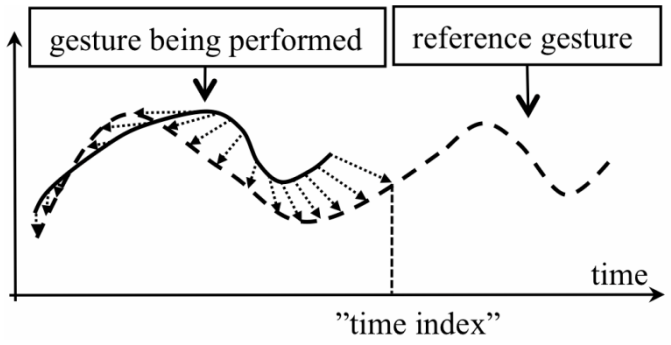

Fig. 8.3 Online time warping of the gesture profile

Second it returns the *likelihood* value that can be interpreted as the probability of the observed gesture being generated by the template. This can be used as a similarity measure between the gesture being performed and the templates [18].

The decoding is computed using the forward procedure [47]. The state probability distribution $\alpha_t(i)$ of a series of observations $O_1....O_t$ is computed as follows, considering the observation probability $b_i(O_t)$:

1. Initialization

$$\alpha_1(i) = \pi_i b_i(O_1) \qquad 1 \le i \le N$$

where π_i is the initial state distribution.

2. Induction

$$b_i(O_t) = \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left[-(O_t - \mu_i)^2 / 2\sigma_i^2\right]$$

$$\alpha_{t+1}(i) = \left[\sum_{j=1}^{N} \alpha_t(j)\, a_{ij}\right] b_i(O_t) \qquad 1 \le t \le T-1,\ 1 \le i \le N$$

where a_{ij} are the state transition probabilities (note that $b_i(O_t)$ is expressed here for the case M=1, the generalization to M>1 is straightforward)

The time progression index and the likelihood associated to the observation series is updated at each new observation from the $\alpha_t(t)$ distribution:

$$\textit{time progression index}(t) = \arg\max\left[\alpha_t(i)\right]$$

$$\textit{likelihood}(t) = \sum_{i=1}^{N} \alpha_t(i)$$

These parameters are output by the system continuously, from the beginning of the gesture. The likelihood parameters are thus available before the end of the gestures, which could allow for the determination of early recognition, as also proposed by Mori [45].

Generally, the computation is run in parallel for k templates, returning k values of the time progression index and the likelihood. The *argmax*[*likelihood(t)*] returns the likeliest template for the current gesture at time t.

8.3 Temporal Mapping

Different types of mappings have been identified, such as *explicit* or *implicit* mapping [63]. In explicit mapping, the mathematical relationships between input and output are directly set by the user. On the contrary, indirect mapping generally refers to the use of machine learning techniques, implying a training phase to set parameters that are not directly accessed by the user [17, 22 and 25].

As shown in Fig. 8.1,, our architecture includes both explicit and implicit types of mapping that are operated simultaneously. Explicit mapping has been largely discussed and is commonly used. We discuss below more specifically the implicit mapping used in conjunction with the *gesture follower*. Precisely, as discussed previously, we introduce a *temporal mapping* procedure by establishing relationships between the gesture and audio temporal profiles. The temporal mapping can be seen as a synchronization procedure between the input gesture parameters and the sound process parameters. Technically, this is made possible by the use of the time progression index that the gesture follower provides continuously: the pacing of the gesture can therefore be synchronized with specific time processes.

Fig. 8.4 illustrates a simple example of temporal mapping: two temporal profiles are mapped together, namely hand acceleration and audio loudness. Please note that hand acceleration and audio loudness were chosen here for the sake of clarity, and that any gesture data or audio processing parameters could be used. In this example the gesture data is composed of three phases illustrated by colored regions: one first phase (dark grey), an exact repetition of this first phase (dark grey), and a third different phase (light grey). The audio loudness data is structured differently: it is constantly increasing up to a maximum value, and then constantly decreasing.

The temporal mapping consists here in explicitly setting a relationship between these two temporal profiles: values of acceleration and loudness are linked together according to their history. The mapping is therefore dependant on a sequence of values rather than on independent single values. For example, vertical lines in Fig. 8.4 all indicate similar acceleration values. In particular, lines 1 and 3 (resp. 2 and 4) have strictly identical values (all derivatives equal) as they correspond to a repeated gesture. One can note that interestingly these identical gesture values are mapped to different audio values, as shown by the values pointed by lines 1 and 3 (resp. 2 and 4). In our case, the mapping clearly depends on the time sequencing of the different gesture values and must be considered as a time process. While gesture reference profiles are generally directly recorded from capture systems, audio input profiles can be created by the user manually, or derived from other modeling procedures. For example, recorded bowing velocity profiles could be used as parameters to control physical models. Using the temporal mapping, these velocity profiles could be synchronized to any other gesture profile. This is especially interesting when the target velocity profile might be difficult to achieve, due to biomechanical constraints, or particularities in the response of the capture system (e.g., non linearities).

This formalization of temporal mapping includes also the possibility to use particular temporal markers to trigger processes at specific times (generally associated with the use of cue-lists). For example, in Fig. 8.4, line 5 (as well as any other) could be used as such a marker. Therefore, temporal mapping can be extended to the control of a combination of continuous time profiles and discrete time events, synchronized to the input gesture data.

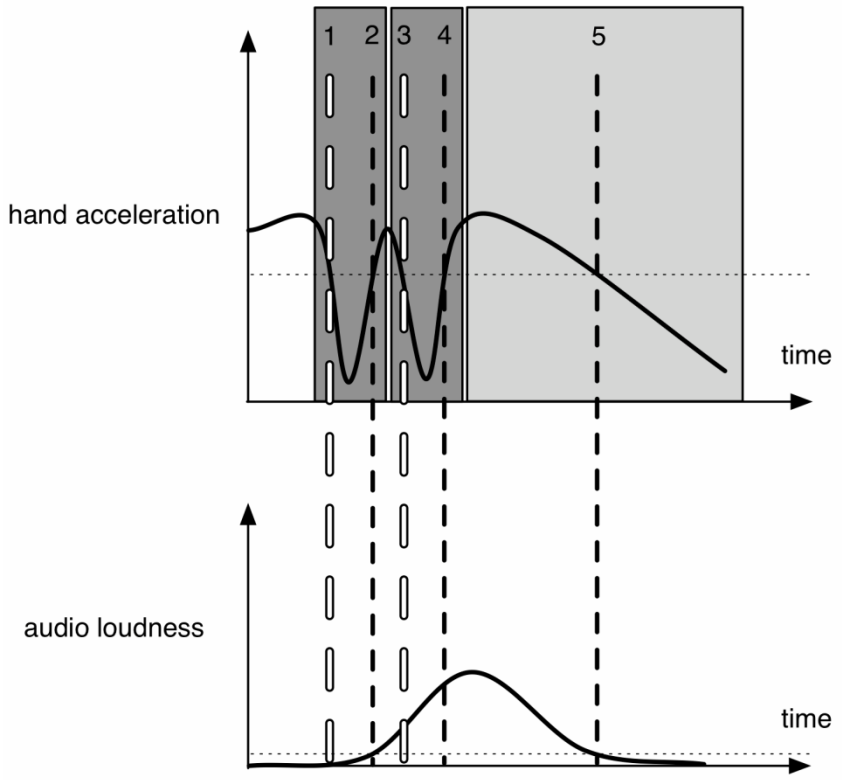

Fig. 8.4 Toy example of temporal mapping between hand acceleration audio loudness. Mapped gesture data is composed of three phases: one first phase (dark grey), a repetition of this first phase (dark grey) and a third different phase (light grey). Vertical lines indicate similar gesture values.

Momeni and Henry previously described an approach that also intrinsically takes into account temporal processes in the mapping procedure. Precisely, they used physical models to produce dynamic layers of audio control [31 and 44]. In our approach, we rather leave users free to define any shape for the audio control, whether based on physical dynamics or designed by other means.

8.4 Example Use with Phase-Vocoder Techniques

The direct combination of the gesture follower with an advanced phase-vocoder system [55] allows for the implementation of a set of applications where a gesture can control continuously the playing speed of an audio file [7 and 10]. In practice, this type of application can be efficiently built using the method illustrated in Fig. 8.5.

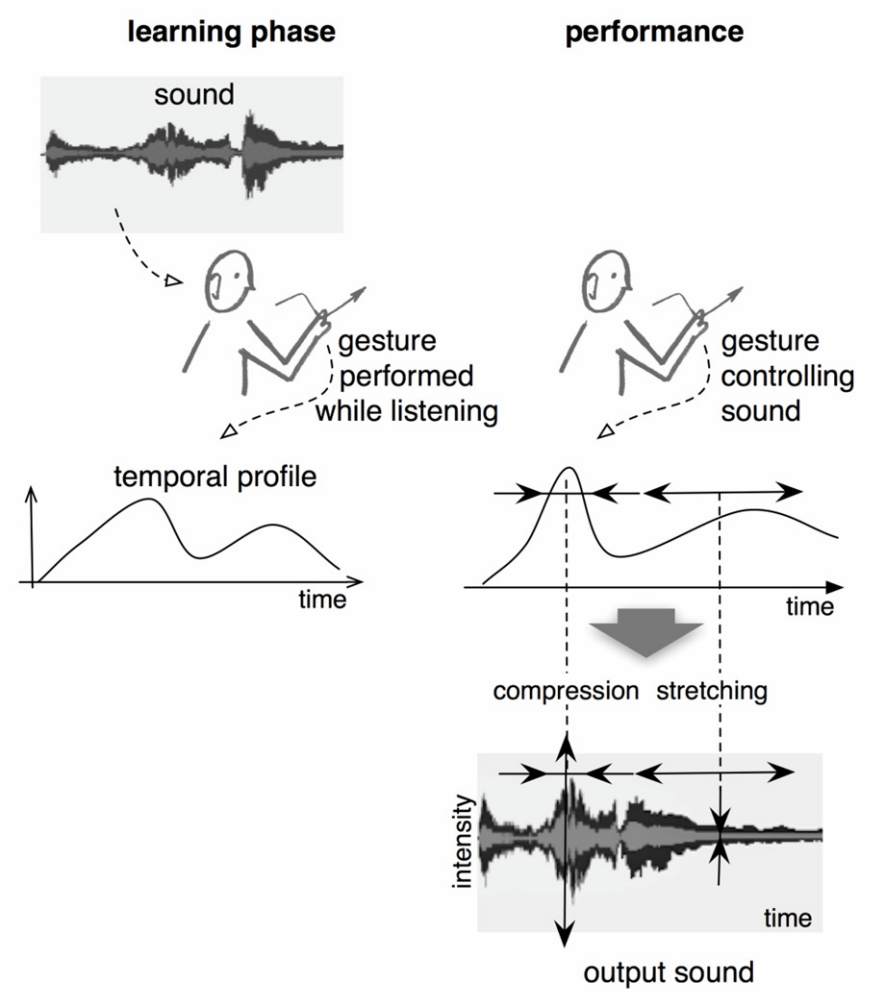

Fig. 8.5 Possible use of the general framework: the user first proposes a gesture while listening to the sound, and then play the sound with temporal and spatial variations.

First, the user must record a gesture while listening to an audio file. This step is necessary for the system to learn a gesture template that is actually synchronous with the original audio recording. The audio recording and the gesture can be several minutes long.

The second step corresponds for the user to re-perform the "same gesture", but introducing speed and intensity variations compared to the reference template. The gesture follower is used to temporally synchronize the gesture with the rendering of the audio file. In other words, the temporal mapping procedure allows for setting a direct correspondence between the gesture time progression and the audio

time progression. The audio rendering is performed using a phase-vocoder, ensuring that only the audio playback speed is changed while preserving the pitch and timbre of the original audio recording.

This application allows for the design of a "conducting scenario", where the gesture chosen by the user can be used to control the playing speed of a recording. From an application perspective, this could be applied to "virtual conducting" applications (see for example [36 and 37]. The advantage of our application resides in the fact that the gesture can be freely chosen by the user (e.g. dance movements) by simply recording it once. Precisely, this application can accommodate directly both standard conducting gestures and original gestures invented by the user, which can lead to novel interaction design strategies.

8.5 Conclusion

We described a general framework for gesture-controlled audio processing which has been experimented with in various artistic and pedagogical contexts. It is based on an online gesture processing system, which takes into account temporal behavior of gesture data, and a temporal mapping procedure. The gesture processing makes use of a machine learning technique, and requires pre-recorded gesture templates. Nevertheless, the learning procedure operates using a single recording, which makes the learning procedure simple and easily adaptable to various gesture capture systems. Moreover, our approach could be complemented using Segmental HMM [2, 12 and 13] or hierarchical HMM [26] in modeling transitions between gesture templates, which is missing in the current approach. This would allow for a higher structural level of sound control.

Acknowledgements

We acknowledge partial support of the following projects: the EU-ICT projects SAME and i-Maestro, the ANR (French National Research Agency) projects EarToy ANR-06-RIAM-004 02)and Interlude (ANR-08-CORD-010). We would like to thank Riccardo Borghesi for his crucial role in software development and the students of the "Atelier des Feuillantines" for contributing to experiments.

References

1. Arfib, D., Couturier, J., Kessous, L., Verfaille, V.: Strategies of mapping between gesture data and synthesis model parameters using perceptual spaces. Organized Sound 7(2), 127–144 (2002)
2. Artieres, T., Marukatat, S., Gallinari, P.: Online handwritten shape recognition using segmental hidden markov models. IEEE Transactions on Pattern Analysis and Machine Intelligence 29(2), 205–217 (2007)
3. Aucouturier, J., Daudet, L.: Editorial: Pattern recognition of non-speech audio. Pattern Recognition Letters 31(12), 1487–1488 (2010)

4. Bell, B., Kleban, J., Overholt, D., Putnam, L., Thompson, J., Kuchera-Morin, J.: The multimodal music stand. In: NIME 2007: Proceedings of the 7th International Conference on New Interfaces for Musical Expression, pp. 62–65 (2007)
5. Bettens, F., Todoroff, T.: Real-time DTW-based Gesture Recognition External Object for Max/MSP and PureData. In: SMC: Proceedings of the 6th Sound and Music Computing Conference, pp. 30–35 (2009)
6. Bevilacqua, F.: Momentary notes on capturing gestures. In (capturing intentions). Emio Greco/PC and the Amsterdam School for the Arts (2007)
7. Bevilacqua, F., Guédy, F., Schnell, N., Fléty, E., Leroy, N.: Wireless sensor interface and gesture-follower for music pedagogy. In: NIME 2007: Proceedings of the 7th International Conference on New Interfaces for Musical Expression, pp. 124–129 (2007)
8. Bevilacqua, F., Muller, R.: A gesture follower for performing arts. In: Proceedings of the International Gesture Workshop (2005)
9. Bevilacqua, F., Muller, R., Schnell, N.: MnM: a max/msp mapping toolbox. In: NIME 2005: Proceedings of the 5th International Conference on New Interfaces for Musical Expression, pp. 85–88 (2005)
10. Bevilacqua, F., Zamborlin, B., Sypniewski, A., Schnell, N., Guédy, F., Rasamimanana, N.: Continuous Realtime Gesture Following and Recognition. In: Kopp, S., Wachsmuth, I. (eds.) GW 2009. LNCS, vol. 5934, pp. 73–84. Springer, Heidelberg (2010)
11. Bianco, T., Freour, V., Rasamimanana, N., Bevilaqua, F., Caussé, R.: On Gestural Variation and Coarticulation Effects in Sound Control. In: Kopp, S., Wachsmuth, I. (eds.) GW 2009. LNCS, vol. 5934, pp. 134–145. Springer, Heidelberg (2010)
12. Bloit, J., Rasamimanana, N., Bevilacqua, F.: Modeling and segmentation of audio descriptor profiles with segmental models. Pattern Recognition Letters 31, 1507–1513 (2010)
13. Bloit, J., Rasamimanana, N., Bevilacqua, F.: Towards morphological sound description using segmental models. In: Proceedings of DAFx (2009)
14. Bobick, A.F., Wilson, A.D.: A state-based approach to the representation and recognition of gesture. IEEE Transactions on Pattern Analysis and Machine Intelligence 19(12), 1325–1337 (1997)
15. Camurri, A., De Poli, G., Friberg, A., Leman, M., Volpe, G.: The mega project: analysis and synthesis of multisensory expressive gesture in performing art applications. The Journal of New Music Research 34(1), 5–21 (2005)
16. Camurri, A., Volpe, G., Poli, G.D., Leman, M.: Communicating expressiveness and affect in multimodal interactive systems. IEEE MultiMedia 12(1), 43–53 (2005)
17. Caramiaux, B., Bevilacqua, F., Schnell, N.: Towards a Gesture-Sound Cross-Modal Analysis. In: Kopp, S., Wachsmuth, I. (eds.) GW 2009. LNCS, vol. 5934, pp. 158–170. Springer, Heidelberg (2010)
18. Caramiaux, B., Bevilacqua, F., Schnell, N.: Analysing Gesture and Sound Similarities with a HMM-based Divergence Measure. In: Proceedings of the Sound and Music Computing Conference, SMC (2010)
19. Castellano, G., Bresin, R., Camurri, A., Volpe, G.: Expressive control of music and visual media by full-body movement. In: NIME 2007: Proceedings of the 7th International Conference on New Interfaces for Musical Expression, pp. 390–391 (2007)
20. Chafe, C.: Simulating performance on a bowed instrument. In: Current Directions in Computer Music Research, pp. 185–198. MIT Press, Cambridge (1989)

21. Cont, A.: Antescofo: Anticipatory synchronization and control of interactive parameters in computer music. In: Proceedings of the International Computer Music Conference, ICMC (2008)
22. Cont, A., Coduys, T., Henry, C.: Real-time gesture mapping in pd environment using neural networks. In: NIME 2004: Proceedings of the 2004 Conference on New Interfaces for Musical Expression, pp. 39–42 (2004)
23. Demoucron, M., Rasamimanana, N.: Score-based real-time performance with a virtual violin. In: Proceedings of DAFx (2009)
24. Dourish, P.: Where the action is: the foundations of embodied interaction. MIT Press, Cambridge (2001)
25. Fels, S., Hinton, G.: Glove-talk: a neural network interface between a data-glove and a speech synthesizer. IEEE Transactions on Neural Networks 3(6) (1992)
26. Fine, S., Singer, Y.: The hierarchical hidden markov model: Analysis and applications. Machine Learning 32(1), 41–62 (1998)
27. Godøy, R., Haga, E., Jensenius, A.R.: Exploring music-related gestures by sound-tracing - a preliminary study. In: 2nd ConGAS International Symposium on Gesture Interfaces for Multimedia Systems, Leeds, UK (2006)
28. Godøy, R., Leman, M. (eds.): Musical Gestures: Sound, Movement and Meaning. Routledge, New Yark (2009)
29. Guédy, F., Bevilacqua, F., Schnell, N.: Prospective et expérimentation pédagogique dans le cadre du projet I-Maestro. In: JIM 2007-Lyon
30. Guédy, F.: Le traitement du son en pédagogie musicale, vol. 2. Ircam – Editions Léo, L'Inouï (2006)
31. Henry, C.: Physical modeling for pure data (pmpd) and real time interaction with an audio synthesis. In: Proceedings of the Sound and Music Computing Conference, SMC (2004)
32. Hoffman, M., Cook, P.R.: Feature-based synthesis: Mapping from acoustic and perceptual features to synthesis parameters. In: Proceedings of International Computer Music Conference, ICMC (2006)
33. Hunt, A., Wanderley, M., Paradis, M.: The importance of parameter mapping in electronic instrument design. The Journal of New Music Research 32(4) (2003)
34. Hunt, A., Wanderley, M.M.: Mapping performer parameters to synthesis engines. Organised Sound 7(2), 97–108 (2002)
35. Jensenius, A.R.: Action-sound: Developing methods and tools to study music-related body movement. PhD thesis, University of Oslo, Department of Musicology, Oslo, Norway (2007)
36. Lee, E., Grüll, I., Kiel, H., Borchers, J.: Conga: a framework for adaptive conducting gesture analysis. In: NIME 2006: Proceedings of the 2006 Conference on New Interfaces for Musical Expression, pp. 260–265 (2006)
37. Lee, E., Wolf, M., Borchers, J.: Improving orchestral conducting systems in public spaces: examining the temporal characteristics and conceptual models of conducting gestures. In: CHI 2005: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 731–740 (2005)
38. Leman, M.: Embodied Music Cognition and Mediation Technology. Massachusetts Institute of Technology Press, Cambridge (2008)
39. Levitin, D., McAdams, S., Adams, R.: Control parameters for musical instruments: a foundation for new mappings of gesture to sound. Organised Sound 7(2), 171–189 (2002)

40. Maestre, E.: Modeling Instrumental Gestures: An Analysis/Synthesis Framework for Violin Bowing. PhD thesis, Universitat Pompeu Fabra (2009)
41. Minka, T.P.: From hidden markov models to linear dynamical systems. Technical report, Tech. Rep. 531, Vision and Modeling Group of Media Lab, MIT (1999)
42. Miranda, E., Wanderley, M.: New Digital Musical Instruments: Control And Interaction Beyond the Keyboard. A-R Editions, Inc (2006)
43. Mitra, S., Acharya, T., Member, S., Member, S.: Gesture recognition: A survey. IEEE Transactions on Systems, Man and Cybernetics - Part C 37, 311–324 (2007)
44. Momeni, A., Henry, C.: Dynamic independent mapping layers for concurrent control of audio and video synthesis. Computer Music Journal 30(1), 49–66 (2006)
45. Mori, A., Uchida, S., Kurazume, R., Ichiro Taniguchi, R., Hasegawa, T., Sakoe, H.: Early recognition and prediction of gestures. In: Proceedings of the International Conference on Pattern Recognition, vol. 3, pp. 560–563 (2006)
46. Myers, C.S., Rabiner, L.R.: A comparative study of several dynamic time-warping algorithms for connected word recognition. The Bell System Technical Journal 60(7), 1389–1409 (1981)
47. Rabiner, L.R.: A tutorial on hidden markov models and selected applications in speech recognition. Proceedings of the IEEE, 257–286 (1989)
48. Rajko, S., Qian, G.: A hybrid hmm/dpa adaptive gesture recognition method. In: International Symposium on Visual Computing (ISVC), pp. 227–234 (2005)
49. Rajko, S., Qian, G.: Hmm parameter reduction for practical gesture recognition. In: 8th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2008), pp. 1–6 (2008)
50. Rajko, S., Qian, G., Ingalls, T., James, J.: Real-time gesture recognition with minimal training requirements and on-line learning. In: CVPR 2007: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8 (2007)
51. Rank, E.: A player model for midi control of synthetic bowed strings. In: Diderot Forum on Mathematics and Music (1999)
52. Rasamimanana, N., Guedy, F., Schnell, N., Lambert, J.-P., Bevilacqua, F.: Three pedagogical scenarios using the sound and gesture lab. In: Proceedings of the 4th i-Maestro Workshop on Technology Enhanced Music Education (2008)
53. Rasamimanana, N.H., Bevilacqua, F.: Effort-based analysis of bowing movements: evidence of anticipation effects. The Journal of New Music Research 37(4), 339–351 (2008)
54. Rasamimanana, N.H., Kaiser, F., Bevilacqua, F.: Perspectives on gesture-sound relationships informed from acoustic instrument studies. Organised Sound 14(2), 208–216 (2009)
55. Roebel, A.: A new approach to transient processing in the phase vocoder. In: Proceedings of DAFx (September 2003)
56. Schnell, N., Borghesi, R., Schwarz, D., Bevilacqua, F., Müller, R.: Ftm - complex data structures for max. In: Proceedings of the International Computer Music Conference, ICMC (2005)
57. Schnell, N., Röbel, A., Schwarz, D., Peeters, G., Borghesi, R.: Mubu and friends: Assembling tools for content based real-time interactive audio processing in max/msp. In: Proceedings of the International Computer Music Conference, ICMC (2009)
58. Schnell, N., et al.: Gabor, Multi-Representation Real-Time Analysis/Synthesis. In: Proceedings of DAFx (September 2005)

59. Schwarz, D., Orio, N., Schnell, N.: Robust polyphonic midi score following with hidden markov models. In: Proceedings of the International Computer Music Conference, ICMC (2004)
60. Serafin, S., Burtner, M., Nichols, C., O'Modhrain, S.: Expressive controllers for bowed string physical models. In: DAFX Conference, pp. 6–9 (2001)
61. Turaga, P., Chellappa, R., Subrahmanian, V., Udrea, O.: Machine recognition of human activities: A survey. IEEE Transactions on Circuits and Systems for Video Technology 18(11), 1473–1488 (2008)
62. Van Nort, D., Wanderley, M., Depalle, P.: On the choice of mappings based on geometric properties. In: Proceedings of the International Conference on New Interfaces for Musical Expression, NIME (2004)
63. Van Nort, D.: Modular and Adaptive Control of Sonic Processes. PhD thesis, McGill University (2010)
64. Verfaille, V., Wanderley, M., Depalle, P.: Mapping strategies for gestural and adaptive control of digital audio effects. The Journal of New Music Research 35(1), 71–93 (2006)
65. Volpe, G.: Expressive gesture in performing arts and new media. Journal of New Music Research 34(1) (2005)
66. Wanderley, M., Depalle, P.: Gestural control of sound synthesis. Proceedings of the IEEE 92, 632–644 (2004)
67. Wanderley, M.: (guest ed.) Mapping strategies in real-time computer music. Organised Sound, vol. 7(02) (2002)
68. Wilson, A.D., Bobick, A.F.: Realtime online adaptive gesture recognition. In: Proceedings of the International Conference on Pattern Recognition (1999)