



# Designing action-sound metaphors using motion sensing and descriptor-based synthesis of recorded sound materials

Frédéric Bevilacqua, Norbert Schnell, Jules François, Eric O Boyer, Diemo Schwarz, Baptiste Caramiaux

## ► To cite this version:

Frédéric Bevilacqua, Norbert Schnell, Jules François, Eric O Boyer, Diemo Schwarz, et al.. Designing action-sound metaphors using motion sensing and descriptor-based synthesis of recorded sound materials. Routledge Companion to Embodied Music Interaction, 2017, 978-1-138-65740-3. hal-01572629

**HAL Id: hal-01572629**

**<https://hal.science/hal-01572629>**

Submitted on 23 Jul 2019

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

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

# **Designing Action–Sound Metaphors using Motion Sensing and Descriptor-based Synthesis of Recorded Sound Materials**

Frédéric Bevilacqua, Norbert Schnell, Jules François, Éric O. Boyer, Diemo Schwarz,  
Baptiste Caramiaux

STMS Ircam-UPMC-CNRS

1, place Igor Stravinsky,

75004 Paris, France

## **Abstract**

In this chapter, we present a general methodology for designing interactive musical systems, using movement sensing and descriptor-based synthesis of recorded sound materials. Importantly, the design principles focus on action–sound metaphors that can be built upon audio features of recorded sound materials and their possible relationships to human movement. We describe the critical design choices, including sensing technologies, movement analysis, and action–sound models, along with application examples.

## Introduction

Movement-based interactive musical systems have been developed in artistic communities since the beginning of electronic music. Several new types of applications have emerged over the last years, such as gaming, sound design, or even rehabilitation (Franinović and Serafin, 2013). In this context, theoretical frameworks such as embodied cognition lead us to reconsider design approaches for creating interactive systems.

We present a general methodology for designing musical interactive systems, using movement sensing and descriptor-based synthesis of recorded sound materials. Importantly, the design principles focus on action–sound metaphors that can be built upon features of recorded sound material and their possible relationships to human movement (Caramiaux et al., 2014a). Our primary argument, that we will develop through several examples, is that the technology-mediated interactions built on these principles allow for gesturally re-enacting audio materials through action–sound metaphors.

The interaction scenarios we are mostly interested in generally imply *continuous* movement and *continuous* sound feedback. The implementation of such action–sound metaphors can be difficult to put in practice using standard movement sensing and sound synthesis. Standard sampling-based synthesis, especially well suited for triggering actions, remains yet generally cumbersome for continuous control. To overcome these difficulties, we propose here to use granular/concatenative synthesis driven by audio descriptors, as implemented for example in the software library called MuBu (Schnell et al., 2009). A typical example of such a sound synthesis technique is corpus-based concatenative synthesis, which has been described in previous work (Schwarz, 2007).

The scope of this chapter is to describe the general method to create movement and bodily interaction using such sound synthesis techniques. These systems can be seen as the fusion of several lines of research we have been conducting at Ircam from fundamental research on movement and interaction to technical developments on motion sensing and sound synthesis.

We first introduce several key conceptual elements that motivate our approach. Second, we present the general method and technical architecture. Third, we describe how this synthesis method can be used for implementing movement–sound relationships in concrete exemplary cases. We end the chapter with a discussion and a proposition for future challenges.

## Action–Sound Metaphors

In previous work, we formalized the design principle of *Mapping through Listening* that considers listening as the foundation and the first step of the design of the relationships between motion and sound (Caramiaux et al., 2014a, 2015). This approach builds upon related work on listening modes and gestural sound descriptions to formalize different types of movement–sound mappings (note that similar approaches have been also described by [Maes et al., 2010] et [Jensenius 2007]). Building on this approach, we propose a general design method for creating movement–sound relationships in technology-mediated interactive systems. In this chapter, we explicitly include in the design step the full chain of technological elements, sound recording, movement sensing, and descriptor-based sound synthesis. We use the term *action–sound metaphor* to stress that the movement–sounds relationships we propose can be related or derived from existing situations (see Wessel et al 2002 for early attempt using such concepts). Our aim is to take advantages of known associations between movement and sound, acknowledging the importance of action-perception in interaction.

### *Action-Perception*

Our approach is motivated by a line of research in cognitive sciences and neuroscience that further supports the role of the body in sound perception and cognition. Theories of embodied cognition emphasize the essential role of the body in cognitive phenomena (Anderson, 2003; Lakoff & Johnson, 1999). O'Regan and Noë (2001) showed that perception should be considered as an active phenomenon, as sense organs are themselves dynamic instruments of exploration of an environment. The theory of embodied cognition has a significant impact on current trends in Human-Computer Interaction research (Kirsh, 2013), and interaction design (Dourish, 2004). In particular, Leman (2008) underlines the importance of “corporeal engagement” in music experience (Leman, 2008, p19).

These theories are supported by recent results in neuroscience. Studies on auditory–motor interactions in the perception and production of music have brought evidence of a strong coupling between action and perception (Haueisen and Knösche 2001, Zatorre et al. 2007; Lahav et al. 2005; Chen et al. 2008).

The notion of action–perception coupling is therefore central to the design of interactive systems where technology mediates our bodily interaction with sound.

## *Gestural Affordances of Sound*

Several authors have argued that sound affords movement, or can elicit its listeners' movements with particular characteristics. According to Godøy (2003), “we mentally imitate the sound-producing action when we attentively listen to music, or that we may image actively tracing or drawing the contours of the music as it unfolds”. This concept is supported by recent experiments investigating motor responses to sound stimuli (Godøy et al., 2006a,b; Caramiaux et al., 2010a, 2014b, Altavilla et al., 2013). Depending on the context, a wide range of strategies for associating gestures to sound are possible, such as mimicking the sound-producing actions or tracing the perceived properties of the sound (Godøy et al., 2006a, Caramiaux et al. 2014b). Recently, Caramiaux et al. (2014b) showed that the identification of the sound source can influence such gestural strategies.

## **Method and Technological Tools**

The general workflow of our applications is described in Figure 1. This workflow brings together several technological tools, which must be chosen and implemented carefully to guarantee the consistency of a chosen action–sound metaphor.

### Selecting or Recording of Sound Materials

A first element in the design of interactions consists in deriving perceived or imagined actions from the sound itself. Sound recordings are chosen or created purposely to contain perceivable features that will be used in the interactions. In addition the recordings may be decomposed into sequences of sound segments that correspond to specific sound events.

### Selecting Movement and Interfaces

The second part of designing action–sound metaphors relies on defining the users movements and actions. These can include for example free movements performed in the air, the manipulation of objects, touching surfaces, as well as vocal input. In the case of using objects or tangible interfaces, the design must embrace their inherent affordances and metaphorical references.

### Dimensionality Reduction

Audio descriptors extracted from the recorded sound materials (Peeters et al., 2004) and raw sensor data describing the user's movements or vocal articulations, are generally represented by streams of scalars or vectors in similar rates (generally from 40 to 200 Hz). At this point, is

often necessary to further reduce the audio and motion data dimensionality to establish meaningful relationships between movement and sound descriptors. This can be performed either manually by selecting and combining several components (e.g. computing vector norm from  $x$ ,  $y$ ,  $z$  components), or using statistical methods such as supervised or unsupervised learning methods (e.g. principal component analysis or canonical component analysis, see Caramiaux et al., 2010a).

The resulting streams of so-called high-level descriptors should ideally correspond to features that could be associated with perceivable sound and movement features. The choice of analysis and data reduction techniques are thus critical in the design insofar as it creates particular “points of view” which allow for establishing the desired metaphors.

## Action–Sound Models

To implement action–sound metaphors we propose three categories of models, *congruent models*, *physical behavior models*, and *adaptive models* (see Figure 1).

### Congruent model

In some cases, the action–sound metaphor can be solely defined by *shared* features that describe both the users' movements and the recorded sound materials, constituting what we define as a *congruent model*. These features are generally built by an adequate analysis and data reduction. In this case, the action–sound metaphor can be implemented as a direct mapping between movement and sound parameters, using these shared features. For instance, in the *Cocktail Shaker* example application presented below, the movement and sound descriptions can be reduced to the an *intensity* feature, which *congruently* applies to the intensity of the users' shaking movement and to the dynamics of the recorded sounds.

### Physical Behavior models

Many action–sound metaphors rely on complex behaviors, involving several physical mechanisms and rules that mediate the relationship between the users' movements and the resulting sound. In some cases, these *physical behaviors* can be implemented by explicit rules and constraints or by using physical models. As described later in the rainstick example, a physical model can mediate the relationship between the inclination of the stick and the sound resulting from the motion of a multitude of virtual small particles inside the stick. More generally, we also consider that constraints and behavior imposed by real physical objects (such as tangible interfaces) can be considered as part of physical behavior models.

### Adaptive models

Apart from *congruent models* and *physical behavior models*, we propose a third category of implementations that aims to automatically adapt a probabilistic model to a sought relationship between movement and sound parameters. Here, the model can be trained through interactive machine learning (or manually adjusted) to create arbitrary action–sound metaphors. Such machine learning techniques can also account for automatically reducing the data dimensionality by finding inherent correlations in the evolution of movement sound descriptors.

## *Technological Components*

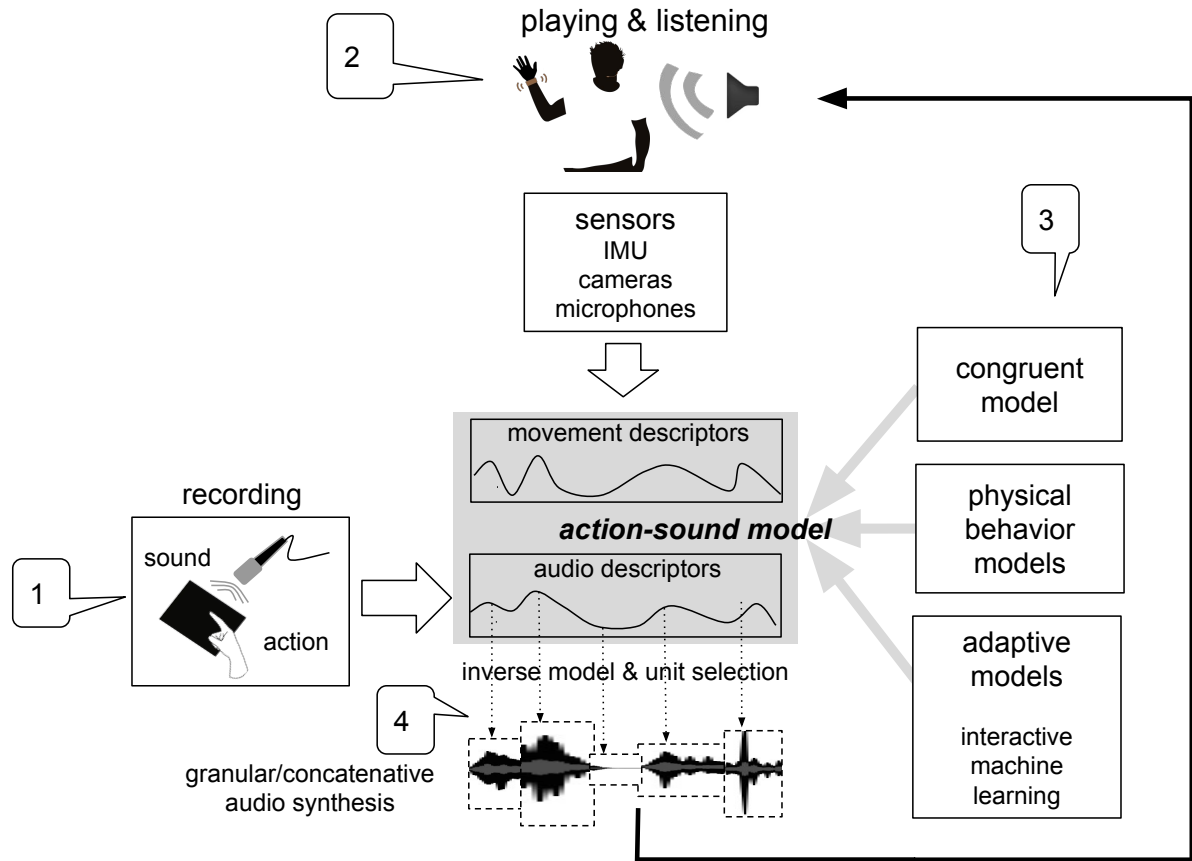
### Motion Sensing

User action and movements are transformed into input data streams directly using various motion capture systems or indirectly using microphone techniques (Miranda and Wanderley, 2006). The applications described in this chapter use either 3D camera systems (such as the *Leap Motion*), or wireless inertial measurement units such as the MO (Modular Musical Objects) (Rasamimanana et al., 2011; Schnell et al., 2013; Bevilacqua et al., 2013). Some of our applications also use the inertial measurement units integrated in mobile phones.

### Audio Analysis and Synthesis

The applications described in this chapter are implemented using the *Max (Cycling '74)* programming and audio–visual processing environment and different audio processing libraries that we have developed. All the applications make use of the MuBu library (Schnell et al., 2009) and some use FTM & Co (Schnell et al., 2005; Schnell and Schwarz, 2005). These libraries provide data structures, signal processing functionalities and visualization tools. The recorded sound materials can be analyzed to extract audio descriptors (Peeters et al., 2004). Most of the functionalities of these libraries have been developed to support the analysis and resynthesis of sound and motion.

Our method is based on descriptor-based sound synthesis — or *corpus-based concatenative synthesis* (Schwarz, 2007), — where a stream of audio descriptors is used to query and play the associated sound grains or segments in a corpus of sounds. Precisely, this is implemented using a *kD-Tree* search in the descriptor space (Schwarz et al., 2009). This method can be implemented with either granular or concatenative sound synthesis.



**Figure 1: General workflow of the design of movement-based interactive applications with recorded sound materials. (1) Selection and recording of sound materials, segmentation and analysis (i.e. extraction of audio descriptors). (2) Capture of user movements, and computation of movement descriptors. (3) The action-sound model produces the evolution of sound descriptors from the incoming movement descriptors. (4) Sound descriptors control granular or concatenative sound synthesis through an inverse model (i.e. selecting sound segments that fit a given description).**

## Example Applications: Interacting with Water Sounds

We present in this section a series of applications that were designed following this framework. While developed in different application contexts, we chose for the sake of comparison a series of applications making use of water sound. These examples illustrate how different action-sound metaphors can be designed starting from a given sound source, using various types of motion sensing and interfaces. We grouped them to follow the different categories of action-sound models between movement and audio descriptors.

### *Examples Based on Congruent Models*

For the first two examples, the perceived evolution of sound features is directly induced by the users' movements. The first example uses inertial measurement units to capture the users'



shaking intensity that is directly used to query sound segments with a *congruent* sound intensity. In the second example, vocal articulations are captured as sound and described by timbral coefficients. These coefficients are used to drive the synthesis of “speaking water textures” which exhibits similar articulations.

## Cocktail Shaker

The shaking metaphor of this example directly relies on the common description of the user's movement and the water sound textures by a shared *intensity* parameter. The movement is captured through an inertial measurement unit that the users hold in their hand. The shaker *intensity* is computed as the low-pass filtered power of the acceleration. The sound materials are segmented and each segment is described by its maximum power (calculated using of a gliding window of 20 to 50 milliseconds) that represent the sound *intensity*. The mapping between these two intensities is obtained by normalizing both intensities to a common range (see also Tanaka et al., 2013). Note that this mapping is to select each sound grain separately, which are then rendered using granular/concatenative synthesis. The sonic result is perceptually very different than if only the global audio volume was modified.

The metaphor of this application can simply be performed through the mimicry of the movements of shaking a cocktail, or further emphasized by embedding the sensor into an actual cocktail shaker.

## Speaking Water

Similar to the cocktail shaker in the previous example, the metaphor of this example is composed by the conjunction of speech articulation and water textures. The implementation is based on a technique known as audio mosaicking, which consist of recomposing an audio stream from frames that follow the evolution of intensity and timbre of another audio stream (Lazier and Cook, 2003, see also [http://imtr.ircam.fr/imtr/Real-Time\\_Audio\\_Mosaicing](http://imtr.ircam.fr/imtr/Real-Time_Audio_Mosaicing), Schnell, N. (2011).). It uses Mel Frequency Cepstral Coefficients (MFCCs) to describe both the speech input captured by a microphone and the recorded sound materials.

The continuous query of audio frames from the recorded water sounds is directly driven by the MFCCs extracted in real-time from the speech input. A linear scaling is derived from example sounds to match the descriptor spaces of speech and water sounds. As in the previous example, these scaled parameters constitute a shared congruent description, applying to both input and output sound.

An interesting variant of this application can be obtained by replacing the speech input by sounds captured from movements performed on a rough surface. The action metaphor here changes from *speaking* to *touching*.

### *Examples Based on Physical Behavior Models*

#### Rainstick

The metaphor here is directly inspired by the traditional rainstick, which creates a homogeneous sound texture of intense rain when a long object is strongly tilted. A “digital” rainstick can be easily implemented using an object, the tilting angle of which is measured using an inertial measurement unit (This case previously described as the Grainstick, developed in collaboration with Pierre Jodlowski (Leslie et al., 2010)). To reproduce the rainstick sonic behavior, it is necessary to simulate the dynamics of the grains falling in the tube: the sound intensity and granularity reacts to the tilting movement with a delayed response, lasting longer than the movement itself. Such a dynamical behavior can be produced using a physical model: temporal envelopes representing the grains density and velocities, resulting from tilting, are used to drive the density and intensity of recorded sound droplets. This represents an example where a temporal model is necessary to properly implement dynamical behavior inspired from the interaction of physical objects

#### Tangible Waterscape

In this application the action–sound metaphor relies on a tangible interface. The users interact with a granular material such as tapioca beads in a flat glass dish using their hands. Moving the different beads creates various water sounds, in reaction to the hand movement's energy and spatial location. The aim is not to simulate a realistic interaction with a water bowl, but to give access, through a tangible interaction, to a large variety of metaphorical interactions related to water, from single drops to waterfalls and ocean waves.

This application uses the interface called DIRTi (see Savary et al., 2013) developed for tangible interaction with a larger corpus of sounds. The beads are tracked by a webcam from underneath, which allows for measuring the relief (i.e. thickness of the beads layer). The segments of the recorded sound materials are associated to a grid of uniformly distributed positions on the 2D interaction surface of the bowl by choosing two audio descriptors as its axes (Lallemand and Schwarz, 2011). This way, each sound segment has a fixed position on

the surface according to its sonic characteristics, e.g. brilliant splashy sounds are to the top-right, and muffled gurgly sounds to the bottom left, and sounds in between will present a smooth progression in sound characteristics, with similar sounds clustering together. The amount of movement at the position of a sound segment is mapped to its playback volume, so that fast movements play loudly, slow movements play softly.

Complementary to the auditory feedback, the tangible interaction can include the full surface of the hands, giving tactile feedback through the physical properties of the interaction material. The distribution of sound timbre at different locations of the bowl stimulates spatial movement explorations. The dynamic response of the system also stimulates the exploration between soft playing to huge washes and waves of sound made with the full hand(s).

### Virtual Water Tank

This example implements another physically-inspired metaphor where users play with their hands in water. This was developed based on the concept of “Audio Virtual Surface” (Boyer et al., 2015) that was developed for sensory substitution applications. In contrast with the previous case, no tangible interface is used here: the same sound material is played by interacting with a virtual water volume. The movement-sound relationship is built using physics-inspired behaviors.

The action–sound metaphor is directly derived from the sound recording of splashing actions at a water surface, from tiny ones, performed with the fingers, to large ones which imply larger and more energetic movements. These movements are detected using a stereo camera (*Leap Motion*) that estimates the hand and fingers positions and velocities in the air. When the hand crosses the virtual tank water surface, the hand and fingers velocities are used to select different sound segments based on their description. First, the velocity is used to select the length of sound segment. Therefore, energetic movements drive longer sounds, giving the sonic impression that the water surface takes longer to return to a stable state. Second, velocity also modifies the audio spectral features (centroid, spread) of the sound segment. Finally, we also added another interaction mode that occurs when the hands are below the surface level. In this case, lateral waving movements control water sound with an additional audio effect (low-pass filtering) simulating sonic action underwater.

While these different interactions are performed in the air without any tangible interface, the very familiar types of sound that responds to the movements (both the clear splashing at the

surface and the muffled underwater sound) can create the illusion of the water volume (Boyer 2014).

### *Example Based on Adaptive Models*

#### DIY: Define the Interaction Yourself

In the previous examples, the relationships between movement and sound are fixed, and the users must adapt their movements to the system. An alternative approach consists in letting users design these movement-sound relationships themselves. In recent work, we developed an approach called *Mapping by Demonstration* that learns a continuous mapping between motion and sound parameters from user examples (Françoise 2015). This approach relies on Interactive Machine Learning techniques (Fails and Olsen 2003; Fiebrink et al. 2011) where we make use of probabilistic models as a user-adaptable tool for designing movement and sound relationships (Bevilacqua et al. 2011, Françoise et al. 2014a). This scenario gives users the opportunity to craft the system according to their own metaphors, or to the actions evoked by their personal listening experience of the sound recordings. Moreover, users can record different gestures for different sounds, that can be mixed, sequenced or composed using recognition and bimanual interaction. This leaves the opportunity to develop an entire sonic environment controlled with a personal vocabulary of actions that can evolve over time and training.

In this scenario, that we built as a public installation (Françoise et al., 2014b; Françoise, 2015), users can design the gestures associated with pre-designed sounds. We use the same recorded water sounds than the ones used in the Tangible Waterscape example. User's hand movements are captured using a stereo-camera (Leap Motion), from which we extract the speed and orientation of the hand as well as custom features of 'shakiness' and hand opening. For each sound, users can record their own gesture (Françoise et al. 2014b), and the sound textures are finally recomposed using descriptor-driven concatenative synthesis.

### **Discussion and Conclusion**

We presented a general method to create interactive audio applications based on action-sound metaphors. The metaphors are implemented by action-sound models which establish

relationships between movement and audio features. We would like to insist that the methodology we presented should not be understood as a method to *control* sound synthesis, but as the design of action–sound metaphors. The presented examples using the same recorded sound materials illustrate how action–sound metaphors may be built on a large variety of different movements and gestures. As described below, the key components of the system — the sound materials, the movements, action–sound model — must be coherently chosen to make a given metaphor effective.

First, the movement interfaces and the possible affordances must be carefully examined and made consistent with the sought action–sound metaphor. Moreover, a choice between free gestures and tangible interfaces has important implications on aspects of effort, perception, and emotion (Tajadura-Jimenez 2015), as shown in the previous examples of applications between *Tangible Waterscape* and *Virtual Water Tank*.

Second, the selection of the sound recordings and their description are essential design choices. In particular, the set of audio recordings should be chosen to afford, after segmentation, a set of possible actions. Similarly, different audio descriptors provide different perspectives on the recording material. Empirically, we found that in many cases the use of a limited number of audio descriptors is sufficient to represent the perceived feature variations. For example, for a sound obtained from scratching a surface, the loudness descriptor correlates with timbral variations (i.e. sounds of higher energy have a more brilliant timbre). Thus, the selection of sound segments by the loudness provides the user with a sound palette that also contains timbral variations that are coherent with the scratching action.

Another essential design choice resides in the action–sound model, which defines the overall behavior of the system. Action–sound metaphor can rely on different categories of models that we propose here: *congruent models*, *physical behavior models* and *adaptive models*. Each of them refers to specific physical behavior models that are helpful to implement the metaphors. As these different models generate different temporal responses to the user input, they might affect differently the agency perception.

As previously said, the movement interfaces, sound material and descriptors must be designed to be coherent with the metaphor. Nevertheless, it should remain clear that, as illustrated with our examples with water sound, several solutions could be valid. Several criteria could serve to compare the different solutions depending on the application contexts. These criteria can span from sensori-motor learning for rehabilitation applications (Boyer, 2015), stimulating in

pedagogical context, or to favor social engagement and playfulness in gaming/entertainment. Notions such as animation or re-enactment has also been proposed by Schnell (2014).

Finally, we would like to end this discussion by presenting several research challenges for embodied interaction.

1. *Building on existing motor skills*

One of our primary design strategies is the re-use of acquired motor skills in novel contexts. By repurposing existing motor skills for interaction with sound, we hope to strengthen the player's action-perception loop. We are currently investigating whether such extrinsic auditory feedback using metaphors of physical behaviors can support sensori-motor learning (Boyer, 2015). In addition, our systems allow for a broader exploration of the motor theory of — sound — perception. While their design is based on a set of actions afforded by the initial recordings, their reinterpretation through bodily interaction necessarily changes the player's perception.

2. *Building on musical knowledge*

The presented implementations of our approach have mostly focused on applications in sonification, gaming, and installations. While such systems could be used as instruments in music performance, our approach would benefit from the integration of additional musical knowledge, such as harmonic and temporal structures derived from existing music styles.

3. *Building on social Interactions*

Finally, social aspects are an important factor of our applications since generally they are discovered, learned and played collectively. We are currently expanding our framework and applications to large group interactions, which remains to be fully investigated both theoretically and technologically.

## **Acknowledgments**

This research has been conducted in the {Sound Music Movement} Interaction team at Ircam. We wish to acknowledge our team collaborators Riccardo Borghesi, Emmanuel Fléty, Joseph Larralde, Sébastien Robaszkiewicz, Benjamin Matuszewski, and external collaborators Roland Cahen, Matthieu Savary, Denis Pellerin, Pierre Jodlowski. This research was supported by the European projects Rapid-Mix (H2020-ICT-2014-1 Project ID 644862), MusicBricks (H2020-ICT-2014-1 Project ID 644871), SkAT-VG (FET-Open grant number: 618067); and by the ANR projects Topophonie (ANR-09-CORD-022), LEGOS (ANR-11-BS02-012), and CoSiMa (ANR-13-CORD-0010), and received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 659232.

## References

- Altavilla, A., Caramiaux, B., & Tanaka, A. (2013). Towards Gestural Sonic Affordances. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME'13)*.
- Anderson, M. (2003). Embodied Cognition: A field guide. *Artificial Intelligence*, 149(1), 91–130.
- Bevilacqua, F., Schnell, N., Rasamimanana, N., Bloit, J., Fléty, E., Caramiaux, B., ... Boyer, E. (2013). De-MO : Designing Action-Sound Relationships with the MO Interfaces. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems*. Paris, France.
- Bevilacqua, F., Schnell, N., Rasamimanana, N., Zamborlin, B., & Guédy, F. (2011). Online Gesture Analysis and Control of Audio Processing. In *Musical Robots and Interactive Multimodal Systems* (pp. 127–142). Springer.
- Boyer, E. O. (2015). *Continuous auditory feedback for sensorimotor learning*. Université Pierre et Marie Curie.
- Boyer, E. O., Pyanet, Q., Hanneton, S., & Bevilacqua, F. (2014). Learning Movement Kinematics with a Targeted Sound. In *Sound, Music & Motion, Lecture Notes in Computer Science Volume 8905* (pp. 218–233). Springer International Publishing.
- Boyer, E. O., Vandervoorde, L., Bevilacqua, F., & Hanneton, S. (2015). Touching Sounds: Audio Virtual Surfaces. In *SIVE 2015, IEEE 2nd VR Workshop on Sonic Interactions for Virtual Environments*. Arles, France.
- Caramiaux, B., Bevilacqua, F., Bianco, T., Schnell, N., Houix, O., & Susini, P. (2014). The Role of Sound Source Perception in Gestural Sound Description. *ACM Transactions on Applied Perception*, 11(1), 1–19.
- Caramiaux, B., Altavilla, A., Pobiner, S. G., & Tanaka, A. (2015). Form Follows Sound: Designing Interactions from Sonic Memories. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 3943–3952). ACM.
- Caramiaux, B., Bevilacqua, F., & Schnell, N. (2010). Towards a gesture-sound cross-modal analysis. *Gesture in Embodied Communication and Human-Computer Interaction*, 5934, 158–170.
- Caramiaux, B., Schnell, N., Françoise, J., & Bevilacqua, F. (2014). Mapping Through Listening. *Computer Music Journal*, 38(34-48).
- Chen, J. L., Penhune, V. B., & Zatorre, R. J. (2008). Listening to musical rhythms recruits motor regions of the brain. *Cerebral Cortex*, 18(12), 2844–2854.
- Dourish, P. (2004). *Where the action is: the foundations of embodied interaction*. The MIT Press.



- Fails, J. A., & Olsen, D. R. (2003). Interactive machine learning. In *Proceedings of the 8th international conference on Intelligent user interfaces* (pp. 39–45).
- Fiebrink, R., Cook, P. R., & Trueman, D. (2011). Human model evaluation in interactive supervised learning. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (p. 147). Vancouver, BC, Canada: ACM.
- Françoise, J. (2015). *Motion-Sound Mapping by Demonstration*. Université Pierre et Marie Curie.
- Françoise, J., Schnell, N., & Bevilacqua, F. (2014). MaD: Mapping by Demonstration for Continuous Sonification. In *ACM SIGGRAPH 2014 Emerging Technologies* (p. 16). Vancouver, Canada: ACM.
- Françoise, J., Schnell, N., Borghesi, R., & Bevilacqua, F. (2014). Probabilistic Models for Designing Motion and Sound Relationships. In *Proceedings of the 2014 International Conference on New Interfaces for Musical Expression* (pp. 287–292). London, UK.
- Franinović, K., & Serafin, S. (2013). *Sonic Interaction Design*. MIT Press.
- Godøy, R. I. (2003). Motor-mimetic music cognition. *Leonardo*, 36(4), 317–319.
- Godøy, R. I., Haga, E., & Jensenius, A. (2006). Playing “ Air Instruments ”: Mimicry of Sound-producing Gestures by Novices and Experts. *Gesture in Human-Computer Interaction and Simulation*, 256–267.
- Godøy, R. I., Haga, E., & Jensenius, A. R. (2006). Exploring music-related gestures by sound-tracing - a preliminary study. In *2nd ConGAS International Symposium on Gesture Interfaces for Multimedia Systems* (pp. 9–10).
- Goudard, R., Genevois, H., Ghomi, É., & Doval, B. (2011). Dynamic Intermediate Models for audiographic synthesis. In *Proceedings of the Sound and Music Computing Conference*.
- Haueisen, J., & Knösche, T. R. (2001). Involuntary Motor Activity in Pianists Evoked by Music Perception. *Journal of Cognitive Neuroscience*, 13(6), 786–792.
- Jensenius, A. R. (2007). *Action-sound: Developing methods and tools to study music-related body movement*. University of Oslo.
- Kirsh, D. (2013). Embodied cognition and the magical future of interaction design. *ACM Transactions on Computer-Human Interaction*, 20(111), 3:1–3:30.
- Lahav, A., Boulanger, A., Schlaug, G., & Saltzman, E. (2005). The Power of Listening: Auditory-Motor Interactions in Musical Training. *Annals of the New York Academy of Sciences*, 1060(1), 189–194.
- Lakoff, G., & Johnson, M. (1999). *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*. Basic Books.

- Lallemand, I., & Schwarz, D. (2011). Interaction-optimized sound database representation. In *Conference on Digital Audio Effects (DAFx)*.
- Lazier, A., & Cook, P. (2003). MOSIEVIUS: Feature driven interactive audio mosaicing. In *Digital Audio Effects (DAFx)*.
- Leman, M. (2008). *Embodied Music Cognition and mediation technology*. The MIT Press.
- Leslie, G., Zamborlin, B., Jodlowski, P., & Schnell, N. (2010). Grainstick: A collaborative, interactive sound installation. In *Proceedings of the International Computer Music Conference (ICMC)*.
- Maes, P. J., Leman, M., Lesaffre, M., Demey, M., & Moelants, D. (2010). From expressive gesture to sound. *Journal on Multimodal User Interfaces*, 3(1), 67–78.
- Miranda, E., & Wanderley, M. (2006). *New digital musical instruments: control and interaction beyond the keyboard*. AR Editions, Inc.
- O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *The Behavioral and Brain Sciences*, 24, 939–973; discussion 973–1031.
- Peeters, G. (2004). *A large set of audio features for sound description (similarity and classification)*.
- Rasamimanana, N., Bevilacqua, F., Schnell, N., Fléty, E., & Zamborlin, B. (2011). Modular Musical Objects Towards Embodied Control Of Digital Music Real Time Musical Interactions. In *Proceedings of the fifth international conference on Tangible, embedded, and embodied interaction* (pp. 9–12). Funchal, Portugal.
- Savary, M., Schwarz, D., Pellerin, D., Massin, F., Jacquemin, C., & Cahen, R. (2013). Dirty tangible interfaces: expressive control of computers with true grit. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems on* (p. 2991). Paris, France: ACM Press.
- Schnell, N., Borghesi, R., Schwarz, D., Bevilacqua, F., & Muller, R. (2005). FTM - Complex data structures for Max. In *Proceedings of the International Computer Music Conference (ICMC)*. Barcelona, Spain.
- Schnell, N., & Schwarz, D. (2005). Gabor, Multi-Representation Real-Time Analysis/Synthesis. In *DAFx*.
- Schnell, N. (2013). *Playing (with) Sound - Of the Animation of Digitized Sounds and their Reenactment by Playful Scenarios in the Design of Interactive Audio Applications*. Institute of Electronic Music and Acoustics, University of Music and Performing Arts Graz.
- Schnell, N. (2011). Real-Time Audio Mosaicing. Retrieved June 20, 2006, from [http://imtr.ircam.fr/imtr/Real-Time\\_Audio\\_Mosaicing](http://imtr.ircam.fr/imtr/Real-Time_Audio_Mosaicing)

- Schnell, N., Röbel, A., Schwarz, D., Peeters, G., & Borghesi, R. (2009). Mubu & friends - assembling tools for content based real-time interactive audio processing in max/msp. In *Proceedings of International Computer Music Conference*. Montreal.
- Schwarz, D. (2007). Corpus-based concatenative synthesis. *IEEE Signal Processing Magazine*, 24(2), 92–104.
- Schwarz, D., Cahen, R., & Britton, S. (2008). Principles and applications of interactive corpus-based concatenative synthesis. *Journ{ées}es d'Informatique Musicale (JIM)*, GMEA, Albi, France.
- Schwarz, D., Schnell, N., & Gulluni, S. (2009). Scalability in Content-Based Navigation of Sound Databases. In *Proceedings of ICMC*.
- Tajadura-Jimenez, A., Bianchi-Berthouze, N., Furfaro, E., & Bevilacqua, F. (2015). Sonification of Surface Tapping Changes Behavior, Surface Perception, and Emotion. *MultiMedia, IEEE*, 22(1), 48–57.
- Tanaka, A., Caramiaux, B., & Schnell, N. (2013). MubuFunkScatShare: Gestural Energy and Shared Interactive Music. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems* (pp. 2999–3002). Paris, France: ACM.
- Wessel, D., & Wright, M. (2002). Problems and prospects for intimate musical control of computers. *The Computer Music Journal*, 26(3), 11–22.
- Zatorre, R. J., Chen, J. L., & Penhune, V. B. (2007). When the brain plays music: auditory--motor interactions in music perception and production. *Nature Reviews Neuroscience*, 8(7), 547–558.