BODY MOVEMENT IN MUSIC INFORMATION RETRIEVAL

Rolf Inge Godøy
FourMs
Department of Musicology
University of Oslo
r.i.godoy@imv.uio.no

Alexander Refsum Jensenius
FourMs
Department of Musicology
University of Oslo
a.r.jensenius@imv.uio.no

ABSTRACT

We can see many and strong links between music and human body movement in musical performance, in dance, and in the variety of movements that people make in listening situations. There is evidence that sensations of human body movement are integral to music as such, and that sensations of movement are efficient carriers of information about style, genre, expression, and emotions. The challenge now in MIR is to develop means for the extraction and representation of movement-inducing cues from musical sound, as well as to develop possibilities for using body movement as input to search and navigation interfaces in MIR.

1. INTRODUCTION

There are strong links between music and body movement: Performers produce sound through movements, and listeners very often move to music, as can be seen in dance and innumerable everyday listening situations. The links between music and body movement have been discussed since antiquity, but it is mostly in the last decade that we have seen more systematic research efforts on this topic within fields such as music technology, music performance, and music cognition [1–3]. Despite this rapidly growing research in various music-related fields, the idea of body movement as an integral and ubiquitous part of both performance and perception of music seems so far not to have had many consequences for music analysis, music theory, and music information retrieval. Based on a quick survey of papers from recent ISMIR conferences as well as on the overview in [4], the papers that directly or indirectly are concerned with body movement seem limited to a few on query by humming and tapping, as well as some on beat tracking and tempo induction. Also, a cross-check on Google Scholar showed that out of 4670 hits on MIR, 3730 included "audio", 1990 "MIDI", while only 21 included "body movement". [1] It seems fair then to conclude that body movement has not been an important topic in MIR contexts.

Based on our own and various international colleagues' work of the past decade, we believe that body movement is not just something that incidentally co-occurs with music, but that body movement is integral to music as a phenomenon. We would go so far as to claim that our experience of music is based on the combination of sound and movement sensations, hence that music is a fundamentally embodied phenomenon [5, 6]. With such an understanding of music, it also becomes clear that sensations of music-related body movements are in fact highly salient features of music, and should be considered alongside various sonic features, e.g. pitch, melody, harmony, and timbre. Exploring music-related body movement then becomes an urgent task also in relation to MIR, and in this paper we shall try to give an overview of the kinds of body movement that could be of interest in MIR and how they can be studied. Finally, we shall present some suggestions for how body movements could be used in interfaces for the search and retrieval of music information.

2. MUSIC-RELATED MOVEMENT

It seems that listeners associate different kinds of body movement with the music they hear, or merely imagine. Here it can be useful to start by making the general distinction between *sound-producing* and *sound-accompanying* movements. Although this distinction may not always be so clear-cut, sound-producing movements are those that contribute to the production of musical sound, and sound-accompanying movements are those that are made in response to the sound being heard [3].

Sound-producing movements may further be divided into *excitatory* movements such as hitting, bowing, blowing, and *modulatory* movements such as those for making a vibrato or various timbral nuances. Associated with sound-producing movements we also have various types of *sound-facilitating*, *expressive*, and *communicative* movements, meaning movements that are not strictly speaking sound-producing but still play an important role in music performance. Sound-accompanying movements, on the other hand, are all kinds of movements that people may make to music such as in

[1] Search conducted 21 April 2009 using Google Scholar in English, and with a syntax of "Music Information Retrieval" + "...".

dancing, marching/walking, swaying, and gesticulating.

In practice, we may often see these different movement types occur together: it is possible to make movements that partly reflect the sound-production, partly are more independent of the sound-production, e.g. mimicking a solo drum passage with the hands at the same time as swaying the whole body to the meter of the music. We may also see performers making movements that are partly necessary for producing sound, and partly more theatrical for the benefit of the audience, e.g. lifting the hand high up before striking a chord on a guitar. This means that music-related movements may be multi-functional in that they serve several different purposes at the same time.

We believe that musical sound itself also conveys salient movement images that are related to listeners' sensations of effort (tense, relaxed, fast, slow, etc.) as well as to kinematics or geometry of musical instruments (register, up/down, position, posture, etc. in relation to instruments). Studies of so-called 'air-instrument' performance such as 'air guitar', 'air drums', and 'air piano' suggest that even listeners with little or no formal musical training are able to have images of sound-producing movements that reproduce both the effort and the kinematics of the imagined sound-production actions, i.e. they manage to follow the spatiotemporal unfolding of instrumental performance quite well as if they were actually playing the music themselves [7].

As for various kinds of sound-accompanying movement afforded by musical sound, a study of 'free dance' to music [2] shows that professional dancers tend to agree when it comes to the sensation of effort or energy in dance movements, although there are variations in the kinematics (geometry) of the movements [8, 9]. Furthermore, studies of 'sound-tracing' show that listeners with variable levels of musical training (ranging from none to professional level training) also seem to spontaneously associate various shapes with the musical sound that they hear [10]. In these studies, listeners were asked to draw on a digital tablet the shape they associated with a sound fragment immediately after they had heard the fragment. Figure 1 shows the sound-tracings of 9 participants to a sound taken from the contemporary western music repertoire. This sound consists of a high-pitched attack on a triangle, followed by a downward glissando on strings, and ending up with a drum roll [11]. The excerpt is rather unconventional with regards to melodic, harmonic, and timbral features, but as we can see from the images of the sound-tracings, there still seems to be some level of consensus between the nine listeners as to the movement shape that was afforded by the sound.

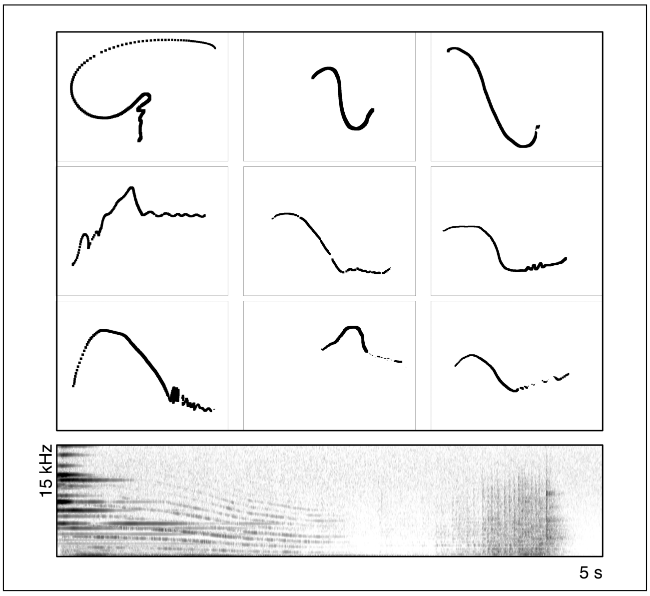

Figure 1. Sound-tracings by nine listeners of the sound fragment built up of an initial triangle attack, a downward glide in the strings and a final drum roll (spectrogram at the bottom) [11].

3. GLOBAL-LOCAL

It does not seem farfetched to suggest that listeners' music-related movements often match well the overall motion and emotion features of the musical sound, e.g. calm music tends to induce calm movements, agitated music tends to induce agitated movements, accentuated music tends to induce jerky movements, etc. The details of the movements may vary, however, something that may be seen both from qualitative annotations [8], as well as from quantitative data. An example of the latter may be seen in how the*quantity of motion* seems to correlate quite well with the dynamics of the waveform of the sound [7]. Similarly, *motiongrams* [3] are useful for displaying movement from video material. Figure 2 shows an example of how a motiongram of the hand movements of a pianist can be used together with the spectrogram of the resultant sound to study relationships between movement features and sonic features in a 20 seconds excerpt from the last movement of Beethoven's Tempest Sonata.

Visual representations such as motiongrams and spectrograms make it possible to move between global and more local perspectives, i.e. facilitates the correlation of music-related movement at different timescales with corresponding sonic features at different timescales. Here it could be useful to identify three different timescale levels when studying sound and movement in music:

Sub-chunk level: the level of perceiving continuous sound (pitch, timbre, and intensity) and movement (location, force, etc.).

Chunk level: sound fragments and actions that are perceived holistically and that may allow for the perception of rhythmical, textural, and melodic patterns, as well as tonal/modal and harmonic features, and importantly, also expressive features.

[2] The only instruction given was to make spontaneous movement to the musical excerpts upon first hearing.

[3] A motiongram is a visual representation of movement in a video, created by spatially reducing frame-differenced video images, see [9] for details

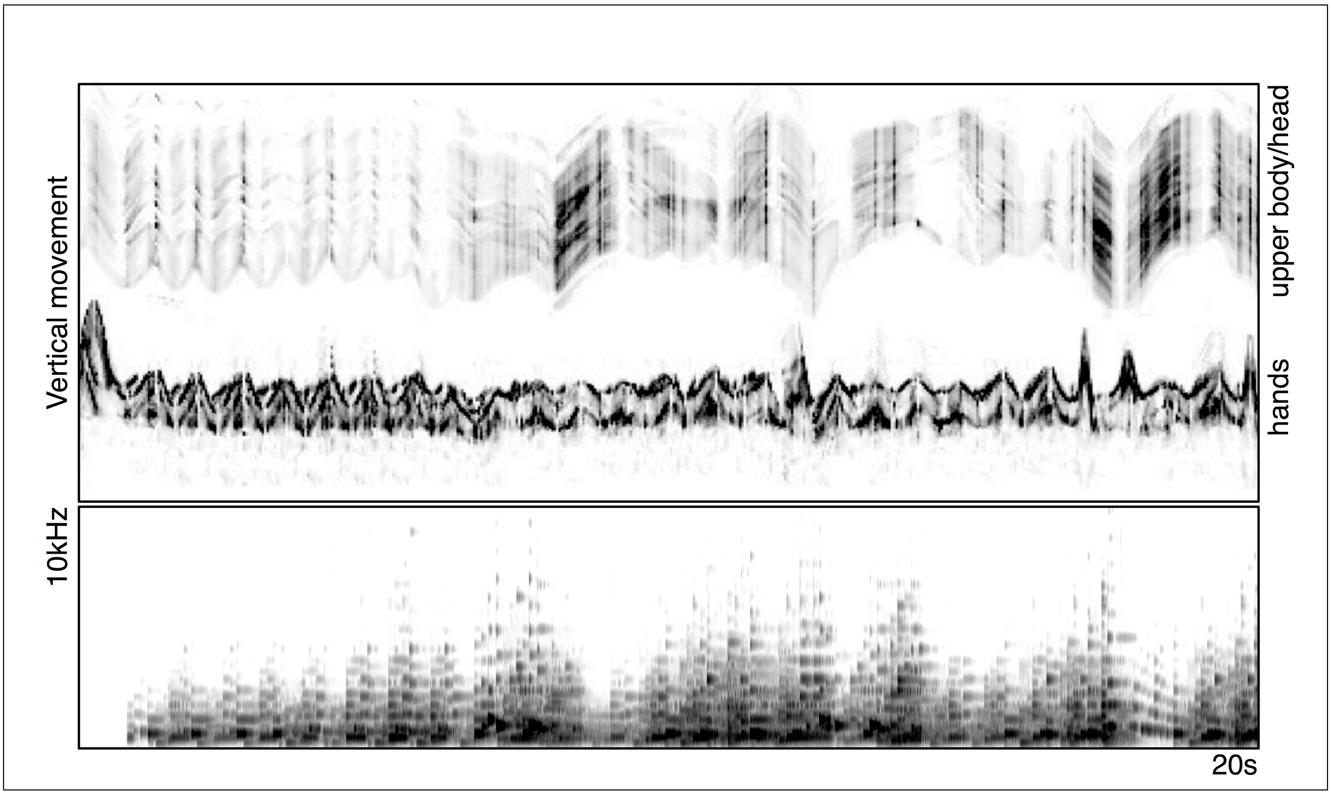

Figure 2. Motiongram of hand movement (top) and spectrogram (bottom) of the corresponding sound in a 20 seconds excerpt (first 30 measures) from the last movement of Beethoven's *Tempest Sonata* performed by François-René Duchable [12]. Notice the correlation between hand movements and the sound, as well as the sway in the upper body.

Supra-chunk level: several chunks are concatenated into larger-scale entities such as whole sections, tunes, movements, and even whole works.

We believe that the chunk-level, in the range of approximately 0.5 to 5 seconds, may be seen as the most important for identification of musical style, mode of performance, as well as emotive features. As suggested by Pierre Schaeffer's work on sonic objects several decades ago [13,14] and recently by work on more traditional western music [15], the chunk level seems to be more important than larger scale levels in music. Interestingly, and probably not accidentally, the temporal size of basic action units fits well with that of sonic objects, as well as with various other constraints on attention and memory, see [16] for a summary.

From what emerges of the sound-movement correspondences mentioned above, we think it is plausible to think of *gestural-sonic objects* in music [17]. This means multimodal units that combine sound and movement so that in addition to various sonic features we also have movement features such as proprioceptive, haptic, and visual images of trajectories and postures. This also means that there are movement-related schemata and constraints at work in gestural-sonic objects, i.e. various biomechanical and neurocognitive constraints such as limits to speed of movement, need for rests, etc., as well as the phenomena of *phase transition* and of *coarticulation*. Phase transitions mean that the speed of movement will lead to different groupings, e.g. speeding up will at some tempo threshold lead to fusion of pulses into a higher order pulse, slowing down will at some tempo threshold lead to fission of pulses into subdivision pulses. Coarticulation means that otherwise distinct sounds and movements will be hierarchically subsumed and contextually smeared so as to produce new emergent sensations, e.g. otherwise singular tone-events and movements fuse into superordinate phrases and movement shapes. Coarticulation seems to be one of the most important elements in the formation of chunks, and furthermore, concerns both the generation and the perception of musical sound [16].

Gestural-sonic images may be flexible, both with respect to resolution or acuity of detail, and with respect to generality by the principle of so-called *motor equivalence*. Motor equivalence means that motor images of singular actions may be generalized so as to encompass different versions of the action, allowing transfers and at the same time preserve basic cognitive schemata across variations. An example this is how the general category of 'hitting' is applicable to all percussion instrument actions, with or without mallets, as well as to all keyboard and struck string instruments.

4. TYPOMORPHOLOGY OF GESTURAL-SONIC OBJECTS

With chunk-level gestural-sonic objects as the basic local focus, we can differentiate various types as well as various features of such objects. Following the pioneering work of Pierre Schaeffer [13, 14], we can proceed in a top-down manner starting with depicting the global features of sonic objects and proceed on to successively finer differentiations of features. The main principle for Schaeffer was the subjective images of sonic objects, and where establishing correlations between these subjective images and the acoustic substrate of the sonic objects was seen as a long-term goal. It is also important to keep in mind that the ambition of Schaeffer was a universally applicable theory, equally valid for sonic objects in electroacoustic, instrumental, or vocal music, and applicable across different genres and musical cultures. Hence, such an approach could be seen as very much in accordance with a more open-ended, universal approach to MIR.

For a start, Schaeffer suggested three main classes of sounds based on their mode of production:

Impulsive: sounds that have a percussion like quality with a sudden onset followed by a decay, i.e. a discontinuous transfer of energy such as in hitting or kicking.

Sustained: a continuous transfer of energy so that the sound would be more or less stable throughout its duration such as in bowing, stroking, or blowing.

Iterative: sounds produced by a rapid series of impulses such as in a drum roll or in a tremolo.

It is the energy envelope of the sound that reflects the underlying assumed mode of sound-production, hence, that these sonic object types are transducers of movement information. This movement information can also be applied to pitch-related information with the following three main types:

Pitched: a more or less clearly perceptible and stable pitch throughout the duration of the sonic object.

Non-pitched: inharmonic or variably noise-dominated sounds with ambiguous or unclear pitch.

Variable: sensation of pitch that varies throughout the sonic objects, e.g. by glissando or vibrato.

Schaeffer combined these three pitch-related types with the three dynamic envelope types mentioned above into a 3 x 3 matrix of basic sonic objects in what he called the *typology*. The typology of sonic objects was a first and rough categorization to be followed by a more detailed depiction of features in what was called the *morphology* of the sonic objects. The morphology is basically concerned with the 'internal' features of the sonic objects such as its various pitch-related, dynamic, and/or timbral evolutions and fluctuations in the course of time. Two of the most prominent features of the morphology are the following:

Grain: fast fluctuations within the sound such as in the 'grainy' sound of a deep bassoon tone or in a flute flatterzunge.

Motion: slower fluctuations within the sound such as in slow ostinato or other textural movements. [4]

These features can be thought of as dimensions of sonic objects, and may also be further differentiated, e.g. the speed and amplitude of the grain fluctuations may be thought of as sub-dimensions, and variations in speed and amplitude may be thought of as further sub-dimensions to these dimensions. The exploration of thresholds for different feature values in relation to sound categories is then made possible, something that is useful for trying to determine categorical thresholds for salient features of sonic objects, hence for sonic features in general in a MIR context.

The typology and the morphology of sonic objects can be combined into an analytic system that for short is called the *typomorphology* of sonic objects. The general strategy here is then that of first attaching metaphorical labels to perceptually relevant (or salient) features of the musical sound, and then proceeding to differentiate various sub-features.

In summary, we believe that most (if not all) features of musical sound may be correlated to some kind of body movement. This is actually the main point of motor theory and embodied cognition, namely that we perceive by correlating whatever we hear (or see) to mental images of movement [6, 7].

5. SUGGESTIONS FOR IMPLEMENTATIONS

Given the abovementioned documentation of links between sound and body movement, the challenge now is to integrate our knowledge of such sound–movement links in audio analysis so that this can be useful in a MIR context.

Several of the features mentioned above can readily be found in audio using traditional analysis techniques. For example, the typological features can be correlated to the amplitude envelope of a sound signal and/or to the pitch contour or fluctuations in the spectral centroid. Details in the morphology, on the other hand, require more studies to be effectively implemented in a machine-based system. While it could be possible to implement this based on analysis of the sound alone, we believe that it may be worthwhile to also look at the movement of performers as well as listeners when they experience music.

As an example, consider the sensation of an undulating or even circular motion that we would assume many listeners would experience in the example illustrated with the motiongram in Figure 2. Although we may find considerable variation in the style of playing this piece, one source of such an undulating motion could be found in the sound-producing actions of the pianist. To an expert musician it might be natural or even obvious to predict from the score that pianists would tend to make this kind of undulating movements, yet it is an element that we believe could be captured and included in MIR as a feature of the music.

Figure 3 shows a graph of the movements of the wrists and elbows of a pianist performing the first 8 measures (with the upbeat figure) of the same piece as in Figure 2.

[4] 'Motion' is sometimes also rendered as 'gait' or 'allure' in English.

The graph is based on recordings with an infrared motion capture system and shows the markers' displacement along the keyboard (i.e. the horizontal plane). This is of course a crude simplification of the richness of the performance, yet we believe it does convey the salient feature of the undulating motion of this piece.

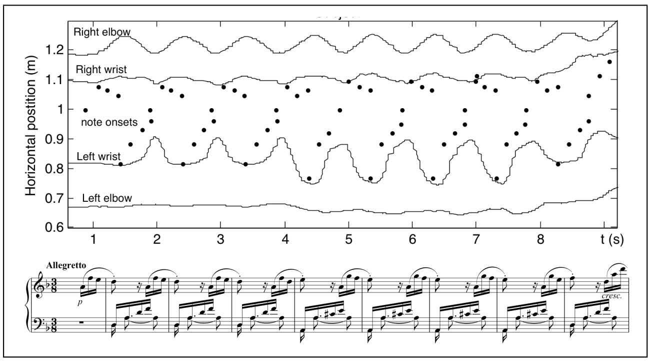

Figure 3. Trajectories of the wrists and elbows of a pianist performing the first 8 measures (and the upbeat measure) of the same Beethoven example as in Figure 2. The marked onset points are recorded from MIDI output from the digital piano used in the study.

Moving towards the analysis of body movement in a MIR context necessitates techniques to represent, store and navigate such movement data. We are here thinking about representations of data in many different forms, e.g.:

- Continuous data from various types of motion capture systems.
- Graphical representations of movement, both static and animated.
- Analyzed movement and gesture data in a structured and symbolic form.
- Various verbal movement metaphors.

Although there exist formats and standards that handle these types of data in other fields than music, we believe it is necessary to develop solutions that are specific to musical applications [18]. One of the most important parts here is to handle synchronisation between movement data, audio, video, MIDI, etc. We are not aware of any solutions that handle this issue in its full complexity, so for that reason we are currently developing the *Gesture Description Interchange Format* [5] (GDIF) as a system for streaming and storing motion capture data [19]. Equally important here is to work out a set of movement descriptors, and sound–movement descriptors, that are useful in a MIR context.

Also, considering that a substantial amount of music is readily available as audiovisual material (e.g. music videos of various kinds), this could be exploited if there were more readily available methods for analyzing both audio and video, and most importantly, for analyzing the *relationships* between features extracted from audio and video. This could then take into account the cross-modal interactions happening in our perception of audiovisual material, as documented in e.g. [20].

Finally, including an embodied perspective in MIR research could also open for new applications of search and retrieval of music through body movement. Using various types of motion capture techniques, ranging from camera-based to sensor-based systems, users could explore a large music collection through body movement. While this could certainly be done in low-dimensional features spaces, we believe that systems that manage to connect complex body movements to complex sound features will open for new and exciting ways of exploring the multidimensionality of musical sound, e.g. as implemented in software for concatenative synthesis [21]. Considering the positive results of the studies of air-performance and sound-tracing as mentioned above, this is something that both novices and experts should be able to do without a too high learning threshold.

It could be useful to regard music-related body movement as a link between otherwise separate elements in western musical thought: the acoustic signal, symbolic notation, and higher level aesthetic and semiotic significations of music. This is because music-related body movement may encompass all these elements at once: On one side the continuous body movement relates to the continuous acoustic signal, with sound-producing movements incorporating the tone events of notational symbols, and with various types of expressive features in the movement touching on aesthetic and semiotic elements. On the other side, music-related body movement contain valuable information of the musical experience that is not present in the audio itself, but which is often available in video material accompanying the sound.

6. CONCLUSIONS

Although we still have a long way to go in exploring music-related body movement and its relationship to musical sound, it seems that we already have reasonable grounds for claiming that sensations of body movement are essential in musical experience. Actually, we would even claim that sensations of body movement are one of the most salient features of musical style and genre, and could for this reason alone be an important element in the development of MIR. When we rather optimistically believe that music-related body movement has great (and mostly untapped) potential for MIR, we are also acutely aware of great challenges here, challenges that may be summarized as follows:

- Development of signal processing methods for extracting movement-inducing cues from audio.
- Development of video processing methods for extracting features of music-related body movement.
- Development of taxonomies and formats for handling such multimodal features in MIR systems.

[5] http://www.gdif.org

- Development of solutions for using body movement in searching, retrieval, and navigation in audio or audiovisual music files.

On the way to this, we need to continue working on what movement sensations listeners have to music, painstakingly building up our knowledge of subjective movement sensations and correlating these with lower-level signal-based features of musical sound.

7. REFERENCES

[1] M. M. Wanderley and M. Battier, eds., *Trends in Gestural Control of Music [CD-ROM]*. Paris: IRCAM – Centre Pompidou, 2000.

[2] A. Gritten and E. King, eds., *Music and Gesture*. Hampshire: Ashgate, 2006.

[3] R. I. Godøy and M. Leman, *Musical Gestures: Sound, Movement, and Meaning*. New York: Routledge, 2009 (in press).

[4] J. S. Downie, "The music information retrieval evaluation exchange (2005–2007): A window into music information retrieval research," *Acoust. Sci. & Tech*, vol. 29, no. 4, pp. 247–255, 2009.

[5] R. I. Godøy, "Motor-mimetic music cognition," *Leonardo*, vol. 36, pp. 317–319, August 2003.

[6] M. Leman, *Embodied Music Cognition and Mediation Technology*. Cambridge, MA: The MIT Press, 2007.

[7] R. I. Godøy, E. Haga, and A. R. Jensenius, "Playing 'air instruments': Mimicry of sound-producing gestures by novices and experts," in *Gesture in Human-Computer Interaction and Simulation, GW 2005* (S. Gibet, N. Courty, and J.-F. Kamp, eds.), vol. LNAI 3881, pp. 256–267, Berlin: Springer-Verlag, 2006.

[8] E. Haga, *Correspondences between music and body movement*. PhD thesis, University of Oslo, 2008.

[9] A. R. Jensenius, *Action–Sound : Developing Methods and Tools to Study Music-Related Body Movement*. PhD thesis, University of Oslo, 2007.

[10] R. I. Godøy, E. Haga, and A. R. Jensenius, "Exploring music-related gestures by sound-tracing - a preliminary study," in *Proceedings of the COST287-ConGAS 2nd International Symposium on Gesture Interfaces for Multimedia Systems* (K. Ng, ed.), (Leeds), pp. 27–33, 2006.

[11] P. Schaeffer, "Sound fragment from cd3, track 13, 20"-29" (no original source indicated) in [14]," in *Solfège de l'objet sonore*, Paris: (with sound examples by G. Reibel & B. Ferreyra), INA/GRM, 1998, first published in 1967.

[12] F.-R. Duchable, "Beethoven Concertos pour piano 1 and 3. A la decouverte des Concertos. François-René Duchable, piano, John Nelson, conductor, Ensemble Orchestral de Paris." [DVD] Harmonia Mundi, 2003.

[13] P. Schaeffer, *Traité des objets musicaux*. Paris: Editions du Seuil, 1966.

[14] P. Schaeffer, *Solfège de l'objet sonore*. Paris: (with sound examples by G. Reibel & B. Ferreyra), INA/GRM, 1998, first published in 1967.

[15] Z. Eitan and R. Granot, "Growing oranges on mozart's apple tree: 'inner form' and aesthetic judgment," *Music Perception*, vol. 25, no. 5, pp. 397–417, 2008.

[16] R. I. Godøy, "Reflections on chunking," in *Systematic and Comparative Musicology: Concepts, Methods, Findings. Hamburger Jahrbuch für Musikwissenschaft* (A. Schneider, ed.), vol. 24, pp. 117–132, Vienna: Peter Lang, 2008.

[17] R. I. Godøy, "Gestural-sonorous objects: embodied extensions of Schaeffer's conceptual apparatus," *Organised Sound*, vol. 11, no. 2, pp. 149–157, 2006.

[18] A. R. Jensenius, A. Camurri, N. Castagne, E. Maestre, J. Malloch, D. McGilvray, D. Schwarz, and M. Wright, "Panel: the need of formats for streaming and storing music-related movement and gesture data," in *Proceedings of the 2007 International Computer Music Conference*, (Copenhagen, Denmark), pp. 13–16, 2007.

[19] A. R. Jensenius, K. Nymoen, and R. I. Godøy, "A multilayered GDIF-based setup for studying coarticulation in the movements of musicians," in *Proceedings of the 2008 International Computer Music Conference*, (Belfast), pp. 743–746, 2008.

[20] B. Vines, C. Krumhansl, M. Wanderley, and D. Levitin, "Cross-modal interactions in the perception of musical performance," *Cognition*, vol. 101, pp. 80–113, 2005.

[21] D. Schwarz, G. Beller, B. Verbrugghe, and S. Britton, "Real-time corpus-based concatenative synthesis with Catart," in *Proceedings of the 9th Int. Conference on Digital Audio Effects (DAFx-06)*, (Montreal), 2006.