Classifying Music-Related Actions

Rolf Inge Godøy[*1], Alexander Refsum Jensenius[*2], Arve Voldsund[*3], Kyrre Glette[#4], Mats Høvin[#5], Kristian Nymoen[#6], Ståle Skogstad[#7], Jim Tørresen[#8]

[*]Department of Musicology, University of Oslo, Norway,

[#]Department of Informatics, University of Oslo, Norway

{[1]r.i.godoy, [2]a.r.jensenius, [3]arve.voldsund}@imv.uio.no {[4]kyrrehg, [5]matsh, [6]krisny, [7]savskogs, [8]jimtoer}@ifi.uio.no

ABSTRACT

Our research on music-related actions is based on the conviction that sensations of both sound and body motion are inseparable in the production and perception of music. The expression "music-related actions" is here used to refer to chunks of combined sound and body motion, typically in the duration range of approximately 0.5 to 5 seconds. We believe that chunk-level music-related actions are highly significant for the experience of music, and we are presently working on establishing a database of music-related actions in order to facilitate access to, and research on, our fast growing collection of motion capture data and related material. In this work, we are confronted with a number of perceptual, conceptual and technological issues regarding classification of music-related actions, issues that will be presented and discussed in this paper.

I. INTRODUCTION

For more than a decade, we have seen a growing interest in research on music-related body motion (Wanderley and Battier 2000; Gritten and King 2006, 2011; Godøy and Leman 2010). One basic tenet in most of this work is that sound and body motion are inseparable in musical experience, and that sonic features may be correlated with body motion features. Recently, we have in our own research focused on what we call *music-related actions*, meaning fragments of body motion in the duration range of approximately 0.5 to 5 seconds, actions closely linked with various sonic features in the same duration range. In our research, we have seen the need to develop a content-based classification scheme for music-related actions, both in view of our own work and in view of establishing a database of music-related actions open to researchers worldwide, enabling a sharing of data that we hope in the future will benefit the international research community.

This is a long-term and extensive project, bearing some resemblance with work in the field of *music information retrieval*, and associated classificatory schemes concerned with so-called *ontologies* of music (e.g. http://omras2.com). However, our challenge is different in that we are concerned with the domain of body motion as well as that of sound, and importantly, with the combination of motion and sound into multimodal chunks, combinations we believe are at the very core of music as a phenomenon.

Our classification and database work is based on the idea that it is necessary to have both data-driven, bottom-up methods and perception-driven, top-down methods, combining quantitative and qualitative approaches and correlating high-level sensations of motion and sound with low-level signal-based data. The top-down approach is based on starting from high-level, subjectively observed and annotated features such as the overall shape of motion and sound features, proceeding downwards into progressively more finely differentiated features. The bottom-up approach needs to take into account the following needs:

1) Means for streaming and storing music-related motion data from various types of sensors and motion capture systems, synchronized with the corresponding audio and video files.
2) Strategies for efficient post-processing of the above-mentioned data and media, as well as feature extraction, classification, visualization, sonification and various types of analyses and annotations of these data and media.
3) A local storage and retrieval system for all data and media mentioned in (1) and (2), including adequate metadata handling, as well as security measures for privacy and ethics considerations.
4) A web storage of some of the data available in (3), with the aim of providing data that other researchers can work with, compare to their own results, and possibly also use for creative applications.

We have previously presented a solution for recording data and media in a synchronized manner (Jensenius et al. 2008), and have also developed several different types of analysis and visualization strategies (Jensenius 2007, Nymoen et al. 2012). Our current efforts are targeted at creating solutions for a local storage, as well as a publicly available database of some of our recordings. A demo of the database will be accessible at our website from July 2012 (fourms.uio.no), and we envisage a continuous updating and adjustment of the various parts of this database over time.

There are, to our knowledge, no publicly available databases of music-related body motion data. Young and Deshmane (2007) announced a database of violin recordings, but it seems not currently to be found online. For more general human motion there are a few database projects, such as the *Carnegie Mellon University Motion Capture Database*, spanning various everyday and sports actions (mocap.cs.cmu.edu). However, our challenge is different in that our studies of action are focused on the linking of sound and action.

Given this background, the aim of this paper is to present our classificatory scheme for music-related actions with the hope of getting feedback from the music perception and cognition community on the various issues that we are working with. We shall in the next section first present some principles on the classification of action features, followed by a section on the classification of relevant sonic features, as well as a section on issues of database design, before making a summary of where we are now and where we are heading with this work in the future.

II. ACTION FEATURES

Given their great variety, and multifaceted nature, it is important to make an intuitive and top-down scheme for classifying music-related actions. Such a scheme should be general and flexible, allowing for as much detail in sub-categories as is needed, yet at the same time be easy to use and remember. For this reason, we shall first distinguish between different *functions* of music-related actions.

Functional classification

Following the basic scheme of various contributions in (Godøy and Leman 2010), we start with the readily observable distinction between *sound-producing* and *sound-accompanying* actions.

Sound-producing actions are all kinds of action that are associated with sound-production. We here distinguish between *excitatory* actions (e.g. hitting, bowing, scratching, blowing, etc.) and *modulatory* actions, i.e. actions that modify the resultant sound (e.g. making a vibrato with the left hand when bowing, changing bow position, opening and closing of mutes, etc.). In most cases of sound-production, there has traditionally been an energy transfer from the body to the instrument. However, with the advent of electronic instruments, this energy transfer link is in many cases no longer there, yet on/off switching and modulatory actions may remain (e.g. in using various input interfaces such as keyboards, pedals, faders, etc.).

Within the category of sound-producing actions, we also find various *ancillary* and/or *sound-facilitating* actions, meaning actions that help in shaping the sound (articulation, phrasing, etc.) or avoid fatigue and strain injury. Furthermore, on the border to sound-accompanying actions, we also find *communicative* actions, such as cues to fellow musicians or theatrical actions in view of influencing the audience. Notably, sound-producing actions may be multifunctional, i.e. serve several purposes at once, such as in a pianist's upbeat motion of hand-arm-shoulder-torso to an accented chord serving both the sound excitation, the communicative purpose of an upbeat cue to fellow musicians, and be communicative of expression to the audience.

The second main category of music-related movement is what we call *sound-accompanying actions*, which includes all kinds of actions that may be made to musical sound. This can be actions such as in dancing, walking, gesticulating, etc., be that in various performance contexts such as on stage or on screen in various multimedia settings, or in a multitude of listening situations, such as at concerts, in private homes, or in listening to music by headphones in public spaces. Sound-accompanying actions usually have some readily observable sound-matching features, such as synchronicity of motion with the rhythm of the music, or hand motion that trace dynamic, rhythmic, melodic, and sometimes also timbral features. In some cases, these sound-accompanying actions tend to imitate sound-producing actions, as in various instances of so-called *air-instrument* performance. In other cases, sound-accompanying actions may reflect more the resultant sound features or some global, affective features induced by the music.

Content classification

A further and equally basic classification applicable to both sound-producing and sound-accompanying actions is that of the *subjects involved*. This may include what type of background they have, e.g. whether they are trained musicians or dancers, and their sociocultural background. It is also necessary to classify whether they appear in a solo or ensemble setting, which type(s) of instrument(s) they play, etc. This will also include information about the context within which recordings were made, such as location, spatial setup, date and time of the recording, etc.

Next it is necessary to classify the content of the recordings themselves. The content is here highly dependent on the motion capture or sensor system used, and include information about the number and types of sensors or markers, and their placement on the body (or bodies). In the case of guitar performance, for instance, it may be interesting to focus on the finger motion on the guitar neck, whereas in a marimba performance, it may be interesting to include whole body motion in addition to the hands/mallets motion.

Furthermore, it is essential to include as much information as possible about the data type such as sampling rate, spatial resolution, synchronization issues, various conversions and/or preprocessing on the raw data, etc., without which the data may become useless in the future.

Timescales

Any music or dance performance will usually include several different types of actions, e.g. sometimes calm, protracted actions, at other times fast, jerky, large amplitude actions. In view of classifying music-related actions on the basis of continuous body motion, it is thus necessary to distinguish between different timescales. We have in the course of our research come to see three main timescale categories here, timescale categories also valid for sonic features, as we shall see in the next section:

- *Micro timescale*, meaning in the 0–0.5 seconds duration range where we may perceive continuous motion features like direction, speed, acceleration, smoothness, jerk, as well as more stationary features such as postures, effector position and effector shape.
- *Meso timescale*, typically in the 0.5–5 seconds duration range, in which we perceive more or less distinct action units or action Gestalts (Klapp and Jagacinski 2011), constituting what we refer to as *action chunks*.
- *Macro timescale*, meaning the timescale of longer stretches, of sections, and even of whole works, consisting of several concatenated meso-level action chunks.

As for the macro timescale, it is relevant to have global-level information about the action data, meaning information that applies to longer stretches of music-related body motion such as whole tunes or whole works. This includes easily accessible overview images of body motion trajectories such as by *motiongrams*, i.e. images showing motion over a shorter or longer duration (Jensenius 2007), as well as low-resolution animations of action trajectories for quick playback. Also some general motion quality features should be available, such as a global measure of *quantity of motion* and *kinesphere* (i.e. meaning the amplitude of motion) (Jensenius et al. 2010), index of *smoothness*, or conversely, of *jerk* (i.e. meaning the derivative of acceleration (Hogan and Sternad 2007)). These are features that can quite easily be derived from raw data and can give useful indications of overall mode of motion, affect, style, etc.

We have good reason to believe that music-related action chunks at the meso-level are the most significant both in the production and perception of music, and they are the main focus of our research. Detecting chunks by computers, however, is difficult because chunking usually depends on a combination of several cues in the signals, as well as a combination of cues in the signal with previously acquired mental schemas (Godøy 2008): living humans are typically never absolutely still, i.e. it is not possible to use periods with little motion as the sole criterion for the start and end of action chunks without some additional principles for thresholding, direction change, disregard of repetitions, etc. As humans we believe we can readily see (and hear) chunk boundaries based on the convergence of a number of cues and schemas, so chunking will thus for the moment mostly need to be done manually. However, we are exploring existing means for doing this automatically (Müller 2007), as well as developing our own schemes for this by studying coarticulation (Godøy, Jensenius, Nymoen 2010) and "reverse engineering" chunking from the effects of coarticulation in human motion.

Chunking

Based on the above-mentioned arguments we believe chunking of music-related actions is a composite phenomenon, but can in our context be understood as concerning the following main phenomena:

- Motion trajectory or shape, often with a more or less clear goal. Action chunks can be understood as centered on so-called *key-postures*, i.e. goals, with continuous motion trajectories between these key-postures. In everyday human body motion these key-postures are typically pragmatic goals such as in lifting an object, hitting a button, etc., whereas in music we understand the goals as salient moments such as downbeats, various accents or peaks.
- Action chunks are conceived and perceived holistically as *Gestalts*, as coherent shapes, and this also goes also for complex (e.g. polyrhythmic) patterns (Klapp and Jagacinski 2011). Action Gestalts assume anticipatory cognition where the whole chunk is preprogrammed and where the otherwise individual actions and sonic events (e.g. singular tone onsets) are fused by coarticulation into one superordinate unit (Godøy, Jensenius, and Nymoen 2010).
- Action chunks are based on constraints of *duration* and *rate*, as is suggested by research on differences between discrete and repeated actions (Hogan and Sternad 2007): if the chunk is too short and the repetition rate is too high, otherwise singular actions will tend to fuse into a new action unit. In other words, we have so-called *phase-transitions* between singular chunks and fused rhythmical motion by the two variables of duration and rate, something that applies equally well to sonic features, as we shall see in the next section.
- Also, chunking of body motion into actions will have to be musically meaningful, hence done in relation to sonic features as well. Typically, chunking is related to meter, i.e. to repeated cyclic motion, however, this cyclic motion need not be isochronous as may be readily observed in more "irregular" metrical and cycle duration patterns e.g. in Norwegian folk music.

Classification models

Thanks to the development of new software tools for analyzing music-related motion capture data (Toiviainen and Burger 2011), it is possible to systematically study a multitude of features of music-related actions that can help us in our classification work. Besides the macro-level, global features mentioned above, we have the means for classifying more local, meso-level chunk features with the following tools (Müller 2007):

- *Shape-based classification and comparison*: this includes both postures and trajectories where the postures can be classified by way of the markers' relative positions, frame by frame, and the trajectories can be classified by slicing the continuous trajectories of the markers into shorter segments.
- *Boolean classification*: a more categorical classification of postures and trajectories based on breaking down the data on the relative position of the markers into binary categories, i.e. left foot in front of right foot, right foot in front of left foot, etc., ignoring details of the motion and building up a compact, "low-resolution", but highly efficient classification for quick indexing and searching.
- *Dynamic time warping*: this enables point-by-point comparisons of actions with different durations by aligning similar events in time with each other, hence, providing a useful means for comparing similar variants of music-related actions and visualizing these with similarity matrices and other means.

In addition, machine learning techniques for classifying music-related actions could be very useful, however this still requires a lot of more development (Glette, Jensenius, and Godøy 2010).

III. SONIC FEATURES

Studies of everyday actions typically designate pragmatic goals for actions, such as grasping, lifting and moving an object (Rosenbaum et al. 2007, Grafton and Hamilton 2007). Music-related actions, with the exception of choreographies that enact everyday tasks such as the famous Barber scene in Chaplin's *The Great Dictator* (Godøy 2010), are typically not goal-directed in the same sense. However, music-related actions can be understood as goal-directed in terms of their resultant sound, such as in hitting a drum or making a fast down-bow motion. In this sense, the *sonic outcomes* could be regarded as the goals of music-related actions. This means that the intended rhythmic, textural, melodic, harmonic, etc. patterns regulate or control the sound-producing actions. But also listening to the sonic results of these actions will in most cases transmit salient information about the actions that generate the sound, e.g. listening to ferocious drum sounds will probably transmit sensations of energetic drumming actions, listening to slow, soft string music will probably transmit sensations of slow and supple bow motion, etc. This means that there is a reciprocal relationship between action features and sonic features, something that we need to take into account when classifying music-related actions.

Actually, the mimicking of sound-production is the main point in so-called 'air-instrument' performances (*air guitar*, *air drums*, etc.). But also in other cases, such as in more free gesticulation or in dance, there will often be similarities with sound-producing actions at a more general level, e.g. beating actions of the hands in a dance in sync with the drum pattern. The so-called *motor theory of perception* (with various variants), suggests that perception usually entails a covert mental simulation of the sound-producing actions believed to be at the source of what we are hearing. For sound-accompanying actions, we can see a continuum from clearly production-related actions to more loose sound feature-related actions, in all cases based on the idea that very many (perhaps most) features in the sound can be related to some kind of action trajectory.

We have in our work observed people making a number of music-related action trajectories based on perceived sonic features such as pitch, dynamics, timbre and texture. As a conceptual framework for these sonic features, we have adapted ideas on feature classification from Pierre Schaeffer's theories of the sonic object (Schaeffer 1966). We believe this can provide us with a very general, perception-based, top-down scheme that accommodates very many action-related sonic features ranging from the overall shape of the sonic object (i.e. its overall loudness, timbre and pitch-related envelopes) down to the minute fluctuations in loudness, timbre and pitch that we find in musical sound (Godøy 2006). This scheme necessitates that we first segment musical sound into shorter and holistically perceived sound chunks, similarly to the abovementioned segmentation of motion into action chunks. The criteria for chunking of sound are as composite as in the case for motion chunking mentioned in the previous section, i.e. a combination of qualitative discontinuities, of repeated patterns, of key-postures, and of shape (Gestalt) and the associated coarticulation, i.e. the contextual fusion of isolated sounds into more superordinate units.

Sonic typology and morphology

The overall loudness envelope of any sonic object is the basis for the initial classification presented by Schaeffer, a classification called the *typology of sonic objects* and which corresponds well with action features:

- *Sustained*, meaning a continuous sound and a corresponding protracted action providing continuous energy transfer to the instrument or the voice such as in bowing, blowing or singing.
- *Impulsive*, meaning a sharp attack, often followed by a longer decay, as typically the case with many percussive sounds, corresponding to an impulsive (sometimes called *ballistic*) action such as in hitting or kicking.
- *Iterative*, meaning a fast fluctuation as in a tremolo or a trill, corresponding to a shaking or rapidly repeated kind of action.

Although quite distinct in terms of sonic features, these categories are dependent on duration and event density, identical to what was indicated above for action categories: if a sustained sound is shortened it may turn into an impulsive sound, and conversely, if an impulsive sound is lengthened it may turn into a sustained sound. And: if an iterative sound is slowed down, it may turn into a series of impulsive sounds, and conversely, if a series of impulsive sounds is accelerated, it may turn into an iterative sound. We thus have phase-transitions here also, identical to what is valid for actions, demonstrating the close connection between sonic and action features, leading us to suggest that musical rhythmical-textural features are just as much action features as more pure sonic features.

These basic loudness envelope categories can be combined into more composite textures, as in instrumental or vocal textures with rapid foreground tones on a background of more sustained sounds, or in various rhythmical-textural patterns as found in dance music with different concurrent periodicities. All elements here may be related to actions, as is evident with the phenomenon of entrainment, i.e. the spontaneous dance-like actions that listeners may make to music. This also goes for other music-related actions such as for instance the abovementioned ancillary actions, actions that evidently are an integral element of music performance and that seem to be quite stable across long timespans and multiple performances of the same musical work. This has recently been the topic of a highly relevant classification project, where on the basis of a database of ancillary actions, a system of dictionaries for these actions has been worked out (Caramiaux, Wanderley, and Bevilacqua 2012).

Actually, all sonic features can be regarded as both being embedded in some action trajectory shape as well as itself having shape. This goes for all pitch-related features such as motives, ornaments, melodies, arpeggios and various other figures, as well as fluctuations within stationary pitches, all dynamic features such as *crescendo*, *decrescendo* and more composite envelopes, all timbral features such as changes from dull to brilliant, smooth to rough/noisy, etc., as well as expressive features of timing with envelopes of acceleration and deceleration, agogic accents, etc. The point is that we can differentiate sonic features at will and in a top-down manner from the chunk-level timescales down to various minute features in the sub-chunk-level timescale inspired by Schaeffer's so-called *typomorphology* of the sonic object, and consistently correlate these features with action features (Godøy 2006). For a more detailed survey of sound-producing actions and the close link between sonic and action features on traditional instruments, see (Halmrast et al. 2010).

Representations of sound

The basic tenet in our work is then that all sounds and sonic features can be seen as included in action trajectories, as related to action shapes. But this means that we also have challenges of representing sonic features as shapes. Traditional music notation may in many cases not be available, and would in any case be limited with respect to how well it represents sonic features as shapes. Spectrograms of recorded sound may then be more informative about shapes of sonic features, but will also often need some amount of processing and additional representations to give good shape-images of what we subjectively perceive. Some useful functions for this are found in e.g. the *Acousmographe* and *Sonic Visualiser* software.

This means that we have to also work on developing useful shape representations of salient sonic features so that we can correlate these shapes with action trajectory shapes. In Figure 1 we have an example of how such a perceptually salient sonic feature can be represented by way of a specific processing. In the top part of the figure, we see the spectrogram of a noise sound with a falling spectral centroid, and with a jagged descending curve indicating the calculated spectral centroid trajectory. In the bottom part of the figure we see the corresponding sound-tracing hand motion made by one subject to this sound.

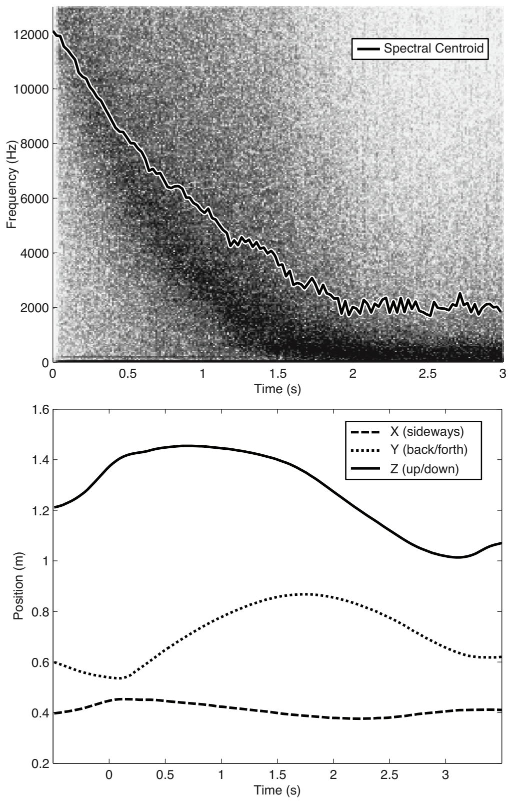

Figure 1. A so-called *sound-tracing* of a noise sound with a descending spectral centroid. The top part contains the spectrogram of the sound with the calculated spectral centroid indicated by the descending jagged curve. The bottom part contains a plotting in three dimensions of the trajectory of the subject's hand motion made to the sound.

IV. DATABASE DESIGN

To advance our knowledge and understanding of music-related actions, we need to consider larger collections of music-related actions and establish a database for this. The basic challenge of developing such a database is that of developing a good classification scheme for all components and levels in the database, in our case, ranging from many different types of raw and processed data and media, to various qualitative descriptions, as presented above. For such qualitative approaches, our database will need to have possibilities for various kinds of annotations (text, graphs, still-pictures, etc.), as well as the ability to play video, animations of sensor data and sound in a synchronized manner.

For more quantitative approaches, we need solutions for streaming and recording data from many different types of instruments (MIDI and other formats), sensor devices (inertial, physiological, etc.) and motion capture systems (inertial and camera-based). All these devices and systems typically output different formats, many of which are proprietary and based on particular commercial software solutions. A main challenge here is therefore to pre-process and format the data to some kind of workable standard to be useful in a database. This includes information about source device, setup (position and number of cameras or other devices) placement of markers on the body, sampling rate, resolution, etc., as well as scaling, to make the raw data comparable with other data within our framework. A draft for such a scheme, tentatively called *Gesture Description Interchange Format* (or *GDIF*, Jensenius et al. 2008), has been worked out and discussed with some of our international partners. A conceptual overview of this scheme can be seen in figure 2.

In addition to various low-level data, also analysis-based data will be possible to include in the database, ranging from low to higher levels of analysis. Low-level features such as different levels of derivatives (velocity, acceleration, jerk) can be useful in finding similar instances of motion qualities. Mid-level features such as global quantity of motion and chunking at various temporal levels can be useful to find different sections in longer files, as well as high-level features of e.g. emotion and affect. Also, there should be analysis-based visualization tools to allow for quick previews of pertinent features, e.g. in the form of a multi-track like playback interface with video, motion trajectories and sound.

In order to make the large amount of music-related actions material accessible for researches, there are also a number of other design issues that we have taken into account in developing our database:

- Metadata formats allowing for as many cross-references/searches as possible
- Tags: both for action features and sonic features as outlined in sections II and III above
- Analysis data from different sources and processing models
- User-friendly representations by various visual means, i.e. pictures, graphics, animations, etc.

The idea is to enter basic metadata into our database together with the data files, and then to have provisions (empty slots) for filling in various analysis-based information afterwards as depicted above in sections II and III.

In addition, there are ethical aspects that have to be taken care of, first of all concerning privacy protection of subjects who have participated in the recordings (regulated by Norwegian law and the research ethics regulatory body), but also with regards to copyright issues. This means that users of the database will have to sign a formal agreement before having access to the database. In this connection, users will also agree to follow the data format schemes so as to allow the material to be easily accessed by other users.

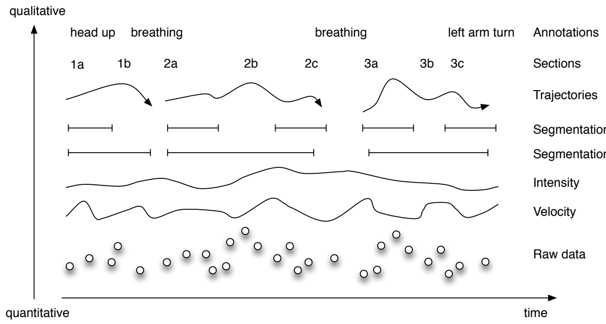

Figure 2. A conceptual overview of the GDIF scheme for the storage, classification, analysis and annotation of music-related motion and sound, starting from raw input data at the bottom to perceptually more pertinent features at the top. The horizontal axis represents time, and the vertical axis represents the bottom-to-top layers, i.e. the quantitative to qualitative dimension.

V. CONCLUSIONS

Much of the recent research on music-related body motion is motivated by the idea that music is a multimodal phenomenon, combining both auditory and motion sensations. This means that 'musical' features may in fact be regarded just as much as action features as sonic features. In our opinion, studying music-related actions is a way of doing musical research, and in this perspective, we need to develop a conceptual apparatus for handling music-related actions in a more systematic manner, as well as technologies that enable us to record, retrieve and analyze such actions efficiently. This is a long-term goal, and we hope our international colleagues will join us in this work by taking part in our discussions and using our coming database as a forum for developing our knowledge on music-related actions. This should hopefully provide us with the basis for an analytic-classificatory framework extensible to all multimedia art, e.g. music videos, film music, computer games, etc., and also inwards to increasingly finer detail, e.g. a multitude of timbral features, all based on the idea of correlating sonic features with action features.

REFERENCES

Caramiaux, B., Wanderley, M. M. and Bevilacqua, F. (2012). Segmenting and Parsing Instrumentalists' Gestures, *Journal of New Music Research, 41(1)*, 13-29.

Glette, K., Jensenius, A. R., and Godøy, R. I. (2010). Extracting Action-Sound Features From a Sound-Tracing Study. In *Proceedings of the Second Norwegian Artificial Intelligence Symposium*. (pp. 63-66). Trondheim: Tapir Akademisk Forlag.

Godøy, R. I. (2006). Gestural-Sonorous Objects: embodied extensions of Schaeffer's conceptual apparatus. *Organised Sound 11(2)*, 149–157.

Godøy, R. I. (2008. Reflections on Chunking. In A. Schneider (Ed.) *Hamburger Jahrbuch für Musikwissenschaft 24* (pp. 117-133). Frankfurt: Peter Lang.

Godøy, R. I. (2010). Gestural Affordances of Musical Sound. In R. I. Godøy and M. Leman (Eds.) *Musical Gestures: Sound, Movement, and Meaning* (pp. 103-125). New York: Routledge.

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

Godøy, R. I. and M. Leman (2010). *Musical Gestures: Sound, Movement, and Meaning*. New York: Routledge.

Grafton, S. T. and Hamilton, A. F. dC. (2007). Evidence for a distributed hierarchy of action representation in the brain. *Human Movement Science 26(4)*, 590–616.

Gritten, A. and E. King (Eds.) (2006). *Music and Gesture*. Hampshire: Ashgate.

Gritten, A. and E. King (Eds.) (2011). *New Perspectives on Music and Gesture*. Hampshire: Ashgate.

Halmrast, T., Guettler, K., Bader, R., and Godøy, R. I. (2010). Gesture and Timbre. In R. I. Godøy and M. Leman (Eds.), *Musical gestures: Sound, movement, and meaning* (pp. 183-211). New York: Routledge.

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

Jensenius, A. R., Nymoen, K., and Godøy, R. I. (2008). A multi-layered GDIF-based setup for studying coarticulation in the movements of musicians. In *Proceedings of the 2008 International Computer Music Conference*, (pp. 743–746). Belfast: ICMA.

Jensenius, A. R., Wanderley, M. M., Godøy, R. I., and Leman, M. (2010). Musical gestures: Concepts and methods in research. In R. I. Godøy and M. Leman (Eds.), *Musical gestures: Sound, movement, and meaning* (pp. 12-35). New York: Routledge.

Müller, M. (2007). Information Retrieval for Music and Motion. Berlin Heidelberg: Springer-Verlag.

Nymoen, K. R. I. Godøy, J. Torresen, and A. R. Jensenius (2012, in press). A statistical approach to analyzing sound tracings. In *Selected proceedings CMMR*. Berlin Heidelberg: Springer-Verlag.

Rosenbaum, D., Cohen, R. G., Jax, S. A., Weiss, D. J., and van der Wel, R. (2007). The problem of serial order in behavior: Lashley's legacy. *Human Movement Science 26*, 525-554.

Schaeffer, P. (1966). *Traité des Objets Musicaux*. Paris: Éditions du Seuil.

Toiviainen, P., and Burger, B. (2011). *MoCap Toolbox Manual.* Jyväskylä, Finland: University of Jyväskylä.

Young, D, and Deshmane, A. (2007). Bowstroke database: A web-accessible archive of violin. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, (pp. 352–357). New York, NY: ACM.

Wanderley, M. M. and Battier, M. (eds.) (2000). *Trends in Gestural Control of Music*. Paris: IRCAM.