Understanding Musical Instants

Rolf Inge Godøy, revised version July 2019

Abstract: We may typically experience music as continuous streams of sound and associated body motion, yet we may also perceive music as sequences of more discontinuous events, or as strings of chunks with multimodal sensations of sound and body motion, chunks that can be called *sound-motion objects*. The focus in this chapter is on how such sound-motion objects emerge at intermittent points in time called *musical instants*, and how musical instants are necessary in order to perceive salient features in music such as of timbre, pitch, texture, contour, and overall stylistic and affective features. The emergence of musical instants is also understood as based on the combined constraints of musical instruments, sound-producing body motion, and music perception, also suggesting that understanding musical instants may have practical applications in making music.

Keywords: continuity, discontinuity, intermittency, body motion, sound-motion objects, constraints

1. Introduction

The focus in this chapter is on our subjective experiences of brief events in music, of fragments of sound and associated body motion, typically in the 0.3 to 3 seconds duration range. Although we regard music as unfolding in time, the claim here is that fragments of music at this timescale may be perceived and conceived holistically and 'in a now', to borrow an expression from Edmund Husserl (Husserl 1991), hence, what may be called *musical instants*.

Musical instants can include any kind of sound fragments, be that instrumental, vocal, electronic, or environmental, and be singular (e.g. a trumpet tone, a single guitar pluck, a hammer hitting a metal rod) or composite (e.g. contain any kind of rhythmic, textural, melodic, harmonic, and timbral pattern), provided that such fragments are short and are perceived holistically as somehow coherent entities (e.g. as a fanfare motif, a rapid harp arpeggio, a waltz pattern, or a *wah-wah* sound of the opening and closing a trombone mute). Musical instants are crucial for our experience of music, because in spite of their rather short duration, they may be sufficient for perceiving not only basic musical features such as timbre, dynamics, and pitch, but often also more high-level salient musical features such as style, aesthetics, sense of motion, and affect, due to the cumulative and 'instantaneous' presence in our minds of sequentially unfolding sound and associated body motion events.

Yet the idea of musical instants is challenging because events that have temporal extension in our environment, i.e. sequentially unfolding sound and associated body motion, are transformed in our minds into something perceived as instantaneous, into seemingly more solid sound and motion images. How events that unfold in time can be recalled as instantaneous overview images is a conundrum that has received much attention in phenomenological philosophy and gestalt theory. The remarkable insights from some late nineteenth and early twentieth century thinkers in these domains on the coexistence of continuity and discontinuity in our experience of the world, can now be extended with more recent findings in cognitive psychology, in particular in human movement science, with insights about *intermittency* in motor control and effort. Intermittency is here understood as the constraint-based disposition of human behavior for discontinuous, point-by-point schemes of motion control and perception. And this is then the main idea of the present chapter, namely *to try to understand the subjective experiences of musical instants as grounded in the predispositions of our organism for intermittent cognition*.

To arrive at such an understanding of musical instants, we need to start out with some considerations of what seems to be a spontaneous and instantaneous perception of significations in listening (section 2). Considering elements of listening will be necessary when we in the succeeding sections go on to an overview of the timescales involved in the generation and perception of different musical features, e.g. duration thresholds for perceiving salient features such as pitch, timbre, texture, style and sense of motion (section 3). The topic of timescales in music is in turn closely linked with various constraints, both of our bodies and our minds, in the generation of music (section 4). Combined elements of timescales and production constraints will then in turn be necessary for understanding intermittency in music (section 5), as well as for understanding how perceptual and motor-related features in music may be thought of as shapes, as indeed instantaneous overview images of that which unfolds in time such as a melodic contour or a timbral transition (section 6). This all converges in what may be called *sound-motion objects*, meaning multimodal fragments of sound and body motion that may be experienced holistically as musical instants (section 7). Finally, it will be concluded that an enhanced understanding of musical instants may be useful in performance, improvisation, and composition, but needless to say, there are many important challenges ahead in securing a more solid basis for understanding the workings of musical instants (section 8).

2. Listening

For a start, it could be useful to look at some elements in the process of listening that seem to contribute to the sensation of musical instants. As suggested by the ecological approach to auditory event perception (Gaver 1993), we may tend to spontaneously assign labels to whatever it is that we are hearing in our environment, typically concerning causality and significance, e.g. 'bell', 'dog', 'car', 'approaching train'. Such auditory event perception seems to not require much reflection as to what we are hearing, nor will there usually be much focus on the sound features as such. Rather, the sound signal, or the 'carrier' of the event information, tends to 'disappear' the moment the meaning of the event has been extracted and processed further in our minds in interaction with pre-existing schemas and categories. By such fast and holistic perception of auditory events, it seems that we have 'auditory instants' in everyday listening, and that this seems to be a spontaneous and effortless capability for perceiving and thinking, similar to the fast *System 1* of Daniel Kahneman (2011)[1].

But as documented in auditory scene analysis research (e.g. Bregman 1990), event perception may require the sound to have a gestalt-like coherent ordering of low-level spectral features in the signal, as well as providing access to cues at higher levels of organization, often also to cues from other sense modalities, in order to work well. Hence, sound events are ontologically composite in the sense that they have multiple

[1] Our ability to provide fast, seemingly instantaneous judgments or answers (e.g. read words on a large billboard or calculate 2 + 2 = ?), is called "System 1" by Daniel Kahneman, and what he calls "System 2" denotes our ability to solve more complex tasks (e.g. fill our a tax form or calculate 17 x 24 = ?), tasks that for most people require more extended and step-by-step reasoning.

components, ranging from low-level physical signal features to various higher-level significations. This ontological complexity also raises issues of top-down schematic processing vs. bottom-up emergence in auditory perception, and the general view now seems to be that of a dynamical interaction between these two streams (Griffiths and Warren 2004, Bizley and Cohen 2013).

Given the top-down influence in listening, it means that our attitude, or what has come to be called our *intentionality* in listening, may be decisive for how we perceive sound. An interesting model for this was proposed by Pierre Schaeffer in connection with the development of his theory of so-called *sound objects* of the *musique concrète* (Schaeffer 1966). Sound objects in Schaeffer's theory are fragments of sound from a variety of sources, e.g. musical instruments, human voice, sound synthesis, or the environment, typically in the 0.3 to 3 second duration range mentioned above. The use of sound objects was initially a pragmatic tool for composition in the *musique concrète,* consisting of mixing the playback of looped fragments on phonograph disks. However, Schaeffer understood that after countless repetitions, the composers' attention tended to be diverted from the more immediate and/or everyday significations of the sound objects (cf. the notion of ecological approach mentioned above) to the more sonic-perceptual features of such looped sound fragments. This shift of attention, within Schaeffer's model, came to be called "reduced listening", and signified an intentional shifting of focus that is also an important feature of phenomenological philosophy (Husserl 1982)[2]. Crucially, this method of intentional listening enhanced the awareness of salient sonic features such as the overall dynamic and pitch-related shapes, what came to be called the *typology of the sonic objects*, and the multidimensional web of internal timbral-textural features, what came to be called the *morphology of the sonic objects* (more on shapes in section 6 below).

The typology concerns the energy envelope of the sound, and can be seen to correlate closely with images of body motion, e.g. the envelope of a percussive sound corresponding with an impulsive body motion (more on this in section 4 below). For this reason, it seems intuitive to extend the idea of sonic objects to also include corresponding body motion, hence, the extension of Schaeffer's concepts to "gestural sonorous objects" (Godøy 2006). This extension was based on the so-called *motor theory* perspective on perception (Galantucci, Fowler, and Turvey 2006), specifically on an adaptation of this theory that I have called *motormimetic cognition*, signifying that images of sound-producing body motion are closely linked with images of the sonic objects (Godøy 2001, 2003).

This means that images of sound-producing body motion often are reenacted in our minds when we listen to, or merely imagine, music. This is based on the fact that musical sound is (traditionally) included in a body motion trajectory, such as hitting, stroking, bowing, blowing, or singing. In this perspective, a sound event does not only include the audible components, but also the motion trajectory images of the sound-producing effector, e.g. the hand from the initial position to the impact with the drum, and the continuation after the impact with a rebound, and then a return to the starting point, or on to a new drum-hitting motion sequence.

In addition to mental images of such sound-producing body motion, motormimetic cognition in listening may also include so-called sound-accompanying body motion, such as in dancing, walking, or gesticulating to the music. Music seems to trigger a number of body motion responses, hence, the idea of multiple gestural affordances of

---

[2] Schaeffer came to see the affinity of his method with phenomenology, but he commented that he and his colleagues, with this pragmatic phonograph loop origin and use of reduced listening, were initially actually practicing phenomenology without knowing it (Schaeffer 1966, 262).

musical sound (Godøy 2010a). In many cases, there will be a similarity of shape between the dynamic envelope of the sound and the evoked body motion shape in the mind of the listener, e.g. a *sforzato* chord in an orchestra evoking the image of a rapid downward hand motion. Furthermore, such schemas may become generic and applicable across similar instances in listening. Thus, perceived patterns of sound-accompanying body motion may contribute to our perceptions of musical sound patterns in listening (Chemin, Mouraux, and Nozaradan 2014).

Importantly, this converges in understanding sound and motion chunks as related in view of gestalt principles, cf. the basic similarities between gestalts in perception and motor control (Klapp and Jagacinski 2011). Such chunks of sound and motion, what we shall call *sound-motion objects*, have features in parallel that are fused into coherent entities, entities that may be subjectively perceived as musical instants (see section 7 below).

3. Timescales

We may experience music as a continuous stream, as an uninterrupted flow of sound and motion sensations. Yet in the course of the unfolding music, we may shift our attention to certain just passed events in the music (e.g. to a high-pitched tone on a trumpet, to a drum fill, to a screaming guitar solo), as suggested by both the event based and sound object based accounts of listening mentioned above. But knowing that musical features are based on temporal unfolding, we may also realize that our perceptions in listening are based on cumulative and atemporal assessments of that which is inherently time-dependent.

A possible solution to this conundrum of temporal continuity and discontinuity in experiences of music could then be to differentiate various timescales involved in music. This means trying to classify salient perceptual features in music in view of their typical durations. The main questions will then be: what are typical minimum durations necessary for perceiving salient features such as the loudness, timbre, or pitch of any musical sound? Or for perceiving the textural, rhythmical, or melodic patterns, as well as the affective and motion-related features of any musical fragment? Clearly, this will create a rather lengthy list of feature-related timescales, extending from that of single impulses and vibrations in the sub-millisecond range to large-scale formal features in the range of several minutes and even hours. But to have a more concise overview of timescales, I have, in previous research on music-related sound and body motion, found it practical to distinguish three main timescales categories in view of their salient perceptual features (Godøy 2010b):

- *Micro*, with continuous features such as pitch, dynamics, and quasi-stationary timbre. These features may typically need very short durations in order to be clearly perceived, i.e. sometimes (dependent on content and context) down to the 20 milliseconds range and even lower (Suied et al. 2014).
- *Meso*, meaning fragments typically in the 0.3 to 3 seconds range as mentioned earlier, usually with clearly perceivable dynamic and pitch-related envelopes, as well as rhythmical, textural, melodic and harmonic patterns. In some cases, fragments with durations in the 300 milliseconds range may be coarsely identified (Gjerdingen and Perrott 2008), again suggesting that factors of content and context play an important role here. Significantly, the lower duration range here in listening (≈300 milliseconds), has also been suggested as the typical duration for intermittent motor control (more on this later).

- *Macro*, meaning several meso-level chunks in concatenation, constituting periods, sections, and whole works. This is not a well-researched area, but it seems that perception of such large-scale forms may allow more variation (i.e. alternative orderings of the elements) than the two other timescales (Eitan and Granot 2008).

These timescales are concurrent, i.e. they coexist in the music and are interdependent in the sense that there will often be contextual effects both upwards and downwards, e.g. that micro timescale features are found at the meso timescale, and that the meso timescale context modifies the perceptual framework for the micro timescale. As listeners, we may also intentionally zoom in and out of details, shifting our attention to different timescales. This is what Schaeffer denoted with the twin terms *context* and *contexture* in sound object research mentioned above (Schaeffer 1966), meaning that a given feature has both a context (e.g. a single tone in the middle of a phrase with several tones) and a contexture (e.g. the internal unfolding of the tone from the attack point to the last part of the decay), and that the interplay of these timescales is a crucial factor of musical experience.

In musical imagery (Godøy 2001), we seem to have the possibility, at will, to zoom between different levels of resolution, and also shift between slow motion and instantaneous reenactments, enabling what Paul Hindemith seemed to have in mind with the following: "If we cannot, in the flash of a single moment, see a composition in its absolute entirety, with every pertinent detail in its proper place, we are not genuine creators." (Hindemith 2000, p. 61).

This means that although the above-mentioned features are time-dependent, we may also think of these features qualitatively as atemporal, such as in assigning pitch and interval labels, rhythmical duration labels, formal labels, etc. As pointed out by Xenakis with his concepts of *en-temps* (in time) and *hors-temps* (outside time), we may freely move between depicting temporal unfolding and more atemporal features when we think about music (Xenakis 1992). This was in particular evident in Xenakis' own musical work demonstrating a close affinity between musical sound and more atemporal architectural shapes, such as in his orchestral work *Metastasis* and the architectural design of the Philips pavilion at the world's fair in Brussels in 1958, where whole passages of orchestral sound are indeed reflected as instants in the graphics of the score, as what we below will think of as the shape element of musical instants.

Interestingly, similar ideas of stepping out of time were previously discussed by Husserl and some of his contemporaries, leading up to Husserl's notion of 'now-points' (Husserl 1991, Schneider and Godøy 2001, Godøy 2008, 2010c, and 2011). For Husserl, these now-points were a necessity in order to extract any meaning at all, because without the stepping out of the continuous flow of time, there would only be an undifferentiated stream of sensations. The phenomenological view of discontinuity seems to agree with several other publications suggesting the need for perceptual coherence at the 'moment' timescale (e.g. of Michon 1978 or of Pöppel 1997).

In summary, we have concurrent timescales in music perception and imagery, i.e. micro timescale features embedded in meso timescale features, and these in turn embedded in macro timescale context, and they may all be part of our subjective experiences of musical instants. However, the meso timescale will be the most important here due to various constraints: physical constraints of both sound and instruments, and constraints of sound-producing body motion and of perception, necessitating that we now look at how these constraints contribute to experiences of musical instants.

4. Constraints

The repertoire of musical sound made by traditional (i.e. pre-electronic) means is indeed very great, yet there are also limitations on what various instruments can output, as well as limitations on what human performers can do. These limitations are the basis for several constraints that are interesting here because they contribute to the formation of musical instants in the form of what could be called *musical quanta,* meaning sound objects with salient shapes in their actual physical unfolding, e.g. the sound of a plucked guitar string, of a cymbal struck with a stick, a rapid trombone glissando (Godøy 2013).

The point of mentioning constraints here is not to in any way reduce the value of music, but rather to recognize that music, as an intrinsic human activity, is in fact piggybacking on certain constraints of instruments, as well as of our bodies and minds. It seems that we are so familiar with how constraints shape music, that we actually expect to find these constraints when we experience music. First, there are constraints of the instruments and sound:

- *Instruments* have a limited repertoire of possible sounds because of their construction, e.g. a plucked instrument can produce different kinds of pluck sounds, but not a bowed or blowed sound (unless using some additional tools).
- *Sound* is typically constrained by the combination of the instrument's physics and the human excitatory and modulatory motion on the instrument in question (or on the human vocal apparatus), e.g. usually limited in duration, event density, loudness, etc. because of human physical limitations.
- *Propagation* of sound, including space, rooms, medium, etc. also constrain sound, e.g. set limits to how far and for how long a sound may resonate, may tend to color the sound, and tend to smear sequences of distinct sounds.

As for sound-producing body motion, there are some significant constraints:

- *Human motion takes time*, i.e. there is no instantaneous displacement of the sound-producing effectors (fingers, hands, arms, feet, lips, tongue), hence there will always be motion trajectories between different postures in sound production, resulting in so-called *coarticulation* (more on this below).
- *Musicians need to optimize motion.* To avoid strain injury and to enable sustained performance, musicians will try to minimize the effort in sound-producing motion by reducing the amount and duration of muscle contraction, exploiting rebounds and conserving momentum.
- *Motor control requirements*, in particular of highly demanding musical passages, will typically need preprogramming and hierarchical control, as there is just not enough time to make decisions continuously (Grafton and Hamilton 2007). In particular, the so-called *psychological refractory period*, meaning the shutting off of new motor control messages within a certain timespan, seems to necessitate gestalt-like holistic conception and control in skilled motion (Klapp and Jagacinski 2011).
- *Slowness of our motor control system*. Now we see research suggesting that there is intermittency in motor control, meaning an uneven distribution of control, in a point-by-point kind of impulse-driven control, or by what has been called "serial ballistic control" (Loram et al. 2014), because our motor control system is too slow for continuous feedback control. This also means that in the phases between these intermittent control points, the effectors are left to carry out the motion without interference until the next intermittent

control point, as is typical of so-called ballistic motion, e.g. hitting, kicking, scraping, rapid stroking, etc.

- *Motor control by key-postures,* meaning optimizing planning and control of motion by preprogrammed shape and position of effectors (fingers, hands, arms, mouth, etc.) at salient moments in time, typically at downbeats and other accented points, and having continuous trajectories in the form of *prefixes* and *suffixes* between these key-postures (Godøy 2008). Derived from the use of so-called *key-frames* and *interframes* in animation (Rosenbaum et al. 2007, Rosenbaum 2017), key-postures may serve as orientation landmarks in time and space, and are particularly interesting in understanding the emergence of musical instants.

Furthermore, there are constraints of sound-producing body motion in the interaction with the instruments, resulting in biomechanically fairly distinct categories as well as correspondingly distinct sound categories:

- *Impulsive*: typically plucking, hitting, rapidly stroking, i.e. motion with burst of effort followed by relaxation as in ballistic body motion, and resulting in a percussive or plucked sound.
- *Sustained*: blowing, bowing, slowly stroking, hence, motion with more continuous effort, resulting in more protracted sound
- *Iterative*: shaking, trembling, rapid back-and-forth stroking or rolling, resulting in various kinds of tremolo and trills.

These are basic categories, and are usually mutually exclusive because of corresponding biomechanical constraints (e.g. impulsive motion is ballistic and discontinuous, whereas sustained motion is continuous). But there are also other salient emergent features due to body motion constraints:

- *Phase-transitions*, meaning that there may be a change of motion category by changes in the duration and rate of motion elements. For instance, if a singular impulsive motion is progressively lengthened, it may change into what is sensed as a continuous motion, and conversely, if a continuous motion is progressively shortened, it will turn into an impulsive motion; and if we have a series of well-spaced impulsive motions which are becoming progressively more densely spaced, these may turn into an iterative motion, cf. an illustration of this in Figure 1. Such category changes due to incremental changes in one or more of its parameters is referred to as *phase-transition* (Haken, Kelso, and Bunz 1985), and is one of the most important principles for chunking motion, hence also for the sensation of musical instants.
- *Coarticulation*, meaning a similar fusion of small-scale motion elements into larger-scale motion and sound chunks, however here emerging from motion between a series of key-postures as schematically illustrated in Figure 2. The principle of coarticulation is well known in linguistics (Hardcastle and Hewlett 1999) and various fields of human movement science (Rosenbaum 2010), and stems from the fact that immediate displacement of effectors is not possible (i.e. that all human motion takes time), and that there will be transition motion between key-postures. For instance: fingers cannot move instantaneously from one piano key to another and will often also require that the whole hand, and even arm and shoulder, move in preparation for the next key depression. Or: the shape of the mouth cannot change instantaneously from that of one vowel to a different vowel but will require a more gradual transition from one shape to another shape. This means that there will be a contextual smearing of motion in coarticulation, both backwards, called *carry*

over effects, and forwards, *called anticipatory effects*. In addition, there will be so-called *spatial coarticulation* in the sense that an effector may recruit motion of several other effectors, e.g. finger motion in piano performance recruiting hand, arm, shoulder, and even whole-body motion. Thus, coarticulation means activation spreading both backwards and forwards in time as well as to other parts of the effector set. It seems that coarticulation may contribute to create coherence both in the sound-producing motion and the resultant sound in music (see Godøy 2014 for details).

Common to phase-transition and coarticulation is the creation of coherence at the chunk level by fusion of otherwise separate elements into gestalt-like units. This may then in turn be seen as the source of emergent sensations of musical instants.

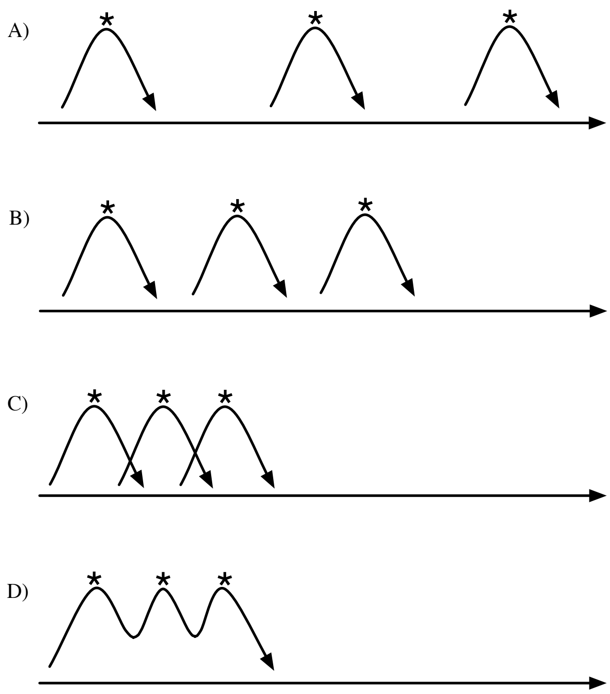

Figure 1: Schematic illustration of phase-transition by decrease of inter-onset distance. Three initially distinct impulsive events in A), when moved closer together as in B), will result in the new back-and-forth motion trajectory in C), hence, make a qualitative change from impulsive to iterative motion.

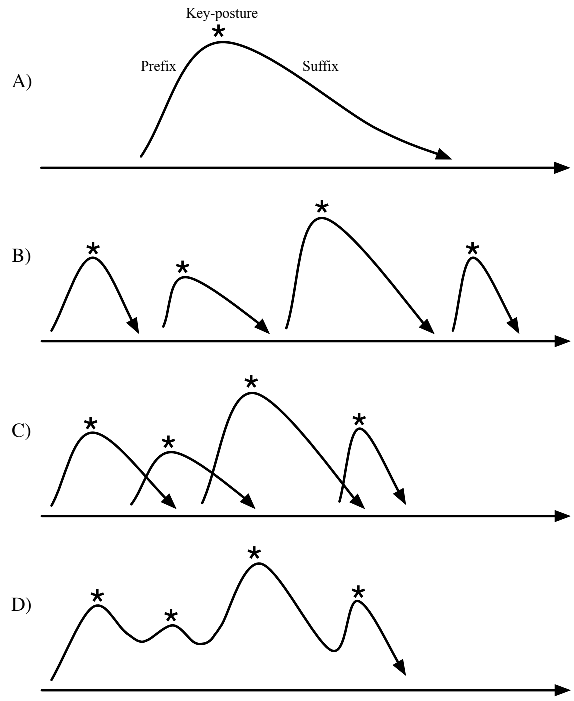

Figure 2: Schematic illustration of one single key-posture with its prefix and suffix, hence with motion returning to the starting point in A), and of four intermittent key-postures each with likewise prefixes and suffixes in B). What happens when these four key-postures are moved closer in C) is that a new and undulating motion trajectory emerges due to coarticulation.

5. Intermittency

From the above-mentioned elements of listening, timescales, and constraints in music, we arrive at the key concept of intermittency in music-related body motion. The concept of intermittency has a background in phenomenology, in particular in Husserl's above-mentioned idea of discontinuity in time perception (Husserl 1991), but there has been a remarkably similar line of thinking emerging in recent years from the domain of human motor control.

For a start, it has been claimed that human motor control is anticipatory, that it works by pre-programming, or in a feed-forward manner (called *open loop* control), because in skilled and fast motion there is not enough time for continuous corrective feedback (for so-called *closed loop* control). The relative sluggishness of our organism's motor control is thus seen as the direct cause of intermittent motor control, and as resulting in our organism instead working by a series of control impulses, or in a point-by-point manner, with pre-programmed motion chunks (Karniel 2013, Loram et al. 2014, Sakaguchi et al. 2015). Although the neurophysiological workings of intermittency remain to be better understood, relevant research seems to offer clear behavioral arguments in favor of intermittency in body motion generation.

It may be hypothesized that intermittency in control also entails intermittency of effort, because control cannot be separated from energy-requiring motion, i.e. there can be no 'pure' control, analogous to 'power steering' in machine control systems, in human body motion. There is evidence from electromyographic (EMG) recordings of muscle activity that effort is not equally distributed throughout body motion chunks (Aoki, Tsukahara, and Yabe 1989). And in line with the mentioned motormimetic approach to music perception, it is hypothesized that this intermittency extends to music perception with projections of schemas from sound-producing body motion onto whatever sound we are hearing, contributing to the formation of sound-motion objects (see section 7 below).

To document these preprogrammed chunks and their triggering in music-related body motion, it is now feasible to study the motion trajectories of these chunks using motion capture technologies, and document the trajectories of effector displacement (of fingers, hands, arms, etc.) in relation to the key-postures on an instrument, i.e. points in the motion of particular significance for motor control (cf. the presentation of key-postures in the previous section). It is also possible to calculate the derivatives of this effector motion, primarily velocity and acceleration (Hogan and Sternad 2007). Typically, points of velocity reversals, e.g. at the striking point in a percussive motion where the direction suddenly changes, may indicate intermittent key-posture points (Godøy 2013).

In addition, it is possible to study the effort in generating these motion trajectories using EMG (electromyographic) recording of muscle activation to learn more about the temporal distribution of such effort. Using EMG is in particular interesting for detecting the non-active segments in time, so-called *pre-motion silent periods* (Aoki, Tsukahara, and Yabe 1989), so as to get a clearer impression of discontinuities of effort.

6. Shapes

A common feature of much recent research on musical sound and on human body motion is that of representing various features as shapes; in sound, we have shapes of the unfolding sound in the so-called time domain, at different levels of resolution ranging from the waveform to more superordinate envelopes, and in the so-called frequency domain of spectral content, we may have shape images of both stationary and more transient components; and in human movement science, we typically have shape representations of postures and of motion trajectories, but also of derivatives showing crucial elements such as the velocity reversals mentioned above. The common feature of these shape representations is that they capture that which unfolds in time in more atemporal, instantaneous images. Said differently: *Shapes are inherently instantaneous* in the sense of being geometric entities that we subjectively perceive, or have in our field of vision, 'in a now' and 'all at once'.

Although we might think of shapes as primarily in the visual domain, shapes have extensive applications in a general and amodal sense. An important body of research on shapes in human cognition has emerged from so-called *morphodynamical* theory (Thom 1983, Petitot 1990, Godøy 1997), with the basic principle that shapes are inherently holistic and can give us a better understanding of various phenomena in our world than more abstract, symbol-based representations. To envisage a complex phenomenon as a shape is not only a matter of scientific method, but may also be an intrinsic feature of perception and cognition in general, as is evident from the large number of shape metaphors used in most areas of human experience.

And because of their generic nature, shapes may mediate between the modalities, e.g. shapes of sound-producing motion may be correlated with shapes of pitch contours. How our minds make such connections between different modalities remains to be further explored, however, we can observe that the use of verbal shape metaphors and graphical shape representations, ranging from score notation to more signal-based shape representations of features, is so widespread in music-related contexts that there can be little doubt about the pragmatic utility of shape images in music (Godøy 2017). Readily available technologies enable the exploration of shape correlations of most sound and motion features, for instance attack transient shapes, formantic shapes, articulation shapes, and various pitch, dynamic, and timbral nuance shapes, with motion-related shapes such as of postures, motion trajectories, motion hierarchies, rhythmical patterns, etc.

We can thus think of two main classes of shapes in our context, those of music-related body motion, and those of music-related sound features. For body motion shapes, we typically have the following:

- *key-posture effector shapes*, e.g. of fingers/hands for playing chords, of mouth for formants in vocal sound
- *effector motion shapes*, e.g. hand moving across keyboard with fingers playing an arpeggio, mouth changing shape from vowel to vowel
- *derivatives of motion*, i.e. shapes of velocity and acceleration

These shapes may be anticipatory or retrospective, may include both quasi-stationary key-posture and motion trajectory shapes, but are typically 'all-at-once', and hence, instantaneous. As for music-related sound features, we often have the following shapes:

- *dynamic contours*: crescendo, decrescendo, tremolo at different speeds
- *pitch contours*: various shapes of successive tones and sub-tone inflections
- *modal shapes*: cumulative pitch images with distinct interval constellations
- *harmonic shapes*: single chords and/or several chords in succession
- *spectral shapes*, both stationary and dynamic timbral features
- *rhythmic shapes*: patterns of sound onsets and durations
- *textural shapes*: wefts of sound events
- *articulation shapes*: accents, staccato, legato
- *timing shapes*: accelerando, ritardando, various fluctuations

The crucial property of shapes here is that they are distributed, i.e. reflect the temporal unfolding of both sound and motion, and are also inherently holistic in that they are perceivable 'all at once'. This means that shapes are non-symbolic in the sense of what Schaeffer called *concrete* as opposed to that which is symbol-oriented and called *abstract* (Schaeffer 1966).

In short, shapes are manifestations of musical instants by being compressed images of that which unfolds on the fragment timescale, typically in the 0.3 to3 seconds duration range. Also, images of shape can work across modalities, hence, be the basis for *sound-motion objects*, i.e. salient multimodal meso timescale fragments of music.

7. Sound-motion objects

The elements of intermittency and of shape converge in what can be called *sound-motion objects*. Based on Pierre Schaeffer's concept of sound objects, sound-motion objects are, as mentioned above, basically a multimodal extension of Schaeffer's idea of sound objects by also including sensations of body motion in musical experience (Godøy 2006). And similar to Schaeffer's notion of object, the object view of sound and motion here is based on the idea that sequentially unfolding features may be kept in memory as cumulative retrospective or prospective images, typically including the overall dynamic, pitch-related, timbre-related envelopes, and the overall trajectories of the sound-producing or sound-accompanying motion.

As an example, think of the sound of someone saying *wah-wah* where there is an overall envelope of timbre and dynamics, as well as an overall envelope of the mouth changing shape. The sound is continuous from the start to the end of the fragment, and so is the motion of the mouth, from the beginning when the mouth start to open and to the end when the mouth is closing. And in the course of this trajectory, the motion results in the distinct spectral shape of the *wah-wah* sound (as we know, on a trumpet or on a trombone, a similar *wah-wah* sound would be made by opening and closing a mute). This means that we now can combine previous insights on gestalt-based object coherence in auditory perception theory (e.g. Bregman 1990) with insights from gestalt-based motor control theory (Klapp and Jagacinski 2011), and also supplement these insights with findings on the efficiency of object attention (De Freitas, Liverence, and Scholl 2014).

The main element here is the combination of intermittency in motor control, i.e. of the previously mentioned *serial ballistic control* (Loram et al. 2014), with a slower and noise-prone motor system, so that intermittent impulses set the musculoskeletal system in motion, in turn producing a continuous sound output. This means that each sound-motion object is triggered by an impulse, and hence, each sound-motion object may be seen as focused on one instant, something I have tried to document with data on velocity reversals in sound-producing body motion, e.g. with the impact-points in piano or percussion performance (Godøy 2013).

We may thus think of sound-motion objects as multimodal, as just as much objects of motion as of sound. This also means that motion coherence may become a schema for sound coherence, hence, the idea that musical gestalts are also action gestalts. Furthermore, we can hypothesize that sensations of musical instants are linked to sensation of intermittent control in body motion, that the serial ballistic control scheme results in a series of motion chunks that due to their very nature of being controlled intermittently, may be perceived as coherent entities and as continuous within their duration.

8. Conclusions and prospects

The main point of understanding the musical instant is to reconcile continuity and discontinuity in musical experience. This can arguably be done by demonstrating that sound and motion are both continuous and coherent at the meso timescale of sound-motion objects, yet that these sound-motion objects are driven by intermittent impulses of control and effort. In summary, sound-motion objects are:

- preprogrammed holistic entities
- impulse-driven
- centered on key-postures
- internally coherent and continuous in time as combined sound and action gestalts

Furthermore, musical instants, based on the intermittency of both control and effort, may provide insights on:

- the human body's strategy for generating fast and accurate music-related motion in a not so fast motor control system by using anticipatory, feed forward, control impulses
- the optimization of skilled music-related body motion by intermittent effort, cf. experts' energy reduction by minimizing muscle contraction times and exploiting conservation of momentum

Insights on the workings of musical instants could have practical applications in music performance, improvisation, and composition by:

- thinking all kinds of rhythmical patterns, including syncopations and polyrhythmic patterns, as single impulse driven patterns (cf. Klapp et al. 1998)
- practicing ornaments as pre-programmed sound-motion objects driven by impulses, e.g. in fast drum set fills (Godøy, Song, and Dahl 2017) or fast violin trills (Godøy et al. 2016)
- conceive of all kinds of textures as piecewise impulse-driven, e.g. tremolo, trills, arpeggios, dance patterns, etc. as impulse-driven sound-motion objects (Godøy 2018)
- provide tools for studying high-level salient features such as of style, aesthetics, affect, and sense of motion, also by focusing on the holistic features of meso timescale sound-motion objects (Godøy 2018)

Needless to say, we have significant challenges here, first of all in documenting the existence of intermittent control in music-related body motion. The cited research (Karniel 2013, Loram et al. 2014, Sakaguchi et al. 2015) has used a variety of more indirect methods for this, and there are also other directly signal-based methods that seem promising (Inoue and Sakaguchi 2015). In my own work and that of colleagues, we have tried to look at features of motion data, in particular of acceleration and velocity reversals, as well as data on trajectory coherence and smoothness, for traces of intermittency.

Presently we are supplementing motion data with EMG data recordings in order to see what muscle activation patterns can tell us about intermittency in both control and effort, as well as to develop an impulse-response type model of intermittent control in sound-producing motion (cf. the work reported in Plamondon et al. 2013, Berio et al. 2017).

But also sound features can tell us something about intermittency and coherence, and we believe various tools from music information retrieval (MIR) can be useful in extracting perceptually salient shape and intermittency data from sound files, data that can be correlated with the corresponding sound-producing and sound-accompanying motion data.

All the above-mentioned research effort will contribute to our understanding of the musical instant as a crucial element in musical experience, including its potential practical applications in performance, improvisation, composition, and sound design. The overall goal here is to demonstrate how music is based on a series of subjectively perceived and conceived instants, as well as how these instants fuse into experiences of larger-scale musical works.

References

Aoki, H., Tsukahara, R., & Yabe, K. (1989). Effects of pre-motion electromyographic silent period on dynamic force exertion during a rapid ballistic movement in man. *European Journal of Applied Physiology*, 58(4), 426-432.

Berio, D., Akten, M., Leymarie, F. F., Grierson, M., & Plamondon, R. (2017). Calligraphic Stylisation Learning with a Physiologically Plausible Model of Movement and Recurrent Neural Networks. In Proceedings of MOCO '17, London, United Kingdom, June 28-30, 2017. DOI: 10.1145/3077981.3078049

Bizley, J. K., & Cohen, Y. E. (2013). The what, where and how of auditory-object perception. *Nature Reviews Neuroscience*, 14, 693–707.

Bregman, A. (1990). *Auditory Scene Analysis.* Cambridge, Mass., and London: The MIT Press.

Chemin, B., Mouraux, A., & Nozaradan, S. (2014). Body Movement Selectively Shapes the Neural Representation of Musical Rhythms. *Psychological Science,* 25(12), 2147–2159.

Chion, M. (1983). *Guide des objets sonores.* Paris: INA/GRM Buchet/Chastel.

De Freitas, J., Liverence, B. M., & Scholl, B. J. (2014). Attentional Rhythm: A Temporal Analogue of Object-Based Attention. *Journal of Experimental Psychology*, 143(1), 71–76.

Eitan, Z., & Granot, R. Y. (2008). Growing Oranges on Mozart's Apple Tree: 'Inner Form' and Aesthetic Judgment, *Music Perception,* 25(5), 397–417.

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, 13(3), 361-377.

Gaver, W. W. (1993). What in the World Do We Hear? An Ecological Approach to Auditory Event Perception. *Ecological Psychology*, 5(1), 1-29.

Gjerdingen, R., & Perrott, D. (2008). Scanning the Dial: The Rapid Recognition of Music Genres. *Journal of New Music Research*, 37(2), 93–100.

Godøy, R. I. (1997). *Formalization and Epistemology.* Oslo: Scandinavian University Press.

Godøy, R. I. (2001). Imagined Action, Excitation, and Resonance. In R. I. Godoy and H. Jorgensen (eds.), *Musical Imagery* (pp. 239-252). Lisse: Swets and Zeitlinger, 237-250.

Godøy, R. I. (2003). Motor-mimetic Music Cognition. *Leonardo,* 36(4), 317-319.

Godøy, R. I. (2004). Gestural Imagery in the Service of Musical Imagery. In A. Camurri & G. Volpe (eds.), *Gesture-Based Communication in Human-Computer Interaction: 5th International Gesture Workshop, GW 2003, Genova, Italy, April 15-17, 2003, Selected Revised Papers, LNAI 2915.* Berlin: Springer, 55-62.

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 Music. In A. Schneider (Ed.), *Systematic and comparative musicology: concepts, methods, findings.* (pp. 117-132). Frankfurt: Peter Lang.

Godøy, R. I. (2010a). 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. (2010b). Images of Sonic Objects. *Organised Sound*, 15(1), 54-62.

Godøy, R. I. (2010c). Thinking Now-Points in Music-Related Movement. In R. Bader, C. Neuhaus and U. Morgenstern (Eds.), *Concepts, Experiments, and Fieldwork. Studies in Systematic Musicology and Ethnomusicology.* (pp. 245–260). Frankfurt am Main: Peter Lang.

Godøy, R. I. (2011). Sound-action awareness in music. In D. Clarke and E. Clarke (Eds.), *Music and Consciousness.* (pp. 231-243). Oxford: Oxford University Press.

Godøy, R. I. (2013). Quantal Elements in Musical Experience. In. R Bader (Ed.), *Sound – Perception – Performance. Current Research in Systematic Musicology, Vol. 1.* (pp.113-128). Berlin, Heidelberg: Springer.

Godøy, R. I. (2014). Understanding Coarticulation in Musical Experience. In M. Aramaki, M. Derrien, R. Kronland-Martinet & S. Ystad (Eds.), *Sound, music, and motion*. *Lecture Notes in Computer Science.* (pp. 535-547). Berlin: Springer.

Godøy, R. I. (2017). Key-postures, trajectories and sonic shapes. In Dan Leech-Wilkinson and Helen Prior (Eds.), *Music & Shape.* (pp. 4-29). Oxford: Oxford University Press.

Godøy, R. I. (2018). Sonic Object Cognition. In Rolf Bader (ed.), *Springer Handbook of Systematic Musicology*. (pp. 761-777). Berlin: Springer Nature.
Godøy, R. I. & Leman, M. (2010). *Musical Gestures: Sound, Movement, and Meaning.* New York: Routledge.
Godøy, R. I., Song, M., Nymoen, K., Romarheim, M. H., & Jensenius, A. R. (2016). Exploring Sound-Motion Similarity in Musical Experience. *Journal of New Music Research*, 45(3), 210-222.
Godøy, R. I., Song, M.-H., & Dahl, S. (2017). Exploring Sound-Motion Textures in Drum Set Performance. Proceedings of the SMC Conferences. ISSN 2518-3672. s 145- 152.
Grafton, S. T. & Hamilton, A. F. (2007) Evidence for a distributed hierarchy of action representation in the brain. *Human Movement Science*, 26, 590–616.
Griffiths, T. D. & Warren, J. D. (2004). What is an auditory object? *Nature Reviews Neuroscience*, 5(11), 887-92.
Haken, H., Kelso, J. & Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. *Biological Cybernetics,* 51(5), 347–356.
Hardcastle, W. & Hewlett, N. (eds.) (1999). *Coarticulation: Theory, Data and Techniques*. Cambridge: Cambridge University Press.
Hogan, N. & Sternad, D. (2007). On rhythmic and discrete movements: reflections, definitions and implications for motor control. *Experimental Brain Research*, 181, 13–30.
Husserl, E. (1982). *Ideas Pertaining to a Pure Phenomenological Philosophy, First Book*. Doredrecht: Kluwer Academic Publishers.
Husserl, E. (1991). *On the Phenomenology of the Consciousness of Internal Time, 1893 -1917*. English translation by John Barnett Brough. Doredrecht: Kluwer Academic Publishers.
Inoue, Y. & Sakaguchi, Y. (2015). A wavelet-based method for extracting intermittent discontinuities observed in human motor behavior. *Neural Networks* 62, 91-101.
Kahneman, D. (2011). *Thinking, fast and slow.* New York: Farrar, Straus and Giroux.
Karniel, A. (2013). The minimum transition hypothesis for intermittent hierarchical motor control. Frontiers in Computational Neuroscience 7, DOI=10.3389/fncom.2013.00012
Klapp, S. T., Nelson, J. M., & Jagacinski, R. J. (1998). Can People Tap Concurrent Bimanual Rhythms Independently? *Journal of Motor Behavior*, 30(4), 301-322.
Klapp, S. T., & Jagacinski, R. J. (2011). Gestalt Principles in the Control of Motor Action. *Psychological Bulletin*, 137(3), 443–462.
Loram, I. D., van De Kamp, C., Lakie, M., Gollee, H. & Gawthrop, P. J. (2014). Does the motor system need intermittent control? *Exercise and Sport Science Review,* 42(3), 117–125.
Michon, J. (1978). The Making of the Present: A Tutorial Review. In J. Requin (ed.) *Attention and Performance VII*. (pp. 89–111). Hillsdale, NJ: Erlbaum.
Petitot, J. (1990). 'Forme' in *Encyclopædia Universalis.* Paris: Encyclopædia Universalis.
Plamondon, R., O'Reilly, C., Rémi, C., & Duval, T. (2013). The lognormal handwriter: learning, performing, and declining. *Frontiers in Psychology* 4(945), doi: 10.3389/fpsyg.2013.00945
Pöppel, E. (1997). A Hierarchical model of time perception. *Trends in Cognitive Science*, 1(2), 56-61.

Rosenbaum, D. A. (2010). *Human Motor Control* (Second Edition). San Diego, CA: Elsevier/Academic Press.

Rosenbaum, D. A. (2017). *Knowing Hands: The Cognitive Psychology of Manual Control.* Cambridge: Cambridge University Press.

Rosenbaum, D., Cohen, R. G., Jax, S. A., Weiss, D. J., & van der Wel, R., 2007: 'The problem of serial order in behavior: Lashley's legacy', *Human Movement Science*, 26(4), 525-554.

Sakaguchi, Y., Tanaka, M., & Inoue, Y. (2015). Adaptive intermittent control: A computational model explaining motor intermittency observed in human behavior, Neural Networks, 67, 92-109, http://dx.doi.org/10.1016/j.neunet.2015.03.012.

Schaeffer, P. (1966). *Traité des objets musicaux.* Paris: Éditions du Seuil.

Suied, C., Agus, T. R., Thorpe, S. J., Mesgarani, N., & Pressnitzer, D. (2014). Auditory gist: Recognition of very short sounds from timbre cues. *The Journal of the Acoustical Society of America, 135*(3), 1380-1391.

Thom, R. (1983). *Paraboles et catastrophes.* Paris: Flammarion.

Xenakis, I. (1992). *Formalized Music* (Revised edition). Stuyvesant: Pendragon Press.