

Chunking in Music by Coarticulation

Rolf Inge Gody, Alexander Refsum Jensenius, Kristian Nymo
fourMs, Department of Musicology, University of Oslo

January 31, 2010

Abstract

In our own and other research on music-related actions, findings suggest that perceived action and sound are broken down into a series of chunks in people's minds when they perceive or imagine music. Chunks are here understood as holistically conceived and perceived fragments of action and sound, typically with durations in the 0.5 to 5 seconds range. There is also evidence suggesting the occurrence of coarticulation within these chunks, meaning the fusion of small-scale actions and sounds into more superordinate actions and sounds. Various aspects of chunking and coarticulation are discussed in view of their role in the production and perception of music, and it is suggested that coarticulation is an integral element of music and should be more extensively explored in the future.

1 Introduction

It is commonly accepted that when listening to music people usually perceive units such as phrases, measures, motives, or other kinds of distinct sonic events, reminiscent of how people perceive phrases, words, or syllables when hearing speech. This unit-formation in music has received much attention in traditional Western music theory, in ethnomusicology, in past and more recent music cognition, as well as in auditory research (see Godøy (2008) for an overview). Often denoted by terms such as *segmentation*, *paraling*, or *punctuation*, we prefer to use the term *chunking* to denote such unit-formation. In the English language the term *chunking*, besides signifying the

cutting up of something into smaller units, also signifies the fusion or transformation of various small entities into larger and more solid units. Our use of the term chunking has its origin in the seminal work of G. A. Miller and what he called *re-coding* in perception (Miller 1956), signifying a transformation of sensory information in people's minds. Studying chunking is then not only about studying the cutting up of streams of sound, but equally much about trying to understand the emergence of somehow stable mental images of musical sound in our minds.

In our research on music-related actions,¹ we have seen evidence for perception of musical sound as closely linked to mental images of sound-producing and sound-accompanying actions. By this we mean that listening to for instance ferocious drumming may result in mental images of energetic hand movement, and that listening to for instance slow and soft string music may result in mental images of slow, protracted bow movements. This in turn means that chunking in music may be closely related to chunking of body movements, and that performers and listeners may tend to relate unit-formation in sound to unit-formation in sound-producing and/or sound-accompanying actions. We are particularly interested in how the phenomenon of *coarticulation*, defined as the fusion of otherwise separate actions into larger action units, may contribute to the formation of chunks in music as suggested by the title of this paper: *chunking in music by coarticulation*. We shall now first have a look at some elements of music-related actions, timescales, and chunking, before going on to some principles of coarticulation in general and in

¹<http://www.fourms.uio.no>

music in particular, and to some remarks on the role of chunking by coarticulation in our understanding of music in general.

2 Music-related actions

The idea of strong links between musical sound and various sound-producing and/or sound-accompanying actions represents what could be called an *embodied* understanding of music perception and cognition (Godøy 2003, Leman 2008, Godøy & Leman 2010), something that should be understood in the light of a general ‘embodied turn’ in the cognitive sciences (see e.g. Gallese & Metzinger (2003)). The fundamental principle of this so-called *embodied cognition* paradigm is that perception and cognition are based on sensations of body movement (see e.g. Wilson & Knoblich (2005), Gallese & Lakoff (2005)). This means that people tend to relate whatever they perceive and/or imagine to mental simulations of associated actions, e.g. in listening to, or imagining, music, speech, or various everyday sounds.

Initially, the idea of strong links between perception and action in listening was suggested several decades ago in linguistics by the so-called *motor theory* of perception (Liberman & Mattingly 1985, Galantucci et al. 2006). Often criticized, the motor theory and various variants of this theory have since the advent of techniques for brain activity observation received increasing support, both as a theory of perception in general (Berthoz 1997), and in cases of sound perception in particular. There are suggestions that the strong links between listening and sensations of movement are neurophysiologically ‘hard wired’ (Kohler et al. 2002, Hickok et al. 2003), but also suggestions that sound-motor links in music perception are enhanced with training as can be seen in the case of expert musicians (Haueisen & Knösche 2001) but also in cases of novices after rather short period of musical training (Bangert & Altenmüller 2003). Also, there is evidence that the visual perception of musicians’ performance movements may significantly alter the perception of the music, either in the direction of enhancing the expressive features

(Davidson 1993, Clarke & Davidson 1998, Camurri & Moeslund 2010), or in the direction of sometimes producing new emergent effects (Vines et al. 2005).

What emerges from this fast growing literature on embodied cognition and sound-motor links is the idea that perceptually salient musical features may be reflected in mental images of various kinds of music-related actions. But the ubiquity of sound-motor links in music can of course also be seen in overt behavior quite simply by observing how people move to music at concerts, in dancing, and in innumerable everyday listening situations. Thus, there is presently a convergence of evidence from a number of different sources for the importance of sound-motor links, altogether suggesting that body movement is in fact an integral part of music as a phenomenon (Godøy & Leman 2010).

When studying the large variety of music-related actions of listeners as well as of performers, it is useful to have some kind of classification scheme that allows us to make more systematic connections between music-related actions and sonic features. As an initial classification, it has been suggested that there are two main groups of music-related actions, *sound-producing* actions and *sound-accompanying* actions (Jensenius et al. 2010).

The first group, sound-producing actions, includes both *excitatory actions* such as hitting, stroking, bowing, blowing, and *sound-modifying actions* such as changing pitch or making vibratos (e.g. by left hand movements on string instruments) or timbre changes (e.g. by shifting bow positions on string instruments or moving mutes on brass instruments), as well as various kinds of *sound-facilitating actions*. Sound-facilitating actions, are not directly sound-producing or sound-modifying, but still important for performers to facilitate expressivity and articulation, or to avoid fatigue and strain injury, and may include actions such as swaying the whole upper body while producing sound with the fingers and arms (Wanderley & Battier 2000, Wanderley et al. 2005).

The second group, sound-accompanying actions, includes all kinds of actions that are not sound-producing, such as listeners moving hands, arms, head, or the whole body, to the pulse of the music or to various contours of the music. However,

often there may be no sharp divide between sound-producing actions and sound-accompanying actions, e.g. as can be seen in the various communicative and/or theatrical actions that performers make when playing instruments (Jenseni et al. 2010).

As for the sound-producing actions, it seems that most listeners, regardless of level of musical training, spontaneously sense quite a lot of the action features that are necessary to produce sound. This may be observed in so-called *air instrument* performance such as *air guitar*, *air drums*, or *air piano*, where listeners reproduce ‘in the air’ what they believe are the appropriate sound-producing actions. Although the amount of detail in such air performances may vary with the level of musical training, we found that even listeners with little musical training had a fairly good idea of the effort and kinematics involved in the sound-producing actions (Godøy et al. 2006b). A similar overall consensus for music-related actions could be observed in studies of so-called *sound-tracing* (Godøy et al. 2006a, Haga 2008). Here listeners were asked to spontaneously draw the gesture shapes they associated with sound excerpts using a digital pen on a graphical tablet. For musical excerpts with few concurrent features there was a fair amount of agreement, but for excerpts that had many features in parallel there was less consensus, and we could also see how musically trained subjects had a more analytic approach when selecting features to draw. In a study of dancers’ spontaneous movement to musical excerpts, there also seemed to be a fair amount of consensus with regard to the overall sense of effort and mode of movement, although again there would be variations in detail (Haga 2008). Other research seems to confirm similar spontaneous association of movement images to salient features of musical sound (Eitan & Granot 2006).

The basic tenet of this embodied view is that any sonic event will be embedded in some kind of action trajectory: tones on a piano, a violin, a drum sound, etc., are all the result of the actions of effectors, e.g. finger, hand, arm, etc., and these actions start before the audible onset of any tone or sonic event and often continues afterwards. For this reason it is interesting to look at excitatory actions and their relation to the sonic results. Following the classificatory scheme sug-

gested in Pierre Schaeffer’s typology of sonic objects (Schaeffer 1966, 1998, Godøy 2006), there are the following basic excitatory categories:

Impulsive: discontinuous effort and transfer of energy, i.e. a short burst of effort, followed by relaxation. Such actions are typically found in percussion performance, and usually result in a sound with a sharp attack followed by a longer or shorter decay.

Sustained: continuous effort and transfer of energy, such as in protracted bowing or blowing, usually resulting in a sound with a more gradual attack followed by a more flat envelope.

Iterative: the fast repetition of sound onsets such as in a tremolo, a drum roll, or a washboard stroking, resulting in a corresponding rapidly fluctuating sound.

Also, in many music-related actions there may be different speeds in parallel. For example, when a pianist plays scales or arpeggios on the piano, the pianist combines fast finger movements with slower wrist and elbow movements, and possibly also with even slower shoulder and torso movements. We may thus speak of music-related actions at different levels of resolution, ranging from detailed to more coarse, and where the fast, detail actions may be included in a more superordinate action trajectory. This is in fact a case of what we call chunking by coarticulation. In our observations of air piano performance, we noticed that such chunking is also readily perceived by listeners, e.g. subjects rendered sweeping scales as sweeping hand, arm, shoulder, and torso movements (Godøy et al. 2006b).

3 Timescales

Although the sound-producing categories of impulsive, sustained, and iterative are quite distinct with respect to both sense of bodily effort and sonic results, there may also be categorical transitions between them. For instance, if an iterative sound-producing action is slowed down beyond a certain

threshold, it tends to become split into singular impulsive actions, and conversely, if singular impulsive actions are speeded up, they tend to become fused into a continuous iterative sound. We thus speak of *phase-transitions* (Haken et al. 1985), meaning that categorical boundaries are crossed by varying the rate and/or duration of the actions and the associated sounds. Such phase-transitions result in the forming, or conversely, the splitting up, of chunks, as has previously been documented in Gestalt theory inspired auditory research (Bregman 1990).

In line with these principles of event rate and duration, there is converging evidence for the existence of what we call *meso-level chunks* in music. Pierre Schaeffer suggested focusing on meso-level chunks in music several decades ago with his idea of the *sonic object* (Schaeffer 1966, 1998), a holistically perceived fragment of sound typically in the range between 0.5 to 5 seconds. Initially a pragmatic tool developed in the early days of the *musique concrète*, the notion of the sonic object evolved into becoming the basis for Schaeffer's theory of music perception. This theory is universal in its scope, applicable to instrumental, vocal, or electroacoustic music, Western and non-Western. From our present day perspective, we see the following main arguments in favor of the sonic object as the perceptually most salient timescale of music:

- *Perceptually salient timbral, dynamic, and pitch-related envelopes* are typically to be found within the 0.5 to 5 seconds range, as argued by Schaeffer (1966, 1998).
- *Rhythm and texture patterns of musical styles* are typically within the same sonic object duration range as can be seen from formalizations of musical style (Cope 1991), and as suggested by listening experiments by Eitan & Granot (2008).

Seeing a privileged status of meso-level chunks or sonic objects in music perception, we are working according to a tentative three-level model of timescales in music-related actions:

Sub-chunk level: the continuous movement trajectories and the corresponding continuous sound.

Duration thresholds for perceiving steady pitch and timbral features, as well as event simultaneity and event order are found on this timescale (see Moore (1995) for an overview of duration thresholds).

Chunk level: the timescale of salient sonic features, as well as the timescale of sound-producing action units.

Supra-chunk level: the timescale of longer contexts such as sections, movements and tunes, where we have concatenations of several chunks.

In line with mainstream theories of echoic memory, short-term memory and long-term memory in music (Snyder 2000), it seems reasonable to assume that sub-chunk, chunk, and supra-chunk levels interact in musical experience so that the sub-chunk level provides the internal sensory features of the chunk, and that the supra-chunk level provides the context and the global experience in music.

4 Principles of chunking

The basic idea of chunking is that sequentially occurring body movement and sound is perceived holistically as a series of units. This means that body movements are compressed into more instantaneously retrievable action images in our minds. Chunking can be understood as a necessity, built into our cognitive apparatus, as well as reflecting essential features of the world. There are several arguments converging to support this idea:

- As was suggested by phenomenological philosophy more than a century ago, people need to somehow break up the continuous stream of sensory experience into meaningful units, otherwise people would only have an amorphous mass of sensations (Husserl 1991). Using the example of the melody, Husserl argued that people can only perceive a sequence of tones as a melody if they can keep several sequentially occurring tones in their consciousness together, 'in a now'. He further argued that perception and cognition proceed by a series of what he called 'now-points',

where each now-point contains a chunk (Godøy 2010).

- Neurophysiological research seems to suggest that attention spans and experiences of the present moment fall within roughly the same time window as for the sonic object, i.e. have a mean duration of approximately 3 seconds (Pöppel 1997). For events or sensations significantly longer than that, Pöppel suggests there will be a shift of attention, as for instance the experience of a foreground-background flip when staring at so-called bi-stable figures, e.g. the Necker cube.
- Mean duration of everyday actions seem to be found in the approximately 3 seconds range, as documented in extensive studies by Schleidt & Kien (1997). To what extent this preference for 3 second action chunks is based on biomechanical constraints (e.g. need for rest, shift of posture, etc.) or motor control constraints is not clear, however as pointed out by Pöppel (1997), there may be a mutual attuning of action and cognition in this 3 second preference.
- The need for chunking sequentially occurring sensations into units has been much studied in linguistics and other cognitive sciences, and models of short-term memory have been suggested to account for these phenomena, typically with durations in the range of a few seconds (Snyder 2000).
- From an auditory perspective it is well known that the sequentially occurring features of sound, in particular the attack segment followed by sustain and decay segments, tend to mutually influence each other, implying that the entire sound chunk is kept in memory, or that there is a kind of ‘resonance’ for the entire chunk, as has been studied in speech perception (Grossberg & Myers 2000).
- In the domain of motor control, the topic of pre-planning of action chunks vs. a more continuous control of action has been hotly debated for more than a century (Elliott et al. 2001). There

is a lucid summary of arguments in favor of pre-planning of action in Rosenbaum et al. (2007), partly reviving the pioneering work of K. Lashley (1951), work which claimed that chunking is a naturally occurring phenomenon in action. According to Rosenbaum et al. (2007), there is for instance the phenomenon of so-called *end-state comfort*, which means that people often start an action in view of how it is going to end. This means that people conceive of an action as a chunk through anticipatory motor control.

- Rosenbaum et al. (2007) further suggest that people tend to plan and execute actions by a series of *goal-postures*, with continuous movements in between. We are currently trying to integrate this idea of goal-postures into our model of music-related actions, where body postures (meaning effector position and shape, e.g. finger, hands, arms, etc. in relation to a keyboard) at accents and other salient points in the music serve as points of orientation.
- Furthermore, anticipatory chunking in musical performance is a sign of expertise. In a study of skill development through musical practice, Drake & Palmer (2000) note that subjects demonstrated “increased anticipatory behavior and a greater range of planning with skill and practice. A strong positive relationship between the mastery of temporal constraints and planning abilities within performance suggests that these two cognitive indicators are closely related and may arise from segmentation processes during performance.” (Drake & Palmer 2000, 1).

But chunking is also related to various qualitative discontinuities in the signal, as suggested for instance by Bregman (1990) and (Schaeffer 1966). Transitions between sound and silence, prominent changes in pitch, timbre, dynamics, etc., may bring about a sense of chunk boundaries. However, qualitative discontinuities in the signal may not be enough in cases where there are competing discontinuities, as in a rapid sequence of sounds, or where there is a sequence of identical sonic events. It is for example known that listeners may tend to project metrical

patterns onto streams of pulses that do not have any acoustic cues for metrical organization (Fraisse 1982). For this reason, we suggest there may be two sources of chunking in music:

Exogenous: fairly clear and/or unambiguous qualitative discontinuities in the signal that may induce sensations of start and end points of chunks.

Endogenous: internally originating chunking schemes, e.g. as in meter projection on a series of non-changing pulses. Endogenous chunking involves volitional and top-down projection of chunking schemes by the perceiver based on his/her prior knowledge and intentional focus, hence may also result in disagreements between perceivers in cases of ambiguity, e.g. in the perception of metrical patterns in music from different cultures.

In musical practice, we furthermore assume that both sources of chunking may be found, depending upon the kind of music listened to and/or the expertise of the listener. However, in both cases we see indications of motor schema for action chunks to be at work, be that as a particular action schema associated with a particular sound, e.g. the chunking of tones in a rapid piano passage, or as a general schema for action organization, e.g. the chunking of sounds into cyclical metrical patterns.

5 Principles of coarticulation

It is commonly agreed that coarticulation entails a subsumption of otherwise distinct actions into more superordinate actions. This also has consequences for perception in that there is a corresponding contextual smearing of individual events into more superordinate events. Coarticulation is a naturally occurring phenomenon in body movement, both in the sense that moving one effector, e.g. finger, hand, tongue, lips, ‘spills over’ into neighboring areas of the body, and in the sense that any movement will be embedded in a temporal context. Furthermore, coarticulation can also be seen as an advantageous phenomenon:

...it is a blessing for us as behaving organisms. Think about a typist who could move only one finger at a time. Lacking the capacity for finger coarticulation, the person’s typing speed would be very slow. Simultaneous movements of the fingers allow for rapid responding, just as concurrent movements of the tongue, lips and velum allow for rapid speech. Coarticulation is an effective method for increasing response speed given that individual effectors (body parts used for movement) may move relatively slowly. (Rosenbaum 1991, 15).

Coarticulation concerns several areas of human movement, but has been most extensively studied in linguistics (Hardcastle & Hewlett 1999). Sometimes also referred to as *continuous speech*, it is interesting to see that coarticulation in linguistics is sometimes modeled as centered around what is called *phonemic goals*, and that there are coarticulated action trajectories to and from these goals (Perkell 2007). A similar understanding of goal-centered coarticulation can be found in instances of fingerspelling (Jerde et al. 2003), in facial animation (Cohen & Massaro 1993), as well as in hand writing (Kandel et al. 1993), and as we shall see in the next section, seems to apply to music as well.

Coarticulation can be understood as *movement context*, i.e. that at any one point in the course of a movement, the position and the shape of the effector is constrained by what to do next as well as by what has just been done, hence that there is a constant contextual smearing going on. This we refer to as *temporal coarticulation*, but we also have what has been called *spatial coarticulation* (Farnetani & Recasens 1999), meaning that in order to move one effector, e.g. a finger in piano performance, it may also be necessary to move the wrist, elbow, shoulder and sometimes also the whole torso, so that there is a spillover effect from one effector to neighbour effectors.

With regard to music-related actions, it is in particular the anticipatory element that is important, meaning that the emergence of the chunk is partly due to the anticipatory movement of the sound-

producing effectors. But there are also backward or carryover effects of coarticulation in that what has been done influences what is done now, and what will be done in the future. Related to this backward effect is that of the response of the instrument where previously excited sounds may smear future sound, if the previously excited sounds are not damped and are allowed to sound until their energy is dissipated. This may stretch the classical definitions of coarticulation, but it may still be argued to be a physical phenomenon of past actions shaping future states. This is also in line with the idea of a mass-spring model, which can be found in music in cases of multiple excitations and incomplete damping between sounds.

6 Coarticulation in music performance

In general, coarticulation is about *continuity*, about movement as continuous and about the human body as made up of interconnected effectors. Coarticulation is thus a natural phenomenon based on constraints of the human body, and it is really just from the symbol-oriented perspective of Western music notation that it becomes something 'added' to, and smearing, something that was purportedly not smeared to begin with. Yet there have been relatively few published studies of coarticulation in music. These studies are mostly concerned with coarticulation in performance, such as for anticipatory finger movement in piano performance (Engel et al. 1997, Jabusch 2006, Jerde et al. 2006, Palmer 2006) and in string instrument performance (Wiesendanger & Kazennikov 2006), and there are also indications of similar anticipatory movements in drumming (Dahl 2006) and in bow movements (Rasamimanana & Bevilacqua 2008).

Our own research on coarticulation has so far focused on pianists' coarticulatory movements of hands, arms, shoulders, head, and upper body in relation to finger actions, i.e. to tone onsets. This includes both *spatial coarticulation* (the degree of effector activation for any single or multiple tone events) and *temporal coarticulation* (contextual smearing by

carryover effects and anticipatory movements). One major challenge in our research is that of capturing the pianists' movements in an as unobtrusive manner as possible in more or less realistic performances of musical excerpts. The most unobtrusive method is to use regular video recordings as the basis for analysis, and one of the present authors has developed a technique for extracting *motiongrams* from such recordings that reveal global movement features (Jensenius 2007). This is based on calculating the *motion image* by subtracting consecutive frames in the video stream, and then averaging each row in the frame to a one pixel wide column that can be plotted over time. The end result is an image that displays movement over time. An example of a motiongram showing coarticulatory movement features in relation to sound can be seen in Figure 1, where an interesting correspondence between the overall movement shapes and piano sound can be seen.

Motiongrams are useful to get coarse overview images of movement, but for more precise data on effector movement, position data from various points on the body are needed. Having experimented with electromagnetic tracking, accelerometers, and infrared motion capture camera systems for position data in research on coarticulation (Jensenius et al. 2008, Godøy et al. 2008, 2009), we have found infrared camera systems to be the most unobtrusive as well as most precise, accurate and fast technology for capturing position data in piano performance.

In this paper we are presenting movement data from recent recording sessions with two professional pianists (both had participated in the earlier motion capture recordings mentioned above), each performing 8 excerpts of piano music containing different technical features (scales, arpeggios, ornaments, leaps, repeated tones/chords, etc.) at different tempi and with different articulations, so as to provide us with a broad selection of material involving coarticulation. Reflective markers were placed on the hands, wrists, elbow and shoulder joints, as well as on the chest, back and lower back, and on a cap on the head of the pianists, as proposed by the "Marker Placement Protocols" of LifeModeler Inc.² The pianists

²http://www.lifemodeler.com/LM_Manual.2007/A_motion.htm

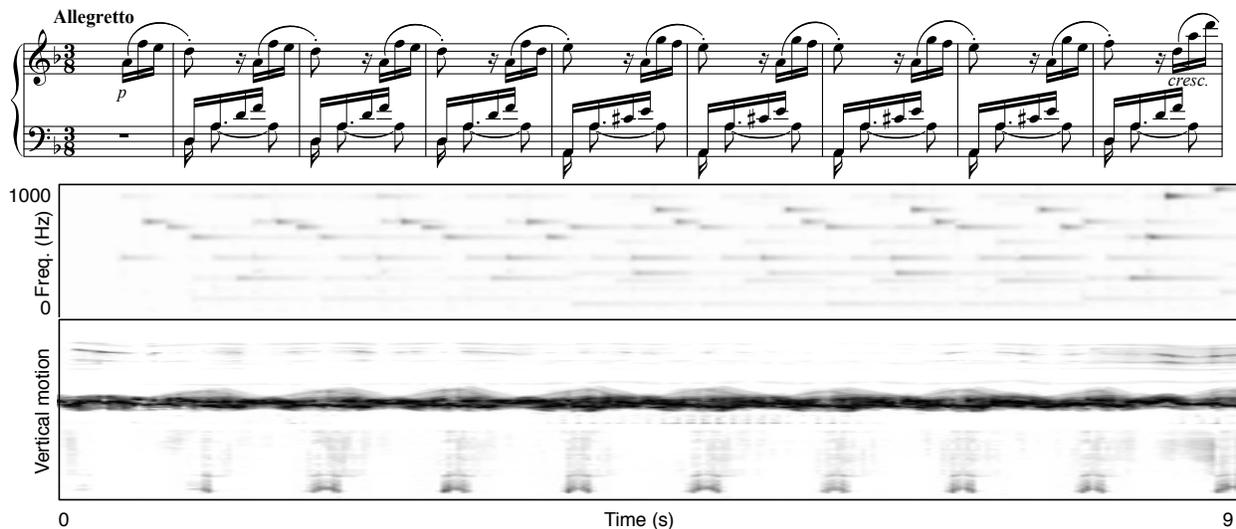


Figure 1: Score (top), spectrogram (middle) and motiongram (bottom) from a performance of the first 8 measures of the opening of the last movement of L. v. Beethoven’s *Piano Sonata nr. 17 Op. 31 no. 2 in d-minor, The Tempest*. The motiongram shows the overall movements of the head (top), hands and arms (middle), and legs and feet (bottom). The movements correlate clearly with the resultant sound.

stated they found performing with markers on their body and wearing a cap with head markers not uncomfortable, and the recording sessions lasted for approximately 45 minutes including some trial recordings. The performances were recorded using a Qualisys infrared motion capture system consisting of 9 Oqus 300 cameras placed in a circle around the pianists performing on a Yamaha P-140 digital piano, an instrument with a low front that allows line of sight to all the cameras, i.e. no occlusion. Position data with a precision of 0.3 millimeters for all markers were recorded at a frame rate of 100 Hz together with sound and video into the *Qualisys Track Manager (QTM)* software. This was then exported to *Matlab*, and various position, velocity, and acceleration data was then calculated and correlated with the tone onset data from spectrograms of the sound files using the *MIR* and *MoCap* Toolboxes (Lartillot & Toiviainen 2007, Toiviainen 2008).

In processing and representing this data, there is the issue of the temporal resolution, i.e. which

timescale to focus on. Zooming into very local timescales will be interesting when studying details of expressivity and articulation (e.g. *staccato* with quick, jerky upward movement of hand, wrist, elbow, as opposed to *legato* with more slow, smooth upward movement of hand, wrist, elbow), but in our present paper on coarticulation with respect to chunking, our focus is on the meso-level timescale (cf. section 3 above). We are thus focusing on coarticulatory movements in approximately the 0.5 to 5 seconds range, and in the following figures the movements of the pianists’ effectors will be plotted in relation to score and spectrogram representations of musical excerpts so as to best represent such meso-level features.

Although we retain the data from all the markers mentioned above for further processing and analysis, we have in this paper selected the wrist, elbow, and shoulder markers for processing and representation as a workable compromise in view of avoiding the cluttered plottings of including the recorded data from all markers. Again to avoid cluttering, we have selected plotting movement trajectories on

accessed January 2010

the horizontal plane, i.e. along the keyboard, because this should indicate coarticulatory movement with respect to finger position for the piano keys to be played. We have also plotted the absolute velocities (i.e. the vector length of the first derivative of the position), smoothed with a Savitzky-Golay FIR filter using a 99 frame window to make the graph more clearly visible. Although the velocity curves is a simplification of the original movement, the representation is interesting with regard to the starting and ending of a coarticulatory chunk: interpreting the points of minimal velocity as *goal-postures* (cf. the above-mentioned discussion and reference to Rosenbaum et al. (2007)), we take this to mean that a typical coarticulatory movement starts from some initial position, accelerates and then decelerates along a trajectory, ending up in a new goal-posture, followed by a new chunk of acceleration and deceleration, a new goal-posture, etc.

Since presenting trajectory and velocity data for all the recorded excerpts would not be possible in this paper, we have instead chosen to present two excerpts that illustrate typical features of chunking by coarticulation. Figure 2 displays the score, spectrogram and the hands, wrists, and elbow joint movements and velocities of one of the pianists performing the first two measures of J. S. Bach's *Fugue in D major, Well-Tempered Clavier I, BWV 850*. The rushes of 32nd notes as well the mordents, pralltrillers, and punctuations in this excerpt calls for a rather well-articulated and 'declamatory' or 'recitative-like' style of playing, something that is reflected in the coarticulatory movements here. The movements of the wrists, elbows, and shoulders indicate optimal ergonomic positioning for the finger actions, and the velocity curves show the acceleration and deceleration in relation to the goal-postures, here in the form of sustained tones (i.e. after ornamental movements are finished).

Although informal observations of pianists' playing style tells us that there are probably great variations in the amount of movement made in performing the same musical excerpts, it could be predicted from the general principles of coarticulation (i.e. biomechanical and motor control constraints of human body movement) that there will be a minimum of similar-

ity in the coarticulatory movement trajectories, and also because of the constraints of distance travelled in a given time (dictated by the rhythm of the music), there also should be some similarity in the velocity curves. Figures 3 and 4 show the scores, spectrograms, and the hands, wrists, and elbow joints trajectories and velocities of the two pianists performing the first 8 measures (with the upbeat measure) of the opening of the last movement of L. v. Beethoven's *Piano Sonata nr. 17 Op. 31 no. 2 in d-minor, The Tempest*. Again there is a coarticulatory inclusion of the tone events in more superordinate hands, elbows, and shoulder trajectories, (variably so) forming cyclical movements in the right and left hands: the right hand with an upbeat movement leading to the downbeat of each measure, and at the downbeat the left hand makes an upward movement ending on the fourth sixteenth note of each measure where the right hand simultaneously starts again the upbeat movement, a pattern that is repeated throughout this excerpt. Here we see again the velocity curves between the goal-postures (at the downbeat for the right arm, fourth sixteenth note for the left arm).

The position and velocity data curves for the two pianists movements also show differences in the shape of these movements. This variation between performers is a general point of coarticulation, as remarked by Engel et al. (1997):

Such anticipatory modifications of the movement were found to be variable from piece to piece and from subject to subject, and it is likely that this phenomenon reflects the task demands of the different pieces and the capabilities and physical constraints (such as fingerspan) of each subject. In fact, we observed a continuum in the extent of anticipatory modifications of finger and hand kinematics, ranging from as much as 500 ms in advance of the time of the last common note to about 15 ms after the time of key-on of this note. (Engel et al. 1997, 198)

Such variations in coarticulatory movement should also be related to the above-mentioned *sound-*

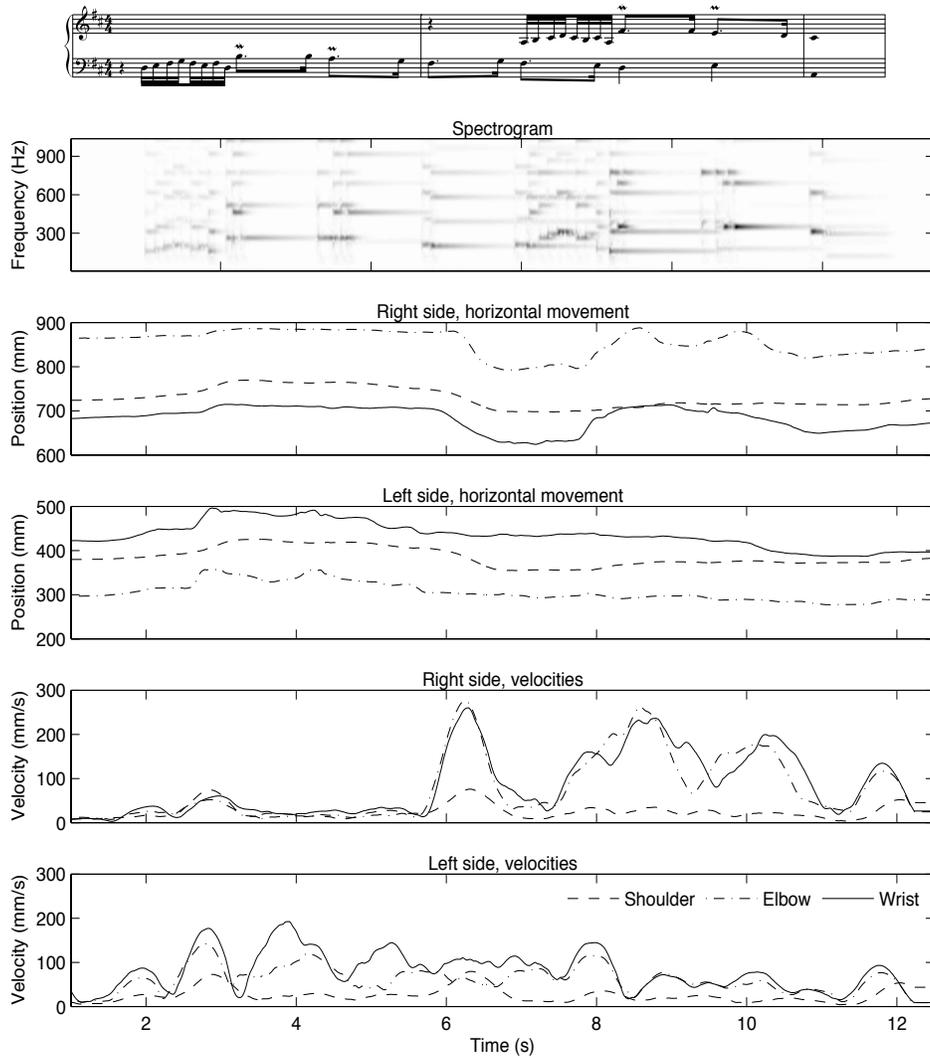


Figure 2: Score and spectrogram of the first two measures of J. S. Bach's *Fugue in D major, Well-Tempered Clavier I, BWV 850* together with plots of the horizontal positions (i.e. along the piano keyboard) and absolute velocities of the left and right wrists, elbows, and shoulders (legend in the bottom subplot). The peak in the velocity of the right hand during the first beat of the second measure (where there is a quarter note rest) is due to the pianist lifting the right hand, which has been resting on the lap during the first measure, up to the keyboard.

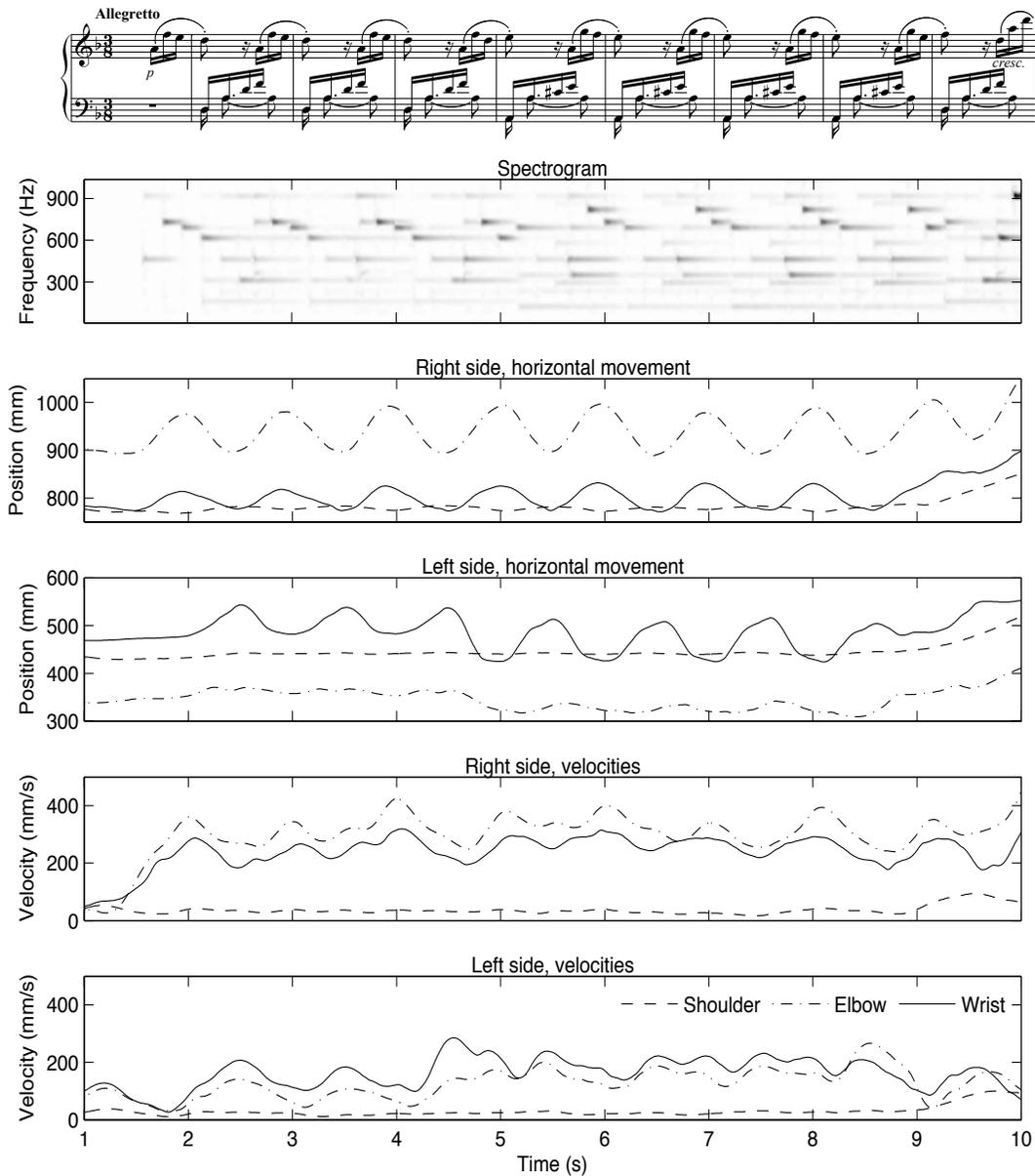


Figure 3: Score and spectrogram of the first 8 measures (with the upbeat measure) of the opening of the last movement of L. v. Beethoven's *Piano Sonata nr. 17 Op. 31 no. 2 in d-minor, The Tempest* together with plots of the horizontal positions (along keyboard) and absolute velocities of the left and right wrists, elbows, and shoulders of pianist A (legend in bottom figure).

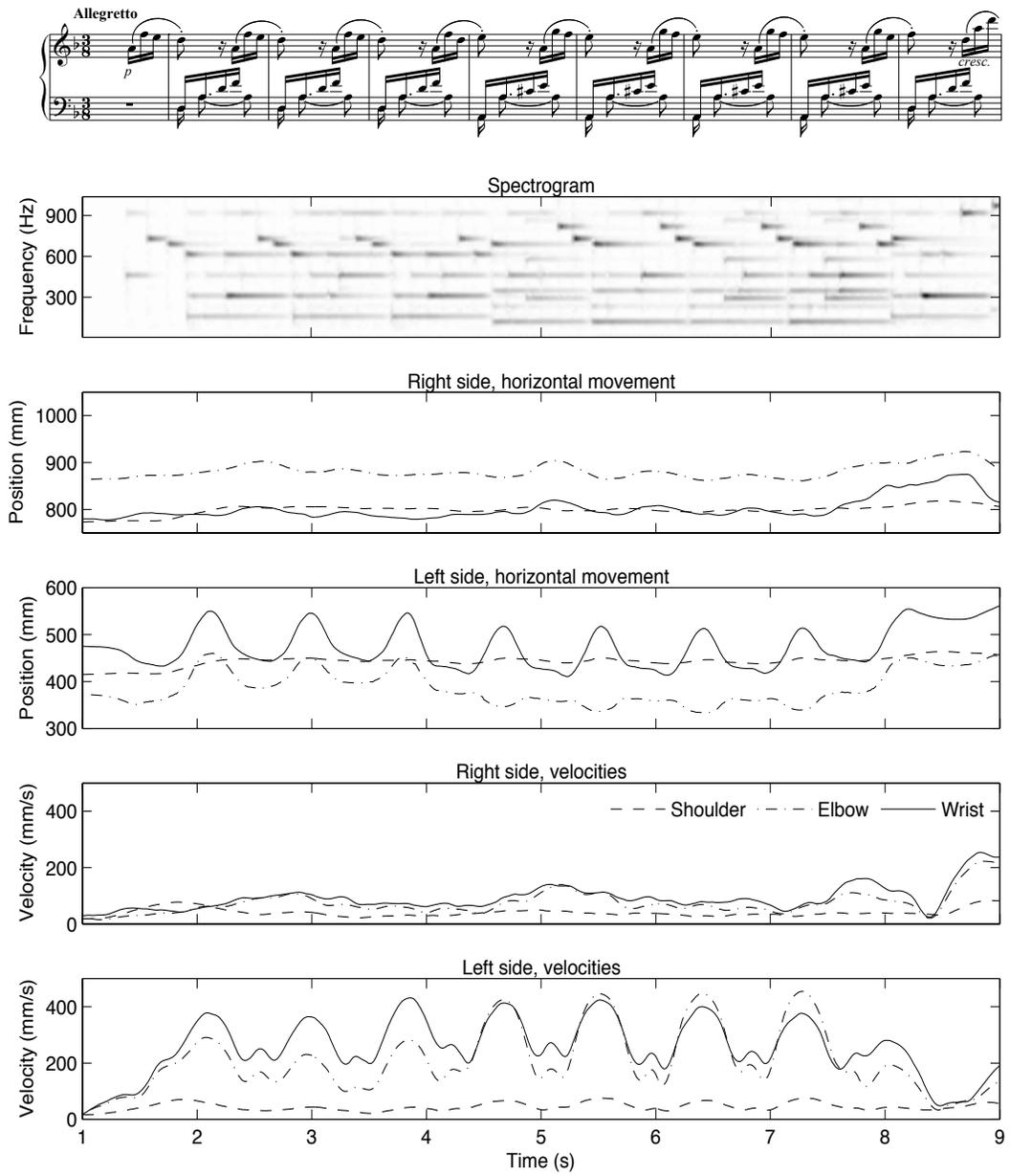


Figure 4: Score and spectrogram of the first 8 measures (with the upbeat measure) of the opening of the last movement of L. v. Beethoven's *Piano Sonata nr. 17 Op. 31 no. 2 in d-minor, The Tempest* together with plots of the horizontal positions (along keyboard) and absolute velocities of the left and right wrists, elbows, and shoulders of pianist B.

facilitating gestures, gestures that may serve different purposes such as facilitating sound-production, avoiding fatigue and strain injury, but also to help shape the music, i.e. the rhythmical organization and the expressive phrasing (Wanderley et al. 2005, Jensenius et al. 2010).

Although there are differences in the velocity curves of two performances of the same Beethoven excerpt in Figures 3 and 4 (pianist A has a more active right arm than pianist B, and pianist B has a more active left arm than pianist A), the overall pattern of an undulating velocity is similar. With the basic understanding of an action chunk as something that starts at one moment in time and ends at a later moment in time, the velocity curve should be an indicator of chunking: if the velocity is minimal, that would indicate some kind of resting position, and if velocity is increasing, is high, or is decreasing, that would indicate there is a movement going on. Furthermore, it must be assumed that these velocity curves are the result of effort, hence is the visible result of an underlying muscle activity. Points of little velocity could be understood as points of relatively little effort and hence as goal-postures and boundaries of the coarticulated chunks. Patterns in velocity and acceleration and the assumed underlying muscle activity could also be related to rhythm perception in general, as remarked by (Luck & Sloboda 2009) in connection with the perception of beats from conductors' gestures. In future research on chunking by coarticulation we shall for this reason try to record EMG data together with motion capture data and look at the relationships between these two data sets.

7 Coarticulation in music perception

Coarticulation entails a contextual smearing both in action and in perception. In speech perception, the contextual smearing by coarticulation may be seen as a problem when learning new languages or when trying to make machine systems for speech transcription. On the other hand, coarticulation may be seen simply as a natural phenomenon in speech, in line

with evolutionary developed features of both the articulatory apparatus and the cognitive faculties for language perception. A similar view of coarticulation could be adopted for music, meaning that coarticulation is the result of the combined constraints of the physics of musical instruments and the sound-producing actions, and has consequences for musical features and music perception: melodic, rhythmic, textural, timbral, etc. features of music can all be regarded as shaped by coarticulation.

However, there is clearly a need for more research on the perceptual effects of coarticulation, both on the chunk level and on sub-chunk level, and in particular with regards to tone-to-tone-transitions. Such tone-to-tone-transitions are manifest in most (if not all) instrumental and vocal performances. In fact, the practicing of smooth transitions between tones is a major activity in training on most instruments, e.g. woodwind and brass, as well as string instruments, and such practicing could actually be understood as practicing coarticulation. Perceptually interesting simulations of contextual smearing in tone transitions and hence of coarticulation can be made with so-called *diphone synthesis*, where it is possible on the basis of spectral analysis and interpolation between the individual sounds to make smooth transitions between them. On the other hand, the lack of coarticulation in digital instruments is one of the reasons they often sound 'unnatural.' In the case of the piano (and other instruments with fairly long decay envelopes), there is of course the contextual smearing caused by overlapping decay envelopes of the tones, in particular with the use of the sustain pedal. But as argued in the previous section, there is also in the case of the piano a clear need for, and perceptual effects of, coarticulatory movement because of the combined instrument-related, biomechanical, and motor control constraints.

To better understand coarticulation in music, we also need more knowledge about temporal integration in the perception of musical sounds, along the lines suggested for temporal integration in the perception of speech sounds by Grossberg & Myers (2000). With the metaphor of 'resonance,' they suggest that sequentially occurring sound features are kept in memory for the duration of a chunk and that there is

a holistic perception of the chunk where the sequentially occurring elements may mutually influence each other, hence their idea of ‘backward effects.’ Interestingly, a similar idea was suggested by Schaeffer with his notion of *temporal anamorphosis* or ‘time warping’ in sound perception, meaning that the entire sonic object with all the sequentially occurring elements may be perceived holistically as one unit, and not as a string of isolated events (Schaeffer 1966, 1998, Godøy 2006).

8 Conclusion

Needless to say, there is a long way to go in understanding coarticulation in music. Yet we already see clear indications of chunking by coarticulation at work in both the production and perception of music. In fact, coarticulation could be regarded as primordial in music, as something that should come before our focus on singular tones, although we may tend to forget this with our habitual symbolic approaches in music research (i.e. by Western notation). A better understanding of coarticulation in music could then teach us more about how experiences of continuity may emerge in music.

Acknowledgments

Many thanks to the two pianists for coming to our lab and making the recordings, and many thanks to the reviewers for their constructive and very helpful comments.

References

- Bangert, M. & Altenmüller, E. O. (2003), ‘Mapping perception to action in piano practice: a longitudinal dc-eeeg study’, *BMC Neuroscience* **4**(26).
- Berthoz, A. (1997), *Le sens du mouvement*, Paris: Odile Jacob.
- Bregman, A. S. (1990), *Auditory Scene Analysis: The Perceptual Organization of Sound*, The MIT Press, Cambridge, MA.
- Camurri, A. & Moeslund, T. B. (2010), Visual gesture recognition. from motion tracking to expressive gesture, in R. I. Godøy & M. Leman, eds, ‘Musical Gestures: Sound, Movement, and Meaning’, Routledge, New York, pp. 238–263.
- Clarke, E. F. & Davidson, J. W. (1998), The body in performance, in ‘Composition-performance reception’, Ashgate, Aldershot, pp. 74–92.
- Cohen, M. M. & Massaro, D. W. (1993), Modeling coarticulation in synthetic visual speech, in N. Thalmann & D. Thalmann, eds, ‘Models and Techniques in Computer Animation’, Springer-Verlag, Tokyo.
- Cope, D. (1991), *Computers and Musical Style*, A-R Editions, Inc., Madison, Wisconsin.
- Dahl, S. (2006), Movements and analysis of drumming, in E. Altenmüller, M. Wiesendanger & J. Kesselring, eds, ‘Music, Motor Control and the Brain’, Oxford University Press, Oxford, pp. 125–138.
- Davidson, J. (1993), ‘Visual perception and performance manner in the movements of solo musicians’, *Psychology of Music* **21**, 103–113.
- Drake, C. & Palmer, C. (2000), ‘Skill acquisition in music performance: relations between planning and temporal control’, *Cognition* **74**(1), 1–32.
- Eitan, Z. & Granot, R. (2006), ‘How music moves: Musical parameters and listeners’ images of motion’, *Music Perception* **23**(3), 221–247.
- Eitan, Z. & Granot, R. (2008), ‘Growing oranges on mozart’s apple tree: ‘inner form’ and aesthetic judgment’, *Music Perception* **25**(5), 397–417.
- Elliott, D., Helsen, W. F. & Chua, R. (2001), ‘A century later: Woodworth’s (1899) two-component model of goal-directed aiming’, *Psychological Bulletin* **127**(3), 342–357.
- Engel, K., Flanders, M. & Soechting, J. (1997), ‘Anticipatory and sequential motor control in piano playing’, *Experimental Brain Research* **113**(2), 189–199.

- Farnetani, E. & Recasens, D. (1999), Coarticulation models in recent speech production theories, *in* W. Hardcastle & N. Hewlett, eds, 'Coarticulation: theory, data and techniques', Cambridge University Press, Cambridge, pp. 31–65.
- Fraisse, P. (1982), Rhythm and tempo, *in* D. Deutsch, ed., 'The Psychology of Music', first edn, Academic Press, New York, pp. 149–180.
- Galantucci, B., Fowler, C. A. & Turvey, M. T. (2006), 'The motor theory of speech perception reviewed', *Psychonomic Bulletin & Review* **13**(3), 361–377.
- Gallese, V. & Lakoff, G. (2005), 'The brain's concepts: The role of the sensory-motor system in conceptual knowledge', *Cognitive Neuropsychology* **22**(3/4), 455–479.
- Gallese, V. & Metzinger, T. (2003), Motor ontology: The representational reality of goals, actions and selves, *in* 16, ed., 'Philosophical Psychology', Vol. 3.
- Godøy, R. I. (2003), 'Motor-mimetic music cognition', *Leonardo* **36**(4), 317–319.
- 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. Hamburger Jahrbuch für Musikwissenschaft', Vol. 24, Peter Lang, Vienna, pp. 117–132.
- Godøy, R. I. (2010), Thinking now-points in music-related movement, *in* R. Bader, C. Neuhaus & U. Morgenstern, eds, 'Studies in Systematic Musicology', Peter Lang, Vienna, pp. 241–258.
- Godøy, R. I., Haga, E. & Jensenius, A. R. (2006a), Exploring music-related gestures by sound-tracing - a preliminary study, *in* K. Ng, ed., 'Proceedings of the COST287-ConGAS 2nd International Symposium on Gesture Interfaces for Multimedia Systems', Leeds, pp. 27–33.
- Godøy, R. I., Haga, E. & Jensenius, A. R. (2006b), Playing 'air instruments': Mimicry of sound-producing gestures by novices and experts, *in* S. Gibet, N. Courty & J.-F. Kamp, eds, 'Gesture in Human-Computer Interaction and Simulation, 6th International Gesture Workshop', Vol. LNAI 3881, Springer-Verlag, Berlin Heidelberg, pp. 256–267.
- Godøy, R. I., Jensenius, A. R. & Nymoen, K. (2008), Production and perception of goal-points and coarticulations in music, *in* 'ASA-EAA Conference', Paris, France.
- Godøy, R. I., Jensenius, A. R. & Nymoen, K. (2009), Coarticulation of sound and movement in music, *in* 'Proceedings of ESCOM 2009: 7th Triennial Conference of the European Society for the Cognitive Sciences of Music', Jyväskylä, Finland.
- Godøy, R. I. & Leman, M. (2010), *Musical Gestures: Sound, Movement, and Meaning*, Routledge, New York.
- Grossberg, S. & Myers, C. W. (2000), 'The resonant dynamics of speech perception: Interword integration and duration-dependent backward effects', *Psychological Review* **107**(4), 735–767.
- Haga, E. (2008), Correspondences between music and body movement, PhD thesis, University of Oslo.
- 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. J. & Hewlett, N., eds (1999), *Coarticulation: Theory, Data, and Techniques*, Cambridge University Press, Cambridge.
- Haueisen, J. & Knösche, T. R. (2001), 'Involuntary motor activity in pianists evoked by music perception', *Journal of Cognitive Neuroscience* **13**(6), 786–792.
- Hickok, G., Buchsbaum, B., Humphries, C. & Muf-tler, T. (2003), 'Auditory-motor interaction revealed by fMRI: Speech, music, and working memory', *Area Spt. Journal of Cognitive Neuroscience* **15**(5), 673–682.

- Husserl, E. (1991), *On the Phenomenology of the Consciousness of Internal Time, 1893 -1917*, English translation by John Barnett Brough Kluwer, Academic Publishers, Dordrecht.
- Jabusch, H.-C. (2006), Movement analysis in pianists, in E. Altenmüller, M. Wiesendanger & J. Kesselring, eds, 'Music, Motor Control and the Brain', Oxford University Press, Oxford, pp. 91–108.
- 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. & Godøy, R. I. (2008), 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.
- Jensenius, A. R., Wanderley, M. M., Godøy, R. I. & Leman, M. (2010), Musical gestures: concepts and methods in research, in R. I. Godøy & M. Leman, eds, 'Musical Gestures: Sound, Movement, and Meaning', Routledge, New York, pp. 12–35.
- Jerde, T. E., Santello, M., Flanders, M. & Soechting, J. F. (2006), Hand movements and musical performance, in E. Altenmüller, M. Wiesendanger & J. Kesselring, eds, 'Music, Motor Control and the Brain', Oxford University Press, Oxford, pp. 79–90.
- Jerde, T. E., Soechting, J. F. & Flanders, M. (2003), 'Coarticulation in fluent fingerspelling', *The Journal of Neuroscience* **23**(6), 2383–2393.
- Kandel, S., Boe, L.-J. & Orliaguet, J.-P. (1993), Visual detection of coarticulatory anticipation or . . . guessing what has not yet been written, in 'Virtual Reality Annual International Symposium, 1993., 1993 IEEE', pp. 148–154.
- Kohler, E., Keysers, C., Umiltà, M. A., Fogassi, L., Gallese, V. & Rizzolatti, G. (2002), 'Hearing sounds, understanding actions: Action representation in mirror neurons', *Science* **297**(5582), 846–848.
- Lartillot, O. & Toiviainen, P. (2007), MIR in Matlab (II): A toolbox for musical feature extraction from audio, in 'International Conference on Music Information Retrieval', pp. 237–244.
- Lashley, K. S. (1951), The problem of serial order in behavior, in L. Jeffress, ed., 'Cerebral Mechanisms in Behavior', Wiley, New York, pp. 112–131.
- Leman, M. (2008), *Embodied Music Cognition and Mediation Technology*, The MIT Press, Cambridge, MA.
- Lieberman, A. M. & Mattingly, I. G. (1985), 'The motor theory of speech perception revised', *Cognition* **21**, 1–36.
- Luck, G. & Sloboda, J. (2009), 'Spatio-Temporal Cues for Visually Mediated Synchronization', *Music Perception* **26**(5), 465–473.
- Miller, G. A. (1956), 'The magical number seven, plus or minus two: Some limits on our capacity for processing information', *Psychological Review* **63**, 81–97.
- Moore, B. C. J., ed. (1995), *Hearing*, Academic Press, San Diego.
- Palmer, C. (2006), The nature of memory for music performance skills, in E. Altenmüller, M. Wiesendanger & J. Kesselring, eds, 'Music, Motor Control and the Brain', Oxford University Press, Oxford, pp. 39–53.
- Perkell, J. (2007), Sensory goals and control mechanisms for phonemic articulations, in 'Proceedings of the 16th International Congress of Phonetic Sciences', Saarbrücken.
- Pöppel, E. (1997), 'A hierarchical model of time perception', *Trends in Cognitive Science* **1**(2), 56–61.
- Rasamimanana, N. & Bevilacqua, F. (2008), 'Effort-based analysis of bowing movements: evidence of anticipation effects', *Journal of New Music Research* **37**(4), 339–351.
- Rosenbaum, D. A. (1991), *Human Motor Control*, Academic Press, San Diego.

- Rosenbaum, D. A., 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.
- Schaeffer, P. (1966), *Traité des objets musicaux*, Paris: Editions du Seuil.
- Schaeffer, P. (1998), *Solfège de l’objet sonore*, (with sound examples by G. Reibel & B. Ferreyra, first published in 1967), INA/GRM, Paris.
- Schleidt, M. & Kien, J. (1997), ‘Segmentation in behavior and what it can tell us about brain function’, *Human Nature* **8**(1), 77–111.
- Snyder, B. (2000), *Music and Memory: An Introduction*, The MIT Press, Cambridge, MA.
- Toiviainen, P. (2008), Mocap tool box manual, Technical report, University of Jyväskylä.
- Vines, B., Krumhansl, C., Wanderley, M. & Levitin, D. (2005), ‘Cross-modal interactions in the perception of musical performance’, *Cognition* **101**, 80–113.
- Wanderley, M. M. & Battier, M., eds (2000), *Trends in Gestural Control of Music [CD-ROM]*, IRCAM – Centre Pompidou, Paris.
- Wanderley, M. M., Vines, B. W., Middleton, N., McKay, C. & Hatch, W. (2005), ‘The musical significance of clarinetists’ ancillary gestures: An exploration of the field’, *Journal of New Music Research* **34**(1), 97–113.
- Wiesendanger, M., B. A. & Kazennikov, O. (2006), Fingering and bowing in violinists: A motor control approach, in E. Altenmüller, M. Wiesendanger & J. Kesselring, eds, ‘Music, Motor Control and the Brain’, Oxford University Press, Oxford, pp. 109–123.
- Wilson, M. & Knoblich, G. (2005), ‘The case for motor involvement in perceiving conspecifics’, *Psychological Bulletin* **1**(3), 460–473.