Geometry and effort in gestural renderings of musical sound

Rolf Inge Godøy

Department of Musicology, University of Oslo, P.B. 1017 Blindern, N-0315 Oslo, Norway

r.i.godoy@imv.uio.no

Abstract. As may be seen at concerts and in various everyday listening situations, people often make spontaneous gestures when listening to music. We believe these gestures are interesting to study because they may reveal important features of musical experience. In particular, hand movements may give us information on what features are perceived as salient by listeners. Based on various current ideas on embodied cognition, the aim of this paper is to argue that gestures are integral to music perception, and to present research in support of this. A conceptual model of separating geometry and effort is presented in order to better understand the variety of music-related gestures we may observe, leading up to some ideas on how to apply this conceptual model in present and future research.

Keywords: Music, perception, gestures, geometry, effort, perception-action, key-postures, movement

1. Introduction

In our research on music-related gestures [1] we have had a particular focus on the spontaneous gestures that listeners make to musical sound. This has been motivated by the belief that perception and cognition of musical sound is intimately linked with mental images of movement, and that a process of incessant motor imagery is running in parallel with listening to, or even just imagining, musical sound. We have called this *motormimetic cognition* of musical sound [2], and see evidence for this in a number of research findings as well as in our own observation studies. Furthermore, we believe *hand movements* have a privileged role in motormimetic cognition of musical sound, and that these hand movements may trace the *geometry* (i.e. elements such as pitch contours, pitch spread, rhythmical patterns, textures, and even timbral features), as well as convey sensations of *effort*, of musical sound, hence the focus in this paper on *geometry and effort in the gestural renderings of musical sound*.

There are many different gestures that may be associated with music (see [3] and [4] for overviews). Using the Gibsonian concept of *affordance* [5], we can thus speak of rich *ges*-*tural affordances* of musical sound. For practical purposes we can in this paper think of two main categories of music-related gestures, namely *sound-producing* gestures (such as hit-ting, stoking, bowing, etc.) and *sound-accompanying* gestures (such as dancing, marching, or making various movements to the music), as well as several sub-categories of these. The distinction between these two main categories as well as their sub-categories may not always be so clear, e.g. musicians may make gestures in performance that are probably not strictly necessary for producing sound, but may be useful for reasons of motor control or

physiological comfort, or may have communicative functions towards other musicians or the audience.

In order to know more about gestural renderings of musical sounds by listeners, we have carried out a series of observation studies where we have asked listeners with different levels of musical training to make spontaneous hand movements to musical excerpts. In these studies we have proceeded from giving the listeners rather well defined tasks with limited gestural affordances onto progressively more open tasks with quite rich gestural affordances. This was done by proceeding from studies of *air-instrument performances* where listeners were asked to make sound-producing movements [6], onto what we have called *sound-tracing* studies where listeners were asked to draw (on a digital tablet) the gestures they spontaneously associated with the musical excerpts and where the musical excerpts were quite restricted as to their number of salient features [7], and finally onto what we have called *free dance gestures* with more complex, multi-feature musical excerpts and rather general instructions to listeners about making spontaneous gestural renderings based on what they perceived as the most salient features of the musical excerpts [8].

The idea of gestural rendering of musical sound is based on a large body of research ranging from classical motor theory of perception [9] to more recent theories of motor involvement in perception in general [10], and more specifically in audio perception [11], as well as in music-related tasks in particular [12], research that converge in an understanding of motor cognition as integral to most areas of perception and cognition.

Obviously, auditory-motor couplings as well as the capacity to render and/or imitate sound is not restricted to hand movements, as is evident from vocal imitation of both non-musical and musical sound (e.g. so-called *beat-boxing* in hip-hop and other music and *scat singing* in jazz). But the focus on hand movements in our case is based not only on innumerable informal observations of listeners making hand movements to musical sound, but also on the belief that hand movements have a privileged role from an evolutionary point of view [13] and from a general gesture-cognitive point of view [14]. Furthermore, we believe that a listener through a process of translation by the principle of *motor equivalence* [15] may switch from one set of effectors to another, translating various sound features to gestures and hence revealing more amodal gestural images of musical sound.

It seems quite clear that even novices can make gestures that reflect reasonably well what is going on in the music when asked to imitate sound-producing actions, although experts tend to make more detailed renderings as reported in [6]. Also when listeners were asked to draw gestures they felt reflected the musical excerpt they heard, i.e. in what we have called *sound-tracing* studies, there was reasonable agreements as long as the excerpts did not have more than one or two prominent features, e.g. an ascending pitch contour or an ascending pitch contour combined with various ornamental ripples, and greater disagreement when the number of concurrent prominent features was increased, e.g. excerpts with several concurrent textural elements [7]. A subsequent so-called interrater study (i.e. a study of agreement/disagreement in judgment) of the resultant sound-tracings seems to have confirmed these agreements and/or disagreements [16]. In the case of 3-dimensional bi-manual movements to sounds, i.e. *free dance gestures*, we got even more varied results, something that we would expect given the greater choice of movement trajectories and feature focus [8, 16].

Rather than despair because of these increasingly divergent and also often rather approximate gestural renderings of musical sound, we shall in the following see how the elements of geometry and effort may be understood as intrinsic to the *perception-action cycle* spontaneously at work in musical experience, and furthermore try to see how gestural renderings of musical sound may be understood as a means for intentional focus in listening, and may even be put to active use in the exploration of musical sound.

It is generally accepted that music is a multidimensional phenomenon in the sense that music has elements such as rhythm, tempo, intensity (often referred to as dynamics), pitch, melody, accompaniment, harmony, timbre, texture, etc., and that these elements in turn may be differentiated into a number of sub-elements. This is one of the reasons for the abovementioned rich gestural affordances of musical sound, as listeners may attend to, and gesturally render, any single or any selection of such musical elements. Also, elements that from a music theory perspective may be thought of as separate, may be fused in actual perception, such as in the well known interdependence of perceived intensity and timbre. This dimensional fusion may even extend to dimensions that 'are not really there', i.e. we may see what the authors of [17] have termed a 'spill over' effect, e.g. that a *crescendo* may also be perceived as an *accelerando* by some listeners even though the tempo was constant.

Although the elements of geometry and effort are inseparable in the sense that we can not have images of geometry (e.g. pitch change, timbre change, etc.) without some image of movement and hence of effort, and conversely, can hardly have images of effort in music without images of movement in space and hence of geometry, it is strategically convenient to separate these elements here in order to be able to better differentiate what features listeners focus on in various gestural renderings of musical sound, as well as to be able to appreciate the sometimes seemingly divergent gestural renderings of musical sound that we may observe: Music is complex and multi-faceted and listeners' attention to features will often vary, so it goes without saying that we may observe seemingly divergent gestural renderings, and our task here is to try to summarize which features are most commonly shared by listeners and which features seem to be more variably rendered by listeners.

It should also be noted that we have initially had a qualitative approach in our studies of gestural renderings of musical sound, meaning that we have proceeded in a top-down manner from overall features to progressively finer sub-features, when analyzing our video material. However, we have also used various sensor technologies and software tools in our studies, as for instance may be seen in figures 1-3. Currently we are using a multi-sensor setup for motion capture together with a multi-stream data storage and retrieval scheme for this motion capture data [18], something we believe in the future will give us more detail data on the various features in gestural renderings of musical sound. But as a point of method, we believe that we regardless technologies will need to have a conceptual apparatus for understanding what we are observing, something I shall try to present in the following sections of this paper.

2. Geometry

Musical sound seems to be a good transducer of various features related to the geometric layout of musical instruments, i.e. of real-world 3-D Euclidian space, such as the left-to-right, low-to-high pitch layout of the piano, or such as the spatial layout of drums. Notions of geometry for pitch-space or for relative-approximate pitch (in the case of non-tuned instruments) are probably learned, but to what extent this left-to-right scheme for pitch ordering is valid across different cultures is uncertain (see [17] for a discussion). In the case of air-piano playing, the gestural renderings seem to reflect quite well not only the pitch space of melodic movement on the piano, but also the relative position and spread of pitches, i.e. deep tones were rendered to the left, high pitches to the right, dense textures with both hands close together, and spread textures with hands spread out across the imaginary keyboard. An example of this may be seen in figure 1 where the subject is giving a gestural rendering of the opening passage of Scriabin's 5^{th} *Piano Sonata* where there is a rapid and rather loud passage running from deep tones up to high tones, and where the texture is quite dense throughout the passage (see [6] for details on the setup and method of this study). The same musical excerpt may be seen rendered in a sound-tracing study in figure 2 where we can see a similar rendering of the pitch profile of this passage (see [7] for details).



Fig. 1. Air-piano rendering of the opening of Scriabin's 5th Piano Sonata, here with four key-frames running left to right, top to bottom.

In the subsequent interrater study of our sound-tracing study [16], 20 subjects with different levels of musical training were asked to rate the correspondences between 18 of the sounds and their respective sound-tracing images from the original sound-tracing study (which consisted of 50 sounds and images [7]) according to a forced choice task of match or mismatch. The subjects in this study were presented with displays of the nine drawings of the sounds (like in figure 2) while the sounds were played back, and the method of forced choice was used in order to collect the spontaneous reactions of the subjects to the correspondences (or lack thereof) between the sounds and the gestural trajectories. The nine drawings of each of these sounds were then arranged according to these match-mismatch judgments, and also the judgments of all the subjects (see [16] for details on this). Briefly stated, the overall agreement for all the sounds-images and subjects in this study was not very strong, although seems to be above chance (but this is open to discussion as to what criteria for above chance agreements are applied). However, looking at the results for individual sounds and their respective tracings, there seems to be some clear tendencies:



Fig. 2. Sound-tracing renderings of the opening of Scriabin's 5th *Piano Sonata* by nine different subjects, time running from left to right in all tracings. Notice that with the exception of the left and middle rendering of the first row, there seems to be more or less agreement about the pitch contour.

Pitch contours are generally more agreed on than single pitches, i.e. pitch-wise stationary, but otherwise (timbrally, dynamically) evolving sounds are not well agreed on. For instance, the sound of a single piano tone was by some of the original sound tracers variously rendered as a straight horizontal line, as a curve (perhaps alluding to the envelope of the tone) or as a single point (perhaps alluding to the singular impact of the finger on the key). Compared with this, a sequence of several piano tones (as in figure 2) resulted in a much higher degree of agreement. In a different study [19], there seems to be a similar agreement on the perception of pitch contours as long as they are fairly simple, i.e. as long as they do not have more than one or very few directions in their contours.

But also other musical elements such as intensity and timbral changes may be conceived as geometric features, something that partly seems to be reflected in the match-mismatch study [16]. Although we may assume that listeners have some experiences of physiological links with intensity and/or timbral changes, as in the case of a *crescendo* with increased amplitude in the hand movements, or in the case of timbral changes related to experiences of changing shape of the vocal tract, or experiences of seeing musicians move mutes, change bow positions, etc., the sound-tracing renderings of these intensity and/or timbral evolutions seem to vary a good deal more than those of pitch contours [7]. In the subsequent interrater study [16], the subjects seemed to agree more on tracings that depicted some kind of envelope or motion than on tracings that only suggested one attack point of the sounds, as in the case of a single cymbal sound or a single trumpet sound. This variety in the sound-tracings of timbral evolutions of sounds can probably be understood as a result of the limitations of a 2D rendering on the digital tablet, i.e. several participants in the original sound-tracing study expressed frustration at being asked to draw timbral evolutions of sounds on a flat surface [7], hence, this could be seen as a source of error both in the original sound-tracing and in the subsequent interrater study.

From all our studies, we believe we may conclude that the geometry of positions, spread, and trajectories in pitch-space is well rendered by hand movements, including elements such as pitch-contours and textural elements, although we may see differences in detail, or in what we could call frequential resolution of the gestures. Gestural renderings of musical excerpts with clearer textural-timbral fluctuations, such as the 'grainy' quality of iterative sounds, suggests that also 'micro features' of sounds are reflected in the gestures, something that has useful applications in sound research [20].

Naturally, with increasing richness of features, e.g. with musical excerpts with more composite textures, there was also increasing diversity in the geometry of the gestural renderings of the musical excerpts, something that we would understand as differences in focus of listening. However, in several detailed annotation analyses of our free dance studies in [16], it seems quite clear that we rarely found gestural renderings that did not at all correspond with some features or salient events in the music. In particular, there seemed in these free dance studies to be good agreements in the gestural renderings at the level of chunks, meaning the geometry of both the overall melodic and overall textural shape.

3. Effort

Musical sound also seems to be a good transducer of the overall activation level in soundproducing gestures: The density, speed, and force of events are well rendered, e.g. rapid passages are rapidly rendered, loud events are rendered by high amplitude gestures, etc., as evident both from quantity of motion estimations and detailed annotations. This seemed to be the case in all our observation studies [6, 7, 8, 16], and although we could see individual variations in the activation levels, we could not see gestural renderings that were clearly contradictory to the overall nature of the music, e.g. we could not see agitated movements to calm music, or calm movements to agitated music.

Furthermore, there seemed to be a fairly good discrimination between sounds with different types of excitatory actions, actions that we believe are based on biomechanical and neurocognitive constraints, and which we have classified as follows:

- Sustained, meaning protracted sounds requiring continuous excitatory effort such as in bowing or blowing. But whereas there is a more or less direct rendering of the events in the case of short, distinct sounds and rapid passages, for sustained sounds there was a tendency to 'fill in', i.e. to make long, slow, and curved gestures. This may be seen in figure 3 where the dancer makes a slow upward gesture with both hands to the sustained sound of the opening of the Lento movement of Ligeti's Ten Pieces for Wind Quintet. The sustained sound does not change in pitch or loudness so as to suggest an upward gesture, yet its sustained character does suggest a continuous movement, and we could very well imagine different directions of such continuous movement to this particular excerpt without any one of them being perceived as in conflict with the effort character of the music.
- *Impulsive*, meaning percussive or other discontinuously excited sounds, also including rapidly rendered groups of tones such as in short glissandi or ornaments.

• *Iterative*, meaning rapid repetition of onsets as well as rapid modulatory movements, e.g. rapid vibrato or tremolo.



Fig. 3. Gestural rendering of five seconds of sustained sound at the opening of the Lento movement of Ligeti's Ten Pieces for Wind Quintet. The dancer makes a slow upward gesture with the arms, starting from arms aligned with the hips and ending up with the arms stretched out horizontally.

These sound types and their associated modes of excitations are quite distinct, requiring quite distinct types of effort and attention. The sustained and impulsive actions seem to be mutually exclusive, but iterative actions may have elements of both sustained and impulsive actions, e.g. in the playing of washboard, maracas, and even drum rolls where the stream of individual strokes is so fast as to require effort quite distinct from singular drum strokes. It seems that we may observe so-called *phase-transitions* between singular impulsive actions and iterative actions, as in the case of the drum roll: With increasing speed there is a transition from singular to iterative actions. Notably, this is valid for other tone repetitions as well, e.g. a rapid rush of tones on any instrument may be perceived as an iterative sound, but slowed down below a certain threshold, the tones may be perceived as sustained.

Furthermore, we may see multiple frequencies simultaneously at work in the gestural rendering of musical sound, e.g. see the combination of high-frequency finger movement with low-frequency hand/arm movement in the rendering of rapid passages of piano music. This subsumption of movements into more superordinate movements (e.g. finger movements as sub-movements of the hand/arm movements) can be understood as *coarticulation*. In coarticulation, the focus is typically shifted to a more superordinate trajectory, actually providing the basis for the parsing of music-related movements into chunks. This process of action chunking is important for understanding the process of auditory chunking as well, because we may often see that the overall trajectory of the chunk, e.g. its pitch contour or its textural spread, is more readily rendered, and thus presumably more robustly perceived, than the detail movements within the chunk.

4. The perception-action cycle of gestural rendering

On the basis of our own observation studies and other research, it seems reasonable to conclude that listeners readily may make gestural renderings that reflect features of musical sound, but that these gestural renderings may be quite variable in resolution or acuity of detail. Seeing such variability in gestural renderings of musical sound, one of the main challenges in studying music-related gestures then becomes that of understanding and appreciating *approximate information*.

This can be done by trying to understand gestural renderings of musical sound as an instance of what has been called the *perception-action cycle* [21]. Various versions of this may be found in different domains of research, but in our context we can understand the perception-action cycle as an incessant process of trying to make sense of what we hear by covertly or overtly making gestures that simulate the generation of what we hear, or that trace the perceived contours (or what is often called envelopes) of what we hear. This means that perception is understood as an *active* process where the point is to proceed by sketches from initially rather coarse images to progressively more refined images. This could also be understood as a process of incessant production of hypotheses as to the causes and features of what we perceive [22], in other words, engaging in a feedback loop of an incessant process of top-down hypothesis-generation followed by bottom-up driven comparison with what we assumed in our hypothesis, successively adjusting and refining our top-down generated hypothesis by each period of the perception-action cycle as schematically illustrated in figure 4. One essential point of this perception-action cycle (as well as in motor theories of perception in general) is that the motor sensation may vary considerably in detail, i.e. may vary between very precise images of sound-production as is often the case with trained musicians [12] and more unspecific sensations of onset events as when listeners wave their hands to the rhythm of the music [8]. In both the expert/specific movements and the non-expert/unspecific movements there will be a coupling between the sounds and actions, a coupling that is one of the core elements of embodied music cognition (see [8] for a further discussion of this).



Fig. 4. Gestural rendering, overt or covert (as mental simulation), of musical sound as an instance of the *perception-action cycle* where the coupling of sound to action and action to sound enhances our perception of musical sound by this incessant circular activity of listening and rendering. Notably, the images of actions associated with the sounds may vary considerably with regards to specificity and/or expert knowledge, yet may still be an essential element in the embodied perception of music.

Accepting that such a perception-action cycle is at work, we can then appreciate the often quite approximate types of gestural renderings that we have seen, where initially rather coarse and sometimes even seemingly divergent gestural renderings may be seen as necessary in the perception and progressively finer discrimination of musical sound. Reminiscent of children's babbling, such approximate gestural renderings also indicate that the holistically perceived chunk of musical sound (e.g. a rhythmical motive, a textural fragment, a melodic fragment, etc.) may be primordial to the singular sound or tone. In other words, the gesture is primordial to the note in music, something that is not obvious in western musical thinking.

This capacity for perceiving chunks better than details remains to be better understood, as does also the time-scales at work here, i.e. how often and at what time-intervals our percepts are updated in such a perception-action cycle. Although we tend to think of both musical sound and music-related actions as continuous in time, there are also indications of an intermittent segmentation of sound and action into chunks in our minds. Briefly stated, both sound and action requires a certain minimum duration in order to be perceived as meaningful, something that is due to constraints of our perceptual-cognitive apparatus, and also actions need to be planned and initiated in advance (see [23] for an overview of recent research on chunking of sound and action). One interesting idea here is to think of action as a combination of key-postures and movement between these key-postures, similar to the distinction in animation between so-called keyframes and interframes [24]. This distinction between key-postures and movements seems to fit quite well with the notion of geometry and effort in our case, in that the key-postures are the position and spread of the effectors (i.e. fingers, hands, arms, feet, etc.) at certain salient points in the music such as downbeats and other accents, and that there are gestures with distinct sensations of effort between these key-postures. In our present research, we have adopted this idea of key-postures and movements between key-postures as a model for both being able to make more wellfounded chunkings of music-related gestures and to have better knowledge of the motion and effort features of music-related gestures [18, 23]. For this reason, we will also in our future research on music-related gestures make use of the conceptual separation of geometry and effort that I have presented in this paper.

5. Conclusion: Thinking music with hands

From various research on music-related gestures, our own as well as that of others, it seems reasonable to conclude that listeners' capacity for making spontaneous gestural renderings of musical sound seems solid and seems to reflect well the geometric layout and the sound-producing effort of musical sound. Actually, we could call this 'thinking music with hands', and we believe this is a phenomenon well worth studying further, as it attests to a spontaneous motor involvement that many (if not most) people have with music. Also, thinking music with hands is interesting in demonstrating how ephemeral and/or fleeting musical sound can become more solidly present as motor images and as visible trajectories and postures, giving us insights into the enigmatic issues of musical memory and capacity for anticipation, i.e. for thinking ahead in musical sound. Further research on this thinking music education, musical composition, improvisation, performance, control of new musical instruments, and in multi-media arts, as well as interesting insights for gesture research in general.

Needless to say, we have so far just scratched the surface of this topic of gestural renderings of musical sound, and we have substantial challenges ahead. These include developing better means for motion capture, better representation of motion capture data, better design for observation studies of how people make hand movements in ecologically valid (i.e. non-laboratory) situations, and better methods for studying the relationship between movements that we can observe and covert, mental sensations of movement that we can not see directly.

References

- 1. http://musicalgestures.uio.no
- 2. Godøy, R. I.: Motor-mimetic Music Cognition. Leonardo 36:4, 317-319 (2003)
- 3. Gritten, A., King, E. (eds.): Music and Gesture. Ashgate, Aldershot (2006)
- 4. Wanderley, M., Battier, M. (eds.): Trends in Gestural Control of Music. Ircam, Paris (2000)
- 5. Gibson, J. J.: The Ecological Approach to Visual Perception. Lawrence Erlbaum Associates, Hillsdale, NJ. (1979)
- Godøy, R. I., Haga, E., and Jensenius, A.: Playing 'Air Instruments': mimicry of sound-producing gestures by novices and experts. In Gibet, S., Courty, N., Kamp, J.-F. (eds.): GW2005, LNAI 3881, pp. 256–267. Springer-Verlag, Berlin (2006)
- Godøy, R. I., Haga, E., Jensenius, A. R.: Exploring Music-Related Gestures by Sound-Tracing: A Preliminary Study. In Ng, K. (ed): Proceedings of the COST287-ConGAS 2nd International Symposium on Gesture Interfaces for Multimedia Systems, pp. 27-33 (2006)
- 8. Godøy, R. I.: Gestural affordances of musical sound (2008, book chapter in preparation)
- 9. Liberman, A. M., Mattingly, I. G.: The Motor Theory of Speech Perception Revised. Cognition, 21, 1-36 (1985)
- Wilson, M., Knoblich, G.: The case for motor involvement in perceiving conspecifics. Psychological Bulletin 131:3, 460–473 (2005)
- 11. Kohler, E., Keysers, C., Umiltà, M. A., Fogassi, L., Gallese, V., Rizzolatti, G.: Hearing sounds, understanding actions: Action representation in mirror neurons. Science 297, 846–848 (2002)
- Haueisen, J., Knösche, T. R.: Involuntary Motor Activity in Pianists Evoked by Music Perception. Journal of Cognitive Neuroscience 13:6, 786–792 (2001)
- Rizzolatti, G., Arbib, M. A.: Language Within Our Grasp. Trends in Neuroscience 21, 188–194 (1998)
- 14. Goldin-Meadow, S.: Hearing Gesture: How Our Hands Help Us Think. The Belknap Press, Cambridge, Mass. (2003)
- Kelso J.A.S., Fuchs A., Lancaster, R., Holroyd, T., Cheyne, D., Weinberg, H.: Dynamic cortical activity in the human brain reveals motor equivalence. Nature 392:23 (1998) 814-818
- Haga, E.: Correspondences between Music and Body Movement (unpublished doctoral dissertation). University of Oslo, Oslo (2008)
- Eitan, Z., Granot, R. Y.: How Music Moves: Musical Parameters and Listeners' Images of Motion. Music Perception 23:3, 221-247 (2006)
- Jensenius, A. R., Nymoen, K., and Godøy, R. I.: A multilayered GDIF-based setup for studying coarticulation in the movements of musicians. In Proceedings of ICMC 2008, Belfast, UK, 24 -29 August 2008 (2008)
- Pirhonen, A.: Semantics Of Sounds And Images Can they be Paralleled? Proceedings of the 2007 International Conference for Auditory Display; June 26-29; Montreal, Canada, 319-325 (2007)
- 20. Godøy, R. I.: Gestural-Sonorous Objects: embodied extensions of Schaeffer's conceptual apparatus. Organised Sound 11:2, 149-157 (2006)
- 21. Neisser, U.: Cognition and Reality. W.H. Freeman, San Francisco (1976)
- 22. Berthoz, A.: Le sens du mouvement. Odile Jacob, Paris (1997)

- Godøy, R. I.: Reflections on chunking in music. In Schneider, A. (ed.), Systematic and Comparative Musicology: Concepts, Methods, Findings. Hamburger Jahrbuch f
 ür Musikwissenschaft, Band 24. Vienna: Peter Lang. pp. 117-132 (2008)
- 24. Rosenbaum, D., Cohen, R. G., Jax, S. A., Weiss, D. J., and van der Wel, R.: The problem of serial order in behavior: Lashley's legacy. *Human Movement Science*, vol. 26, 4, 525-554 (2007)