

Proceedings of the
3rd International Conference on
Music and Gesture

Centre for Interdisciplinary Research in Music Media
and Technology

Schulich School of Music

McGill University

Montréal, Québec, Canada

March 5 – 6, 2010

Tanna Schulich Hall

Edited by

Marcelo M. Wanderley, McGill University

Jon Wild, McGill University

Sha Xin Wei, Concordia University

3rd International Conference on Music and Gesture

Scientific Committee

- Emmanuel Amiot - Lycée Arago, Perpignan
- Rick Ashley - Northwester University
- Marc Battier - Sorbonne University
- William Caplin - McGill University
- Isabelle Cossette - McGill University
- Jane Davidson - University of Western Australia
- Sylvie Gibet - University of South Brittany
- Rolf Inge Godøy - University of Oslo
- Anthony Gritten - Middlesex University
- Robert Hatten - Indiana University
- Alexander Refsum Jensenius - University of Oslo
- Elaine C. King - University of Hull
- Geoff Luck - University of Jyväskylä
- Pat McCreless - Yale University
- Petri Toiviainen - University of Jyväskylä
- Marcelo M. Wanderley - McGill University
- Sha Xin Wei - Concordia University
- Jonathan Wild - McGill University

ISBN: 978-0-7717-0673-8

February 2010

Schulich School of Music, McGill University

Copyright remains with the authors

Friday, March 5th 2010

8h45 **Welcome**

9h00 **Session I**

- 9h00 Highlighting structural issues in musical performance with optical finger tracking, J. MacRitchie, N. J. Bailey, G. Hair
- 9h25 Performer's body motion and phrase structure: The role of velocity magnitude, B. Buck, N. J. Bailey, J. MacRitchie, R. Parncutt
- 9h50 Facial expressions in speech and singing, S. R. Livingstone, C. Palmer, W. F. Thompson, M. M. Wanderley
- 10h15 Coordination of facial muscle activation, intraoral pressure and mouthpiece force during trumpet playing: Preliminary findings, V. Fréour, T. Bianco, I. Cossette, F. Bevilacqua, R. Caussé

10h40 **Coffee Break**

11h10 **Session II**

- 11h10 Coordination of musical parameters and the role of gesture between co-performers, N. Romano, S. Brennan
- 11h35 Actual versus perceived embodiment of a rhythmic pulse, G. Luck, P. Toiviainen, M. R. Thompson
- 12h00 Choreo-musical space frames and implications of multi-channel audio in dance performance, P. V. Swendsen

12h25 **Lunch Break**

14h20 **Session III**

- 14h20 The fingerboard instruments: Reframing lutherie without strings, A. Freed
- 14h45 Constructing a personalizable gesture-recognizer infrastructure for the K-Bow, R. Fiebrink, M. Schedel, B. Threw
- 15h10 Concurrence and counteraction in musical gesture and form, D. A. Stewart, J. Malloch
- 15h35 Enphonic Graphomania, P. H. Harrop

16h00 **Coffee Break**

- 16h30 **Keynote I** – Gesture, mathematical language (and music?), Prof. Brian Rotman, Ohio State University

17h30 **End of day I**

Saturday, March 6th 2010

9h00 **Session IV**

- 9h00 Dmitri Shostakovich's deformational sonata-form gestures and the interpretive implications, S. Reichardt
- 9h25 Heightened harmonic gesture in Wagner's music dramas, D. H. Smyth
- 9h50 Defaulting to gesture: How listeners extract meaning in the absence of cognitively accessible musical organization, R. S. Hatten
- 10h15 Whither gestural analysis? P. McCreless

10h40 **Coffee Break**

11h10 **Session V**

- 11h10 Gesture and agency in responsive media, X. W. Sha
- 11h35 Movement prosody - emergence of collective action, S. P. Gill
- 12h00 Computer evaluation of musical gesture in sound structure, C. Dobrian

12h25 **Lunch Break**

14h20 **Session VI**

- 14h20 Key-postures and trajectories in music-related gestures, R. I. Godøy
- 14h45 Analysis of musical gestures using empirical mode decomposition, P. Toiviainen
- 15h10 Study on gesture-sound similarity, B. Caramiaux, F. Bevilacqua, N. Schnell
- 15h35 Encoding emotion: How performers manipulate tempo locally to convey affect, M. Swoboda, M. Wagner

16h00 **Coffee Break**

16h30 **Keynote II** – Thinking gesturally, Dr. Anthony Gritten, Middlesex University

17h30 **End of the conference**

Highlighting structural issues in musical performance with optical finger tracking

J. MacRitchie, N. J. Bailey, G. Hair
Centre for Music Technology, University of Glasgow

Gestural recognition in musical performance demonstrates links between direction, velocity and performer's intentions and interpretation of the composed structure of the music (Davidson & Correia 2002, Wanderley et al. 2005). Low level movements present in finger-tracking particularly are becoming interesting in both piano and clarinet performances, revealing information about anticipatory movement (Goebel & Palmer, 2006) and stylistic identifiers in velocity and acceleration (Dalla Bella & Palmer, 2006). We aim to use intricate details of finger movement in piano performance to demonstrate performance choices in terms of the low-level structures of the music.

A portable monocular image-tracking system designed for use in analysing piano performances is described. A high frame-rate camera captures birds-eye view images of the hands at the keyboard, whilst an ultra-violet light fluoresces the passive UV-reflective paint markers on the joints of each finger. 3D information is estimated from the coordinates of reference markers near the wrists.

Even at >50fps, concert pianists' gestures are sufficiently rapid that remedial action is required to establish accurate tracking. Over and above a basic system using blob tracking and correlation functions, various algorithmic improvements are assessed and their relative benefits towards automating the data acquisition process established. The incorporation of aspects of a skeletal model is particularly beneficial, increasing the tracking reliability by at least 40%. This finger-tracking system can also be used to estimate the curvature of fingers as they strike the keys as well as fingering patterns, informing piano pedagogy.

The optical tracking system forms part of a multi-modal acquisition system, recording movement, MIDI and audio from six professional performances of Chopin's B flat minor sonata finale op.35. This is a problematic composition in terms of its structural analysis; multi-modal performance analysis is used to uncover differences and similarities between performers in the opening bars. We ask whether bar 5 marks the introduction of a new theme or whether this is an extension of the ideas set out at the very beginning – a point of contention for the traditional musical analyses in existence. 3D motion in conjunction with tempo and dynamics reveal how performers group sets of quavers, as well as accents and stresses on particular notes. Results suggest five out of the six performers do not consider bar 5 of the sonata finale to be a major structural boundary.

References

Davidson, J.W. & Correia, J. S. (2002) Body movement. In R. Parncutt & G. E. McPherson (Eds.) *The Science and Psychology of Music Performance* (pp. 237–250). Oxford: Oxford University Press.

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.

Goebel, W. & Palmer, C. (2006). Anticipatory motion in piano performance. *Journal of the Acoustical Society of America*, 120, 3002.

Dalla Bella, S. & Palmer, C. (2006). Personal identifiers in musicians' finger movement dynamics. *Journal of Cognitive Neuroscience*, 18, Supplement, G84.

Performers' body motion and phrase structure: The role of velocity magnitude

B.Buck¹, N. Bailey¹, J. MacRitchie¹, R.Parncutt²

¹Centre for Music Technology, University of Glasgow

²Centre for Systematic Musicology, University of Graz

Musicians utilise expressive nuances through body movement to highlight structural features of the notated score, an example being to express phrase structure, which is hierarchically organised. The magnitude of the *ritardando* in performance has been shown to be dependent on the hierarchical level of the phrase (Todd, 1992). An audience can determine the phrasing structure of a piece from simply viewing a muted performance (Davidson, 1993; Wanderley, 2001; Vines et al., 2006; Nusseck & Wanderley, 2009; Buck et al, 2009). Investigations into the ancillary movements of clarinetists have revealed that observers make use of performers physical motion to make judgements of musical tension, phrasing intensity and fluency (Vines et al., 2006; Nusseck & Wanderley, 2009). This is consistent with the view that musical experience is multi-sensory, not residing in the aural element alone.

Where previous investigations have focused on the perception of musical tension and affect, here we directly analyse bodily motion of 9 professional pianists performing two selected Chopin preludes (A major & B minor; Op.28: 6&7). We compare and analyse each performance in terms of overall magnitude of velocity and acceleration and, finally, relate both body motion and perceived phrasing to traditional structural analysis. By using a stationary instrument such as the piano we remove the confound of breathing motions that arise with a wind instrument. Employing 9 performers and 15 observers, for 2 preludes, we help to ensure a range of interpretations.

We hypothesise that an audience's interpretation of musical structure is dependent on the velocity of the performers overall movement. We propose that the performer moves less towards the end of each phrase. It is also anticipated that this amount of movement relates to the *accelerandi* and *ritardandi*, with the size of effect dependent on the hierarchical level of the phrase structure.

18 3D motion capture recordings (comprising 2 performances for each of 9 performers, capturing 3D motion of 38 markers placed on the performers' head and upper body; adapted from Cutti et al., 2005's gait model) were analysed to assess the variation of motion expressed by each performer during the two performances. To assess the magnitude of velocity and acceleration, one sternum and four head markers are analysed for each performer. The 18 performances were also presented to 15 'audience' observers in a visual only condition. Observers were asked to move a slider to the right as a phrase reached the middle and return to left as the phrase ended (for detailed analysis see MacRitchie et al, 2009).

Preliminary analysis of one performer suggests overall body movement speeds up at the start of a phrase and slows towards the end. This may coincide with variance of motion direction, producing a visual cue for the audience. Analysis is proceeding on selected

performances rated by experts for performance quality. This research confirms previous findings as to the multi-sensory nature of music perception. Initial analyses show that the magnitude of the velocity and acceleration of body motion directly relates to the hierarchical level of the phrase structure.

References

Todd, N. (1992) The dynamics of dynamics: A model of musical expression. *Journal of the Acoustical Society of America*, *91*, 3540–3550.

Davidson, J. W. (1993) Visual Perception of Performance Manner in the Movements of Solo Musicians. *Psychology of Music*. *21*, 103–113

Vines, B.W., Krumhansl, C., Wanderley, M. M. & Levitin, D. (2006) Cross-modal Interactions in the Perception of Musical Performance. *Cognition*, *101*, 80–113

Wanderley, M. M. (2001) Quantitative Analysis of Non-Obvious Performer Gestures. *Gesture and Sign Language in Human-Computer Interaction: International Gesture Workshop*. Heidelberg, Germany: Springer Verlag, pp. 241–153.

Nusseck, M. & Wanderley, M. M. (2009) Music and Motion - How Music-Related Ancillary Body Movements Contribute to the Experience of Music. *Music Perception*, *26*(4):335–353

Goebl, W. & Palmer, C. (2009) Synchronization of Timing and Motion Among Performing Musicians. *Music Perception*, *26*(5):427–438

Cutti, A.G., Paolini, G., Troncossi, M., Cappello, A. & Davalli, A. (2005) Soft tissue artefact assessment in humeral axial rotation. *Gait and Posture*. *21*, 341–349.

MacRitchie J., Buck B., & Bailey N. (2009) Gestural communication: Linking multimodal analysis of performance to perception of musical structure. In *Proceedings of the International Symposium on Performance Science 2009*, Auckland, New Zealand. ISBN 978-94-90306-01-4.

Buck B., MacRitchie J., & Bailey N. (2009) Perceptual recognition of embodied musical structure. Paper presented at *The Musical Body Conference*, University of London. UK.

Facial expressions in speech and singing

S. R. Livingstone¹, C. Palmer¹, W. F. Thompson², M. M. Wanderley³

¹ Dept of Psychology, McGill University

² Dept of Psychology, Macquarie University

³ IDMIL/CIRMMT, Schulich School of Music, McGill University

Facial expressions are an integral part of nonverbal human communication and convey important information about the speaker's emotional state and intentions. In singing, facial expressions communicate both structural and emotional information (Thompson & Russo, 2007), and in speech, movement of the head conveys important linguistic information (Munhall, Jones, Callan, Kuratate, & Vatikiotis-Bateson, 2004). Do these movements play a direct role in the vocal production of song? Or do they supplement the acoustic signal with an additional modality for communication?

Few studies have investigated the role of facial expression in emotional singing. Recently, Livingstone, Thompson, and Russo (2009) observed that singers' emotional intentions could be identified from facial features. Participants were shown video recordings of a model target singing linguistically neutral statements, each with three emotional intentions (happy, sad, and neutral). Participants had to sing back the target sentences with the same emotion. Participants had four distinct phases of facial expressions: perception, planning, production, and post-production. Qualitative movement differences between the phases suggested that facial expressions complemented the acoustic signal. However, the study did not separate movements specific to vocal production of sound from those specific to emotional state.

We address how emotional states are communicated in facial movements, and whether those movement patterns differ in the production of speech and song. We describe a study in which singers perform an emotional singing and speaking task. Stimuli that reflect neutral English statements are each produced in natural (unpaced) speech with 5 emotional intentions: very happy, happy, very sad, sad, and neutral. Song stimuli contain the same statements paired with a single isochronous 7-note melody. On different trials, participants speak or sing the statement with the specified emotion. Participants are recorded with a microphone, video camera, and motion capture equipment. To confirm the desired emotion was produced, utterances are rated by a separate group of participants using a 5 point scale of emotion, and a 3 point scale of response confidence. Syllable onsets and facial movements are aligned across participants for identical statements using Functional Data Analysis techniques. We expect that a subtraction of the motion profile of neutral statements from those used in emotive statements will yield "ancillary" movements specifically used for the communication of emotion. Furthermore, we expect that ancillary movements will be larger in the "very happy vs happy", and "very sad vs sad" conditions. Finally, we expect that the same motion profile will emerge in spoken and sung versions of the same statements, with more exaggerated versions in song.

References

Livingstone, S. R., Thompson, W. F., & Russo, F. A. (2009). Facial expressions and emotional singing: A study of perception and production with motion capture and electromyography. *Music Perception, 26*, 475-488.

Munhall, K., Jones, J., Callan, D., Kuratate, T., & Vatikiotis-Bateson, E. (2004). Head movement improves auditory speech perception. *Psychological Science, 15*, 133-137.

Thompson, W. F., & Russo, F. A. (2007). Facing the music. *Psychological Science, 18*, 756-757.

Coordination of facial muscle activation, intraoral pressure and mouthpiece force during trumpet playing: preliminary findings

V. Fréour¹, T. Bianco², I. Cossette¹, F. Bevilacqua², R. Caussé²

¹ Department of Music Research, Schulich School of Music, McGill University

² Ircam, UPMC

Sound production during trumpet playing implies the interaction between the performer and his instrument which requires the coordination between physiological and physical parameters under the performer's direct control. We consider this sound control as a gesture achieved through setting of the lip mechanical properties (mass and stiffness) and their aerodynamic excitation provided by the respiratory system.

In the literature, studies particularly focus on control parameters such as intraoral pressure and force applied by the lips on the mouthpiece (Barbenel 1988, Fletcher 1998). Moreover, some works (White and Basmajian, 1973) conducted on facial muscles recruitment during playing describe coordination between muscle activation and sound production.

Whereas previous studies describe the magnitude of these control parameters in quasi-static regime of oscillation (sustained tones), the present account focuses particularly on the time domain evolution of the control parameters during note transients, and on the coordination of the parameters between subsequent notes. These observations have been made possible by combining two measurement systems from two different laboratories. Our set up includes: facial surface electromyography, sound (Respiratory mechanics lab-CIRMMT, Schulich School of Music, McGill University), intraoral pressure and force applied on the mouthpiece (Instrumental Acoustics laboratory, IRCAM).

Electrical activity of two facial muscle groups, respectively associated with smiling and depressing the lower lip (zygomaticus and depressor anguli oris-DAO) was measured with surface electromyography (EMG). The force applied by the mouthpiece in its axis was recorded thanks to a force sensor mounted on a specially shaped mouthpiece receiver. Intraoral pressure and radiated sound were recorded using, respectively, a pressure transducer and a microphone mounted to the trumpet at a constant distance from the instrument bell. 6 high-level trumpet players performed some equally spaced quarter notes at the dynamics pp, mf and ff with two different accentuations (tenuto, staccato).

Temporal profiles taken into account are: 1-Radiated sound envelope, 2-intraoral pressure, 3- envelope of EMG activity, 4- mouthpiece force.

Preliminary results reveal good consistency in between subjects in terms of intraoral pressure profile. An overshooting of intraoral pressure is observed before the sound production during tenuto notes whereas the increase of pressure occurs monotonously in the case of staccato. Moreover, experimental data confirm that EMG activity distributes asymmetrically to tone production, with a higher activity prior to tone commencement (as in Heuser and McNitt-Gray, 1991). EMG activity also shows complementary behavior to mouthpiece force, suggesting that lips' stiffness setting is regulated by the joined control

of their muscular tension and compression.

These first analyses revealed different control parameter patterns following accentuation. Notable differentiations were observed with intraoral pressure. Further analysis must focus on the coordination between all variables, aiming at identifying some specific features relevant to each type of articulation.

Acknowledgements

This work was partially supported by the Short-Term Scientific Mission funding program of the COST Action IC0601 on Sonic Interaction Design and Canadian Foundation for Innovation.

References

Barbenel, J.C., Kenny, P. & Davies, J.B. (1998). Mouthpiece forces produced while playing the trumpet. *Journal of Biomechanics*, 21, 417–424.

Fletcher, N. H. & Tarnopolsky, A. (1999). Blowing pressure, power, and spectrum in trumpet playing. *Journal of Acoustical Society of America*, 105, 874–881.

Heuser, Frank & McNitt-Gray, Jill L. (1991). EMG potentials prior to tone commencement in trumpet players. *Medical Problems of Performing Artists* 6, 51.

White, E.R. & Basmajian, J.V. (1973). Electromyography of Lip Muscles and Their Role in Trumpet Playing. *Journal of Applied Physiology*, 35, 892–897.

Coordination of musical performance and the role of gesture between co-performers

N. Romano¹, S Brennan²

¹ University of Illinois at Urbana-Champaign

² Stony Brook University

Musicians need to communicate in order to effectively coordinate an ensemble performance. However, musicians are at the distinct disadvantage of not being able to use speech, which is the primary tool in coordination of other instances of joint action. Musicians must instead rely on non-verbal cues to coordinate their performances. This study examined the role of gesture and expertise in coordination of musical performance in pianist-violinist dyads. Although not a comprehensive list of gestures employed during musical coordination, the gestures looked at in this study were head nods and exaggerated sound production movements. Participants were recorded during performance using two cameras, one focusing on each participant. Head nods and exaggerated sound production movements were counted for each performer. For violinists, exaggerated sound production movements were coded as bow arm movements that had a larger range of motion than normal. For pianists, exaggerated production movements were coded as hand movements that were a greater distance away from the keyboard than normal. Asynchrony was coded as either early, late, or no entrance at all by one or both players at pre-determined critical points in the piece, defined as beginnings or endings of phrases or sub-phrases. The data show that the more frequently individuals in a dyad gestured, the fewer asynchronies their performance contained. There was also a positive correlation between experience and gesture frequency. Interestingly, the frequency of coordinating gestures was a better predictor of synchrony than the combined experience of the violinist-pianist dyads. These results suggest that gesture plays an integral role in the coordination of musical performance and that ability to coordinate a performance using gesture increases with experience. Future research into coordinating gestures in music performance will provide insight into the communicative nature of music and may help musicians maximize their performance potential.

Actual versus perceived embodiment of a rhythmic pulse

G. Luck, P. Toiviainen, M. R. Thompson
University of Jyväskylä

Background

When tapping in-time with point-light representations of rhythmic human movement, people tend to synchronise with certain kinematic features of the movements presented, most notably with peaks in acceleration along the trajectory. When moving spontaneously to auditory rhythmic stimuli, however, people tend to synchronize vertical movements, specifically peaks in high downward velocity, with the rhythm. In the former case, a visual pulse is perceived as being embodied by peaks of acceleration along the movement trajectory; in the latter case, an auditory pulse is actually embodied as peaks in vertical velocity. This suggests either that people embody rhythmic features differently depending on the precise nature of the task (indicating the pulse, or following the pulse), or that perceived embodiment of such features differs from actual embodiment.

Aims

Here, we examine differences between perceived and actual embodiment of a rhythmic pulse.

Method

Participants tapped in-time with point-light representations of rhythmic full-body movement derived from motion-capture recordings of adults moving spontaneously to beat-driven music (stimuli were visual-only). From the tapping data, kernel density estimation was used to determine tapping probability density, and six stimuli for which there were clear periodic tapping responses from participants were selected for further analysis. From the movement data, instantaneous velocity and acceleration of joint locations, as well as angular velocity and acceleration of limb segment direction vectors and joint angles, were estimated by numerical differentiation. These features were then correlated with tap point timing across participants.

Results

Participants most frequently tapped in-time with acceleration components, followed by velocity components. Jerk components were also occasionally synchronized with. Moreover, tapping regularity was found to correlate with the internal phase coherence of the movement. When analysed overall, correlation coefficients remained low, peaking at .25. However, when each stimulus was analysed separately, coefficients peaked at .6. In summary, participants tended to tap in-time with acceleration (both vertical and along the trajectory), and mainly with acceleration of the extremities (hands, feet, head).

Conclusions

Perception of rhythmic movement differs from embodiment. When embodying, velocity is synchronized with a pulse; when perceiving, acceleration is taken as the cue for synchronization. This is true both when the movement is produced without reference to an external timing signal (such as conducting gestures), and when originally synchronized with such a signal (such as when dancing to music).

Choreo-musical space frames and implications of multi-channel audio in dance performance

P. V. Swendsen
Oberlin Conservatory of Music

Since the middle of the twentieth century, electroacoustic music and dance have been increasingly partnered. While recent scholarship increases our understanding of important artistic processes and outcomes within both fields individually, rarely does it examine the meeting of these two disciplines. In this paper, I examine how scale and physical space are altered when electroacoustic music and dance are combined. In particular, I discuss the implementation of multi-channel audio in dance performance.

Scale and space are particularly important choreo-musical compositional parameters. Mismatches between the scale and space of music and dance can muddy the perceptual waters, while shared exploration can create for the audience an engagingly transmodal experience. In *Living Electronic Music* (2007, 97-98), Simon Emmerson contextualizes scale in a listening environment through “the idea of the frame (a defined area of interest) applied progressively from the largest to the smallest scale:” landscape, arena, stage, and event. Landscape and arena are considered “field” space frames, and stage and event are “local” space frames. “Local controls and functions seek to extend (but not to break) the perceived relation of human performer action to sounding result. Field functions create a context, a landscape or an environment within which local activity may be found.” Emmerson’s terminology is useful in considering how scale impacts the “perceived relation” of dancer and sound. How can those relations be extended, and at what point are they broken? How does scale manifest itself in pieces involving electroacoustic music and dance, and how can we maximize its potential as a composable parameter? How does the visual and physical presence of the dancers define space frames for the electroacoustic composer?

Multi-channel playback systems, while increasingly ubiquitous in electroacoustic concert spaces, present a unique challenge when used with dance. Dance does not benefit from the same first- person audience space enjoyed by cinema, nor does it provide the sound-producing actions of live instrumental performance. The visual data transmitted via dance is the actual expressive content, not the means of production. Dance demands the audiences’ visual attention and transfixes their gaze. The constant frontward pull of the visual undermines the use of spatialized audio, which threatens to disconnect the dancers from what the audience perceives as their stage- defined world. In Emmerson’s terms, it confuses the boundary between stage and arena. Indeed, even dramatic gestures in stereo space can detract from the spatial relationships at play in the onstage environment. I do not argue for the dismissal of multi-channel diffusion in dance performances, but rather encourage careful thought about the spatial relationships and frictions it may produce.

Electroacoustic music and dance collaborate to define and explore all levels of scale and space. In doing so, they create for the audience a uniquely visceral sense of space over time. This sense is sometimes intimate and sometimes expansive, and has the capacity to surpass, in nuance and variety, what either music or dance might achieve on its own.

The fingerboard instruments: Reframing lutherie without strings

A. Freed
CNMAT, UC Berkeley

Recent attempts to extend organology to add useful classifications of electrophones and music controllers have rightly focussed on music performance gestures. The naive approach of classification by gestural types (strum, pluck, hit, slap etc..) fails because gestures say more about musicians and their music than their instruments. The routine development in institutions of "extended" techniques and the drive for uniqueness and theatricality in popular music practice results in each instrument being associated with an increasing number of different performance gestures defeating useful, differentiating classifications. Also, Hendrix's enkindling of the guitar guitar, or Townsend's windmill strum are gestures associated with a time, an individual, and a repurposed ritual but such gestures could be used on any instrument. The british sport of piano bashing where the winning team is the fastest to destroy and put an entire piano through a letterbox reminds us of the wide range of expected and unexpected gestures that may be associated with musical instruments.

A more useful approach is classify by structural attributes looking from the controller out toward the performer. The example I will focus on is fingerboard controllers and instruments. This family includes the guitar and cello but not the harp, dividing traditional chordophones in a way that makes sense with common performance practice. All these instruments can be strummed by virtue of their strings but the hammer-on and pull-off are not available on the harp. More interestingly this classification organizes the early electrophones and early and recent controllers. The Ondes Martenot, Hellertion Trautonium, and Theremin cello, synthesizer ribbon controller, Continuum finger board and Guitar Hero are members of the fingerboard instruments but the Theremin and organ are not.

Like all effective classification systems this one yields some annoying but interesting edge cases like the koto and iPhone. The koto player doesn't interact with a fingerboard directly but the pillars mounted on the sound board, dividing the string are an essential affordance. The iPhone has both proximity sensing (a la Theremin) and a surface for finger interaction whereas most of the other fingerboard instruments have a chord, strip or switch between the player and the fingerboard.

An interesting feature of fingerboard controllers is that they can be differentiated from each other by structural features integrated to discretize gestures - usually those that involve pitch. These fiducials may be visual (slide guitar, iPhone), mechanical (frets), statically haptic (dimples of the Ondes Martenot), or actively haptic (lamella of the Trautonium).

The fingerboard classification allows us to frame David Wessel's Slabs and my own recent controllers as members of a family of related controllers with a long tradition rather than as "unique inventions" whose newness has to be confronted. On the Slabs the fiducials are a rectangular grid of slits with the interaction pads tiled between them in two

dimensions. My "Big Guitar" can be seen as a hybrid koto, fretless bass guitar and Hellertion. My Tablo fabric drape controller shares features of the Hellertion and Theremin cello with a distinguishing annular fingerboard. The 12-stringless cello developed with Frances Marie Uitti has features of ribbon controllers but with position/pressure sensing strips on both sides of the fingerboard and a structurally independent array of rods for two-bow interactions.

This framing suggests the fruitful approach of design space exploration for future controllers instead of the common (romantic) idea of a quest for "invention" of a disruptive breakthrough new controller.

Also this framing suggests an atypical answer to the question of which field new musical controller design and development belongs in, i.e. the plastic arts. One definition (wikipedia) of the plastic arts is that they "involve the use of materials that can be molded or modulated in some way often in three dimensions." This new lutherie stands alongside architecture, textile arts and sculpture involving issues of both concrete forms/structures and engagement of multimodal interactions.

Constructing a personalizable gesture-recognizer infrastructure for the K-Bow

R. Fiebrink¹, M. Schedel², B. Threw³

¹ Princeton University

² Stony Brook University

³ Keith McMillen Instruments

Keith McMillen Instruments sells a series of properly weighted and balanced Bluetooth sensor bows for violin, viola, cello and bass—the K-Bow. The K-Bow has eight sensors, the values of which are communicated via OSC: x-, y-, and z-axis accelerometers, hair tension, grip pressure, bow location along frog to tip and bridge to fingerboard, and tilt. We have been constructing a software system to perform recognition of traditional and non-traditional bowing gestures using the K-Bow. In constructing a supervised-learning-based gesture recognizer, we have employed both a traditional supervised learning workflow (i.e., constructing a gesture database and optimizing model performance on that database) and a less traditional, interactive workflow in which we iteratively add training examples, re-train, subjectively evaluate, and modify the gesture recognizer in real-time. Notably, while a K-Bow performer (our end user) may run our own trained gesture recognizer “out of the box,” he or she may also interact explicitly with the learning system in order to easily create, personalize and augment a library of known motions.

We will describe the trajectory of our experience working together as a software engineer, a composer, and a music technology researcher to implement a robust system for interactively re-trainable gesture recognition, using the standardized K-Bow hardware and building on the Wekinator system for real-time, interactive learning. In our experiences, both off-line and real-time machine learning are integral to constructing a robust and flexible gesture recognition system, and appropriate interfaces for these tasks can aid designers and users alike. We will situate this work with respect to prior research on bowing gesture recognition. We will describe the infrastructure built for our own work, including a method for systematically evaluating features computed from the bow sensor outputs, as well as a real-time gesture recognizer prototyping interface optimized for a string player who uses both hands to play. We will discuss our goal of providing the end user (composer or performer) with an interface for modifying the learned gesture recognition model according to her musical expertise and intuition, while not requiring prior machine learning expertise; we will offer an evaluation of how our current user interface supports that goal. Informed by our experiences with the K-Bow, we will then discuss how applying machine learning for gesture recognition differs from more traditional machine learning applications in interesting and non-trivial ways. For example, we may be interested in multiple overlapping labels (e.g., up bow or down bow, fast or slow bow, marcato or détaché); we are also ultimately concerned with not only acquiring labels for gestures, but understanding the degree of confidence the model has that a given label is correct. Importantly, the ability of the performer to add more examples (rather than change the learning algorithm) to improve labeling accuracy distinguishes this task from more traditional machine learning applications.

Our presentation will conclude with a live demonstration of a performer using the K-Bow with a pre-trained gesture recognizer, modifying the gesture recognizer to accommodate her personal style, and adding a new gesture to the recognizer's vocabulary.

Concurrence and counteraction in musical gesture and form

D. A. Stewart¹, J. Malloch²

¹ Digital Composition Studios, ² Input Devices and Music Interaction Laboratory
Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT)
McGill University

Our contribution to the theme of 'music and gesture' involves describing gestural controller mapping principles from the perspective of the performer and composer. One of the objectives of our work is to effect a transformation from lab-based gestural controller into " musical instrument " through the development of performance practices, music notation standards, compositional language and a body of solo and ensemble repertoire.

Central to this effort is the continuing development of the soprano t-stick.¹ The t-stick is a physical input device for controlling parameters of a sound synthesis algorithm and is capable of sensing where and how much of its surface is touched by the performer, as well as detecting gestures such as tilting, shaking, squeezing or twisting (Malloch & Wanderley, 2007).

Since this device outputs only sensor data rather than sound, it is necessary to develop multi-layered mapping strategies before the t-stick can be played. This is a lengthy and difficult task, which must take into account the aesthetics of culturally-embedded technology, the visual impact of performance gesture, performance ergonomics, the physical format of the controller and the range and responsiveness of the sensors, among other characteristics.

Our research, development, and creation using gesture-controlled audio systems demonstrate that well-designed mapping layers enable a performer to present his or her instrument in a consistent fashion, reinforcing recognisable / identifiable playing techniques within the cultural context of a digital musical instrument and effectively combining a visual performance with associated sounds. Moreover, a thoughtful approach to mapping can result in a consistent learning experience for both the novice and experienced digital instrumentalist. The playing technique becomes self-evident and meaningful as he or she develops competency with the instrument - the playing techniques, themselves, demand that a performer augments his or her playing skills in order to arrive at more satisfying levels of musical expressivity.

It is our belief that the future of electronic music includes a dynamic and vigorous live-performance element rather than more visually static representations, and that this can only be achieved through an ample amount of training. Our work leverages and explores the congruence and counteraction among physical gestures, musical gestures and the

¹ www.vimeo.com/1776361, www.vimeo.com/6390249

compositional project at every step, from the conception of an instrument and musical composition to the eventual performance and reception of the instrument / musical work.

In addition to presenting a discussion on the above, we will illustrate how insights from the evolution of the soprano t-stick can inform the creation and development stages of other devices, including other t-stick models and Serge de Laubier's Méta-instrument.² Furthermore, we will discuss how the attributes of the t-stick may theoretically inform our understanding of digital musical instruments in relation to other devices and control surfaces such as the Hands,³ the Wii game controller and the ubiquitous laptop computer.

Reference

Malloch, J. & Wanderley, M. M. (2007) The T-Stick: From Musical Interface to Musical Instrument. In *Proceedings of the 2007 International Conference on New Interfaces for Musical Expression*, New York City, USA, pp. 66–69.

² www.pucemuse.com

³ www.crackle.org/TheHands.htm

Enphonic Graphomania

P. H. Harrop

Department of Architecture, University of Manitoba

This paper will discuss the development of an interactive drawing interface developed as a device for the co structured development of drawing and sound pieces through the gestural and material actions of traditional drawing techniques. The Enphonic Graphomania device is based on a conceptual variant of a surrealist drawing technique known as Entopic Graphomania, developed by the Romanian surrealist poet, Dolfi Trost (Trost, 1945). Loosely based on this precedent, this paper will outline issues of trans media and trans disciplinary production (in this case sound and drawing) through the mediation of two material and temporal conditions by a gestural act specific to the drawing medium.

An often overlooked, and rather difficult venue for the development of gestural interfaces resides in the production of artistic works. This largely is a question of gestural translation of limited frameworks of movement and time intervals in the evolution of works of art such as painting and drawing. Where dance, and performative gestures reside in a sympathetic temporal scale of traditional performance action, the minute and slow evolving act of drawing tend to be ill-suited for an interactive framework. Yet the act of drawing is one of a haptic interface between body, instrument (a pencil) and a substrate (paper). Most importantly, however the act of drawing is a time dependant, gestural practice that embodies the qualities of precision, rhythm, cadence and emphasis. The challenge then in creating an interactive interchange between sound (music) and drawing is in devising apparatus (both software and hardware) that explore the interchangeable gestures. A device that addresses both of these media in a simultaneous and interchangeable work would parallel the gestural terms of reference in their production such as: intensity, rendering, layering, rubbing, erasing (for drawing) with modulation, additive synthesis, amplification and fading (for sound). This project examines this question through a reciprocal, interconnected and interdependent sound interface || drawing interface. As well, it will investigate the implications of a surrealist generating technique both from the perspective of gestural drawing, but as well within the context of kinematic and dynamic research in gestural interpretation of drawing movements.

As an interface, the Enphonic Graphomania, resembles and behaves as a traditional drawing board. The tools are traditional: pencil, archival paper and an eraser. Like the surrealist game, 32 sensor points are randomly identified on the page, marked with graphite and the drawing exercise consists of developing a density of lines between these points. The conductive nature of graphite presents a variable and highly mutable set of minute voltage differences as the drawing evolves (or rather as the conductive material accumulates on the page). With the evolving (and devolving) conductive material as an input, algorithms and software were developed to interpret these subtle changes using wave modulation, additive synthesis and timed decaying conditions. Rather than developed as a control device, the Enphonic Graphomania attempts to parallel the gradual

development of a drawing with the gradual evolution of an algorithmic score. The intention is to create a reciprocal condition of the co structured evolution of a work of art. The drawing is evolved through sonic conditions that are provoked by the gesture, while the conditions of the drawing are evolved to develop the sonic piece. The aim is an interdependent piece, mutually evolved through the craft of a practiced hand.

Reference

Trost, D. (1945). *Vision Dans Le Cristal, Oniromancie Obsessionnelle et Neuf Graphomanies Entoptiques..* Paris: les Éditions de l'Oubli.

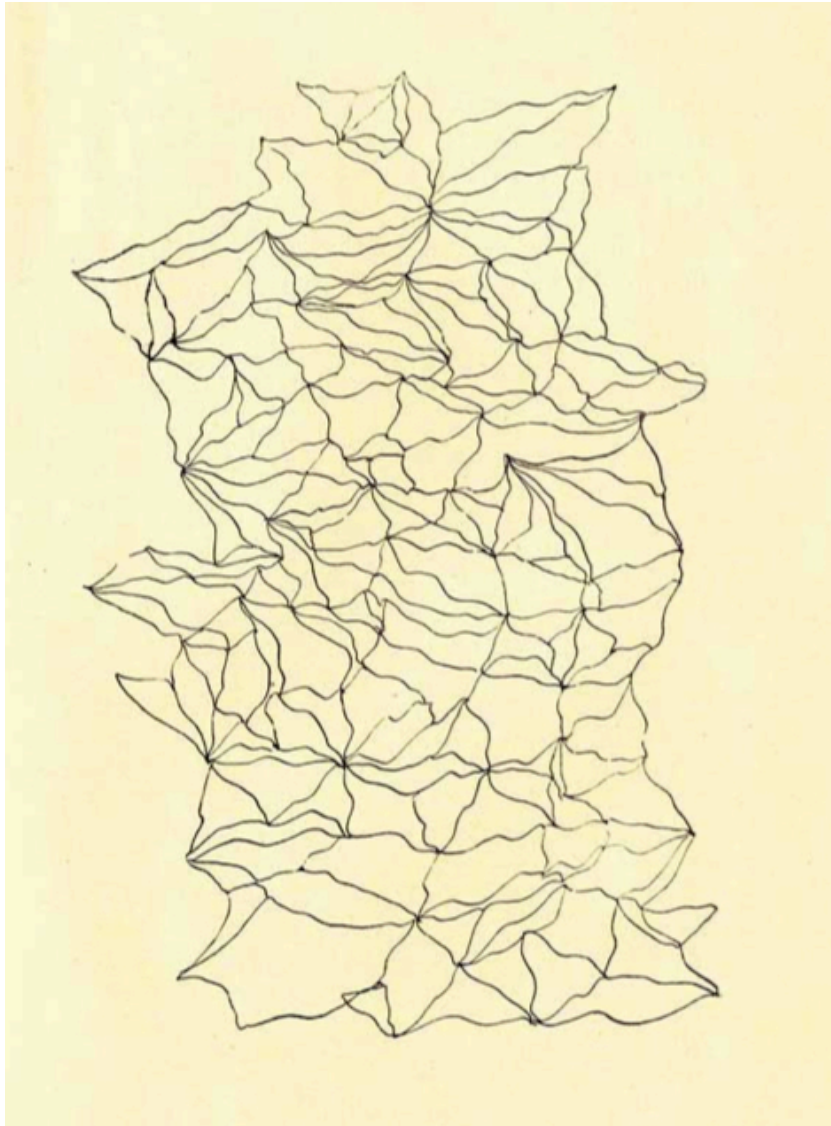


Figure: Enphonic Graphomania v1.1

Gesture, Mathematical Language (and Music?)

Brian Rotman
Ohio State University

That bodies speak has been known for a long time.” (Deleuze) Speech has multiple intimacies with gesture. Not merely physiological but, as a mode of thought-affect, gesture precedes speech (evolutionarily and developmentally), alternates with it (emblems), accompanies it (gesticulation), and is embedded inside speech (tone). Indeed, any kind of sense-making and expression -- film, architecture, theatre, dance, music – is likely to exhibit significant liaisons with gesture. Mathematics, a written language ‘spoken’ by the body, is no exception. Such is the contention of Gilles Chatelet, who locates gesture and its diagrammatic capture as the source of mathematical thought. The founding moves of geometry and arithmetic bear this out. My interest is in the fact that the abstract language of categories articulates mathematics precisely in terms of diagrams. Might categories therefore provide a framework for actualizing the virtual objects of mathematics, of realizing them gesturally, in physical -- visual or sonic or haptic -- form?

Dmitri Shostakovich's deformational sonata-form gestures and the interpretive implications

S. Reichardt

School of Music, University of Oklahoma

Dmitri Shostakovich clearly remained committed to the formal structures of the previous centuries. In his works composed between 1936-1964, Shostakovich consistently wrote movements in dialogue with eighteenth-century sonata-form archetypes. Yet, while referring to these archetypes his music breached historically defined structural boundaries; these deformational procedures can be understood as a means of creating new dramatic content. This paper will discuss rhetorical gestures Shostakovich consistently used that create ruptures within the normative form and the interpretive implications of his alterations of inherited archetypes.

In investigating Shostakovich's normative procedures with respect to sonata forms, two deformational forms become apparent—one used consistently in first movements and a second appropriate for finales. Michael Mishra has described the arch-sonata form that Shostakovich uses in the first movement of his Fifth Symphony as the composer's "trademark" form. While creating an arch form through reversal of the primary (P) and secondary (S) thematic material in the recapitulation, Shostakovich also consistently blurs major divisional boundaries, especially at the juncture of the development and recapitulation. Often the recapitulation emerges out of a developmental haze that is created using P material. The gesture of undercutting the structural boundaries between the development and recapitulation creates an emerging recapitulation. Yet, underneath the blurred divisional boundaries created by the gesture, lies a rotational cycle, enhancing the form beyond a simple arch structure. Thus, beneath the teleological arch-sonata lies a cyclical rhetoric. Despite Mishra's label of the arch-sonata as Shostakovich's "trademark" sonata form, Shostakovich reserves the use of this form for first movement sonata structures. His finale sonata structures have two significant deviations. First, a large-scale cyclical layer is added to the form, as themes from previous movements often return at the climax of the development in finales or allusions to previous material are made in recapitulations. Second, instead of an elided move from the development to the recapitulation, in finales Shostakovich composes a rhetorical gesture of disintegration. Here at the end of the development, after the climactic return of musical material from previous movements, the music will disintegrate to almost nothing, often a single pitch or even silence. The recapitulation is thus forced to literally restart the music. This gesture of disintegration has profound implications on the expressive meaning of the recapitulation depending on the music that follows. Shostakovich does not limit this rhetorical gesture of disintegration to finales, as he will also use the gesture, with an *attacca* marking, when connecting penultimate movements with finales. Using the gesture to bridge movements affects the expressive meaning of any large-scale cyclic returns found in the finale.

Clearly Shostakovich's creation of new musical gestures and manipulations of received gestures produce moments for the emergence of original expressive meanings. The rhetorical gestures provide a means for the alteration of the inherited structure of sonata form, which has provocative implications with respect to interpretation. Combined together, structural and rhetorical idiosyncrasies generate locations within the music that form a nexus for the interaction of expression, structure and meaning.

Heightened harmonic gesture in Wagner's music dramas

D. H. Smyth

College of Music and Dramatic Arts, Louisiana State University

No composer before Wagner so deftly united musical language and dramatic gesture. In "Opera and Drama," he described at some length the contribution of Gesture which for him included all manner of things visible, from appearance and demeanor to facial expression and movement to the full communication of that which was "unspeakable" through the workings of prose and music alone. He wrote "the most highly roused expression of the words [provided] the first incitement to intensify the gesture." In this paper, I explore the relationship between heightened dramatic gesture and extremes of harmonic disjunction in Wagner's mature music dramas. I contend that arresting gestures and the bold harmonic progressions that accompany them contribute significantly to the effectiveness of the drama as seen and heard in the moment and as experienced as a whole.

Anthony Newcomb's 1981 article "The Birth of Music out of the Spirit of Drama: An Essay in Wagnerian Formal Analysis" calls for reconciliation between those who (like Lorenz) place too much emphasis on the shaping forces of the music alone, and those who (like Dahlhaus) essentially deny the existence of forces capable of binding spans of any considerable length. Building on groundwork laid by David Lewin, a number of music theorists have recently offered studies of Wagner's harmonic language based on neo-Riemannian operations and parsimonious voice leading. While much of this work is valuable and engaging, most of it does not rise to Newcomb's challenge, largely because it frequently addresses only brief passages, and it often pays scant attention to the text and broader dramatic circumstances. Furthermore, these theorists tend to concentrate on continuities that allow music to cohere and flow smoothly.

This paper explores an alternative way of viewing some of the most shocking tonal disjunctions in Wagner's music, and relating these to distinctive dramatic gestures. It lays out a modified version of the neo-Riemannian Tonnetz which minimizes, rather than maximizes, common tone retention among adjacent triads. This allows comparison between triads and keys that lie at maximal distances as measured on the circle of fifths, and provides a simple metric for such comparisons. Of special interest are instances of what Richard Cohn has named "hexatonic poles" (E major and c minor, for example, lie seven "clicks" apart on the circle), which he has shown to be signifiers of the "uncanny." But Wagner employed other surprising harmonic gestures both at discrete points of high dramatic tension, and across huge time spans by virtue of recapitulated themes and episodes and the tonalities associated with them. The measurement of maximal dissimilarity among keys (as well as chords) provides a new perspective from which to consider the musical discontinuities Wagner so pointedly invoked to support the heightened dramatic gestures of his characters. Tracing examples of specific, scripted gestures that recur across long time spans will show how Wagner exploited the power of the unexpected upon both the eye and the ear.

Defaulting to gesture: How listeners extract meaning in the absence of cognitively accessible musical organization

R. S. Hatten
Indiana University

In this paper I presuppose the following:

1. Some 20th-century music lacks a stylistic pitch syntax (primarily defined as regularities or rules constraining syntagmatic relationships and connections among pitches, pitch patterns, and pitch structures).

2. Some 20th-century music possesses a compositional organization that is not cognitively entrained through listening (or audiation) in other words, listeners are not able to hear (or aurally imagine from the score) sufficiently to generalize (even tacitly) appropriate style types, principles, and processes; or there is too little hierarchical structuring to facilitate memory; or the time scale of events defeats one's ability to process similarities and differences; or the patterning one does hear is insufficient to support functional implications, etc.

In the absence of sufficient stylistic cues to understand pitch/durational music languages (those styles involving discriminable and memorable motivic/harmonic/contrapuntal/voice-leading/ melodic/metric/rhythmic patterning), listeners will nevertheless attempt to interpret expressive meaning by defaulting to the level of the sounding musical gesture, with gesture defined as "significant energetic shaping through time" (Hatten, *Interpreting Musical Gestures, Topics, and Tropes*, 2004). Through culturally-grounded iconic and indexical resemblances of musical gestures to the dynamics of emotions, actions, dramatic events, etc., listeners will attempt to

- a. understand characteristic affective content
- b. infer degrees and kinds of agency
- b. interpret the succession of gestures and their affective content, as distributed among various agencies, in terms of certain larger schemes such as dramatic trajectories (or narratives) motivating various sized spans of the composition, from phrases to entire movements or works.

Musical gestures in the absence of sufficient stylistic (systematic, rule-based, functional, and pattern-governing) constraints will of course be subject to a wider range of interpretation. In such cases, knowledge of the composer's intent, as revealed through titles, programs, treatises, aesthetic or compositional manifestos, and the cultural/historical context of the work's creation (and production/performance), may be crucial in guiding a more stylistically appropriate interpretation of musical gestures. And in some cases, the emergent effects of gestures may (even intentionally) supersede compositional or pre-compositional organization.

I will demonstrate these points with brief reference to examples chosen from the music of Messiaen, Boulez, Carter, Cage, Feldman, and Reich.

Whither gestural analysis?

P. McCreless
Yale University

Gesture is a word that we musicians—whether composers, performers, theorists, ethnomusicologists, or historians—use all the time, often unthinkingly, as a kind of default. It is one of those words that is indispensable to our musical discourse—a word that says something that a more rigorous theoretical language cannot, and thus a word that we are loath to give up. This is not to say that we do not need theories of musical gesture: we do, and a number of capable theorists have taken impressive steps toward same, as is abundantly evident in the papers from the two previous Music and Gesture conferences, in the book *Music and Gesture*, and in other published work of a number of prominent theorists.

The present paper is not about developing a formal theory of gesture. Rather, its focus will be on the rough-and-tumble world of our normal, everyday usage of the word as musicians and musical scholars. This world is captured nicely by the following simple exercise: go to the database that we all know as RILM, type in the word *gesture* as a keyword, hit search, and see what you get. As of late September of 2009, the exercise produces 1708 hits, and in a quick perusal of a few hundred of them, one encounters a staggering variety of topics, from elementary-level string pedagogy to Baroque musicorhetorical figures, to performance traditions in Verdi and Wagner, to “gesture control” in electroacoustic music, to gestural aspects of all sorts of musics around the world. It would, of course, be insane to try to make sense out of all this in a twenty-minute paper—or even a book or a set of books. But the exercise does raise a question worth pursuing: why is the word *gesture* so essential for us? What is the payoff of using it, rather than some more formal terminology? The strategy of the paper is to show, first, though only briefly, the explosion of topics produced by the RILM exercise—as a way, simply, of becoming aware of “what’s out there.” The paper will then quickly zero in on a small number of carefully chosen examples: from music theory and analysis, from composition of electroacoustic music, from popular music studies, and (perhaps) from ethnomusicology. A single example must suffice here. David Lewin, in an analysis of a passage from Brahms’s Horn Trio, in *Generalized Musical Intervals and Transformations*, graphs an intervallic/transpositional structure that he calls a “nuclear gesture with pickup.” What does Lewin gain by designating the phenomenon as a gesture, rather than, say, a motive, a structure, or an idea? It is the premise of the paper that a careful thinking through of how authors use the term, in specific instances, across a wide variety of situations, will offer clues as to why it is so ubiquitous, why we need it so, and how we might best theorize it.

Gesture and agency in responsive media

X. W. Sha

Topological Media Lab, Concordia University

Accounts of language, languaging, and the construction of knowledge have turned on logic, on semiotics, on information and cybernetics. Can we appeal instead to notions of embodiment and materiality that do not reduce to these other categories? Scholars such as Katherine Hayles and Mark Hansen have recently emphasized the need to rethink embodiment and materiality in an era saturated with digital media,² but attempts to reinterpret such notions have tended to slip back into the semiotic or informatic categories they try to exceed. Questions of gesture and agency in the presence of emerging technologies of performance have cast such notions in sharper relief. Given these concerns, it seems reasonable to expect that an account of gesture and agency informed by some direct experience with real-time computational media as applied to experimental performance could offer more insight to this inquiry. Moreover, since formal accounts of agency have tended to be framed within the categories of the cognitive, if not the linguistic, it seems that experimental projects with music, time-based media and performance would constitute fertile ground for phenomenological study.

Motivated by present concerns with embodiment and materiality, I pose the question: How do our notions of gesture and agency mutate in the presence of real-time, dynamically varying computational media? I refract this question through the experiences of building and playing gesturally controlled musical instruments, and generally the gestural control of musical pattern in corporeal movement and structured light fields, as well as sonic texture. (Of course this raises the phenomenological question of musical pattern.) What I bring into the conversation is a study of interaction and digital media, and some years of experience with building simulations, visualizations of differential geometric processes, and with responsive media spaces. My claim is that such responsive media spaces, both in their construction and in the experiences that they sustain, call into question linguistic and informatic models of gesture and open new ways to understand gesture and agency as embodied, a-linguistic experience. Part of the strategy has been to materialize the argument in the same responsive spaces that we study, and to reflect upon their design and their performance. Consequent to this approach I suggest a way to understand gesture, agency, and free play, and consider in what manner we can constitute material embodiment.

In this essay I first describe the environments that we have built over the past 9 years as prototypical examples of responsive media spaces, some gestural features of such environments, and the qualities of gesture on which I would like to focus, explaining why one may profitably defer resorting to linguistic categories. I then describe the materialization of gesture in physical movement and computationally mediated response, drawing for concreteness on the particular emergent technologies of computational media: real-time digital sound and music. And finally, based on this richer notion of open and multiply completed topological gesture, I draw a set of implications regarding agency in the presence of responsive media.

Movement prosody - Emergence of collective action

S. P. Gill

Computer Laboratory & Centre for Music and Science
University of Cambridge

In this paper, we explore the emergence of patterns of movement prosody that arise when more than one person shares a space of interaction with another, as a kind of musicality and dance. In linguistic analysis, there is the notion of movement of information carried out by cues of feedback such as 'uh huh', 'ok', 'ummm', the meaning of which is conveyed by how they are said as well as when (Carletta et al, 1996; Shimojima et al. 1997). Overlaps in speech may either cause interruptions requiring a person to give the other a turn at speaking, or not affect the speaker's turn [because that speaker is holding the floor]. An analysis of how our bodies engage and cue ways of understanding and providing feedback (Gill et al, 2000) show that the linguistic model of feedback and turn-taking is not sufficient as it is unable to account for continuous movement. Our bodies shape patterns of rhythmic synchrony as collective action, and the pull of one body with another (in synchrony and entrained) or withdrawal of one body from another (out of sync) gives us a sense of how we are making sense of each other and the nature of our commitment to be together. The most powerful form of synchrony is simultaneous motion, Condon said that this was so vital to human sociality that every culture has devised the short cut to it, in the form of greetings. Shaking someone's hand, touching noses, kissing, bowing, is awkward if it is not simultaneously synchronised. The qualities of music performance is considered in order to provide some understanding of rhythmic patterns in human interaction, as music is always made with others, whether it's a soloist rehearsing with an audience or a group of musicians. Unlike speech analysis that can be based on utterances and their placement in relation to each other, music is necessarily understood as taking place in time and meter and cannot be analysed outside of it. Some of the patterns of rhythmic synchrony one finds in body movement is akin to the accents and moments of crescendo found in music performance, and these seem to help us to maintain attention and reach a heightened state of emotion and resonance. There is a deep connection between body movement and music and we know that when we listen to music, motor neurons fire off in the brain (Hickok et al, 2003; Schneck and Berger 2006). In some languages in the world, there is no distinction between the word for music and that for dance. A lot of research has been undertaken on the relation between music and language (Cross and Woodruff, 2009; Rebuschat, P. et al, 2009), evidenced some believe in motherese (mother-baby talk) (Miall and Dissayake, 2003; Trevarthen and Malloch, 2009), where vocal gestures of exaggerated and melodic speech are musical in quality. The facial expressions are also exaggerated with the vocal sounds. The movements of sound and gesture are used to enculture the baby into the interaction patterns of sociality and culture, as collective and synchronised patterns. The body holds and expresses both sonic systems of music and speech, and this expression lies in collective and synchronized movements across people. This work seeks to attain a better understanding of how the body/gesture and sound move together in sense-making drawing upon relationships between music, gesture, and speech.

References

Carletta, J., Isard, A., Isard, S., Kowtko, J., Doherty-Sneddon, G. & Anderson, A. (1996). HRCRC dialogue structure coding manual. *Technical Report HCRC/TR-82*, Human Communication Research Centre, University of Edinburgh, June.

Cross, I. & Woodruff, G. (2009). Music as a Communicative Medium. In R. Both and C. Knight (Eds.) *The Prehistory of Language*. pp 113–144. Oxford: Oxford University Press.

Gill S. P., Kawamori, M., Katagiri, Y. & Shimojima, A. (2000). The Role of Body Moves Dialogue. In RASK: *Internationalt tidsskrift for sprog og kommunikation*, vol.12, pp.89–114.

Hickok, G., Buchsbaum, B., Humphries, C. & Muftuler, T., (2003). Auditory-Motor Interaction Revealed by fMRI: Speech, Music, and Working Memory in Area Spt. In *Journal of Cognitive Neuroscience*, 15(5):673–682.

Miall, D.S. & Dissayake, E., (2003). The Poetics of Babytalk. In *Human Nature*. 14(4):337–364

Rebuschat, P., Rohrmeier, M., Hawkins, J., Cross, I. (Eds.) (2009). *Language and Music as Cognitive Systems*. Oxford: Oxford University Press.

Schneck, D. J. & Berger, D. (2006). *The Music Effect: music physiology and clinical applications*. London: Jessica Kinglsey.

Shimojima, A., Katagiri, Y. & Koiso, H. (1997). Scorekeeping for Conversation-Contruction. In *Proceedings of Munich Workshop on Semantics and Pragmatics of Dialogue*, pp. 172–194.

Travvarthen, C. & Malloch, S. (2009). *Communicative Musicality: Exploring the Basis of Human Companionship*. Oxford: Oxford University Press.

Computer evaluation of musical gesture in sound structure

C. Dobrian
University of California, Irvine

Gesture is any motion that, by certain immanent characteristics, conveys meaning. As it is used in musicological discourse—by composers as well as theorists—gesture refers not so much to the physical action of a performer as to ways of characterizing musical content; the content itself implies motion, and that motion conveys characteristic meanings.

In the design of interactive computer music systems and the composition of interactive computer music, the tracking and analysis of these musical gestures—characteristic motions discerned within musical attributes—provides a promising challenge. I propose that there are in fact ways that one can clearly and empirically define and identify "gesture" in musical content, often with conceptual models and tools similar to those used for tracking and identifying physical gestures. The analysis of musical gesture as "meaningful motion" can be applied to many aspects of music: melodic contour, note speed and density, loudness, level of dissonance, etc., the shapes produced by measuring changes in these aspects, and derivation of data about change, rate of change, etc. within a particular feature or set of features. The existing techniques for tracking and analyzing the physical gestures of a performer can therefore be applied similarly for tracking and analyzing the gestural nature of the music itself. The new insights thus gained into the nature of musical gesture can be applied in interactive music systems to enhance the expressivity of computer music.

Such a study becomes culturally useful only insofar as it sheds light on musical understanding. Although tracking "motion" in musical attributes can be empirical, the analysis of, and attribution of significance to, these attributes is a necessarily subjective and interpretive process. Clearly the cultural background, cognitive skills, and emotional predisposition of the listener play a role, as do the aesthetic intentions of the composer and performer. Computer programs and computer programmers are neither neutral nor objective, and no one can claim knowledge of a "universal" lexicon of musical gestures.

Computer evaluation of gesture may be divided into tasks such as measurement, segmentation, identification, and taxonomy. What are the elements of musical gesture and how can a computer best discern them? How can a computer know when a gesture begins and ends? How can different, unforeseen gestures be compared and classified? How can subtleties within categories be evaluated? And perhaps most significantly, how can a computer, once it has identified a gesture, attribute meaning to it? This article proposes criteria and groundwork for the tracking, measurement, and analysis of "gesture" in the musical content of sound structure, and the use of that analysis in interactive computer music.

Key-postures and trajectories in music-related gestures

R. I. Godøy

Department of Musicology, University of Oslo

One major challenge in studying music-related gestures is that of understanding how we break down continuous streams of sound and movement into somehow meaningful units, into what we call *chunks*. Our answer to this challenge is to regard chunks as first of all centered on certain subjectively experienced salient moments in time, moments where we find what is called *key-postures*. The notion of key-postures is often used in film animation to denote salient postures in the flow of motion, implying that continuous movement trajectories are subordinate to these key-postures. In our context, key-postures denotes the position and shape of sound-producing effectors, e.g. the position and spread of the hands on a keyboard, the position and angle of the bowing hand-arm in relation to string instruments, the mallet-hand-arm in relation to percussion instruments, etc. at salient moments in time. These key-postures are synchronous with experienced salient events in the sound such as downbeats, other kinds of accents, or melodic or timbral-textural peaks. We hypothesize that these key-postures may serve as points of orientation both in the generation and in the perception of music-related gestures.

One core element here is to understand chunks of music-related gestures as holistically conceived and perceived units based on *anticipatory* cognition, meaning that we subjectively somehow have an image of the entire chunk in our minds at once. We believe there is now converging evidence for anticipatory elements in motor control in general (e.g. various models of pre-planning, the so-called *end-state comfort* effect, variants of the so-called *equilibrium point hypothesis*, etc) as well as in music performance. Furthermore, we hypothesize that the continuous gesture trajectories between these key-postures are subordinate to the key-postures. The trajectories before and after such key-postures can be thought of as respectively *prefixes* and *suffixes* to these key-postures, and are also often contextually smeared by the phenomenon of *coarticulation*. Notably, this is all on a rather local timescale with chunk durations typically between 0.5 and 5 seconds, similar to the duration range for sonic objects in Pierre Schaeffer's theory, but these chunks may of course be concatenated into longer sequences.

In our ongoing work on music-related gestures, we are trying to see how motion capture data of performers' sound-producing gestures can be fitted into this model of key-postures surrounded by prefixes and suffixes, looking both at the anticipatory and coarticulatory elements as well as at the velocity profiles of these movements with respect to the salient events in the sound, as suggested by other researchers' studies.

Analysis of musical gestures using Empirical Mode Decomposition

P. Toiviainen
University of Jyväskylä

The analysis of musical gestures recorded with a motion capture device often requires the use of some kind of decomposition method. Such methods are typically used to investigate hierarchical modes in music related movement, to decompose performer's gestures into instrumental and ancillary ones, or to examine synchronization between performers. To obtain a decomposition in the frequency domain, methods such as the Fourier transform and the Wavelet transform are frequently used. Similarly, typical decomposition methods in the time domain include Principal Components Analysis, Independent Component Analysis, and the Periodicity Transform. While all these methods can provide useful information about movement gestures, they have an inherent limitation related to the assumptions they make about the data. In particular, all these methods assume that the data to be analyzed be stationary and linear. A time series is stationary if its probability distributions do not change over time. Linear time series are generated by an underlying linear process, as a result of which they can be decomposed into a superposition of simple solutions (e.g. sine waves). The stationarity assumption can be relaxed to piece-wise stationarity by using frame-based analysis, but this approach is still limited by the uncertainty principle, that is, the tradeoff between frequency and temporal resolution.

Musical movement is often non-stationary and non-linear. For instance, a performer or listener may change their body posture in a non-stationary manner during the process of performing or listening. The aforementioned methods, therefore, may not always provide results that are sufficiently accurate. Empirical Mode Decomposition (EMD) is a data-driven, adaptive signal decomposition method that does not make any a priori assumptions about the stationarity or linearity of the signal. EMD decomposes the signal into a set of intrinsic mode functions that are amplitude- and frequency-modulated signals with zero mean, and which often have meaningful physical interpretations. These properties of the intrinsic mode functions enable precise quantification of instantaneous phase (Hilbert-Huang transform), thus allowing, for instance, accurate analysis of phase synchronization between performers or listeners.

The presentation discusses the principles of EMD and illustrates how it can be used to investigate various aspects of musical gestures, such as the presence of metrical hierarchy in spontaneous movement to music and synchronization between performers. Comparisons of EMD with other analysis methods will also be presented.

Study on gesture-sound similarity

B. Caramiaux, F. Bevilacqua, N. Schnell
IRCAM, CNRS-UMR STMS

Methodology

The dominant paradigm in the design of gestural sound control is based on mapping gesture parameters to sound synthesis parameters. The main difficulty resides in choosing an appropriate mapping strategy between low or high level parameters (Hunt & Wanderley, 2002). Our long-term goal is to propose adaptive methods to set such a mapping automatically from gestures performed while listening to a sound. Such methods require a co-analysis of the performed gesture and the listened sound.

This paper focuses on notions of similarity between gesture signals and audio signals. Quantitative analysis of gesture data together with audio data were proposed mostly based on second-order moments as correlation applied between sensor values and sound description parameters (see (Caramiaux et al, 2010), (Luck & Toiviainen, 2006)). These methods permit to highlight how a human can synchronize with the tempo and allows for the determination of the most important features in gesture in relationship to the sound. However, these methods suffer from important limitations since the relationships between sound and gesture are considered as linear or instantaneous.

To overcome such shortcomings, this study proposes to define similarity measure as a higher-order statistical measure between data. Inspired by previous works (Foote, 1997) in the Information Retrieval (MIR) field, we propose to use Mutual Information (MI) between signals. Actually, similarity measures will not be computed on the signal itself but on its probability distribution encoding higher-level aspects as degree of prediction. Inspired by this literature, we are interested in the extraction of the information content of gesture and sound.

Practically, the first step in the modeling is to estimate the probability distribution function of each signal. The gesture signal is typically the temporal evolution of kinematic variables and we use audio descriptors computed on a sliding window to describe the sound. The second step is the use of an appropriate divergence (e.g. Kullback-Leibler divergence) to measure how much one signal can be explained by the other. If the divergence returns zero or lower than a minimal ϵ -value (corresponding to an objective criteria), then the signals are assumed to be similar. On the contrary to correlation-based methods, here the gesture is no more constrained to be synchronized with the audio signal.

Results and Discussion

In this section we present experimental results obtained applying this methodology on real data. These data has been collected in May 2008 in the University of Music in Graz. For the experiment 20 subjects were invited to perform gestures while listening to a sequence of 18 different recorded sound excerpts. The sound corpus included a wide variety of sounds. The gestures were performed with a small hand-held device that

included markers for a camera-based motion capture system recording its position in Cartesian coordinates.

We consider sampled overlapping windows of the gesture signal and the audio descriptor signal, \mathbf{G}_i and \mathbf{D}_i where i is the time index (in samples) since the beginning. Each windowed signal is a stochastic process of N random variables. We compute their variance on the window: $(\sigma_1^k, \dots, \sigma_N^k)$ where k is the frame number. The resulting variance vector σ^k is considered as a set of independent and identically-distributed random variables on which we estimate the density function. Since we estimate the variance signals as Gaussian mixtures, we choose the KL-divergence to measure the similarity between the two probability density functions (see (Banerjee et al, 2005)).

Using the aforementioned data, we analyze the gestures performed while listening to three different sounds: a sound of an ocean wave; a solo flute (from Sequenza I for Flute by L. Berio); a sound of a crow's caw. These sounds correspond to three distinct morphologies. In conclusion, we find that the KL-divergence is minimized when we consider the gesture performed on one sound and the corresponding audio descriptor whereas its value is higher for a mixed case including a gesture or an audio descriptor from other sounds. Lastly, we compute the KL-divergence considering the audio descriptor of a sound (e.g. the wave) and the corresponding gesture performed by each candidate. This gives a similarity evaluation of each performance that we can compare with videos in order to discuss its pertinence.

References

- Banerjee, A., Merugu, S., Dhillon, I. S. & Ghosh, J. (2005). Clustering with Bregman divergences. *Journal of Machine Learning Research*, 6:1705–1749.
- Caramiaux, B., Bevilacqua, F. & Schnell, N. (2010). Towards a gesture- sound cross-modal analysis. In *Lecture Notes in Computer Science*. Springer Verlag, (to appear).
- Foote, J. (1997). A similarity measure for automatic audio classification. In *Proceedings AAAI 1997 Spring Symposium on Intelligent Integration and Use of Text, Image, Video and Audio Copora*, American Association for Artificial Intelligence.
- Hunt, A. & Wanderley, M. M. (2002). Mapping performer parameters to synthesis engines. *Organised Sound*, 7(2):97–108.
- Luck, G. & Toiviainen, P. (2006). Ensemble musicians synchronization with conductors gestures: An automated feature-extraction analysis. *Music Perception*, 24(2):189–200.

Encoding emotion: how performers manipulate tempo locally to convey affect

M. Swoboda, M. Wagner
McGill University

It is well established that we communicate emotion in music through the manipulation of rhythm (timing), timbre, stress and pitch (Frick 1985, Palmer 2006, Juslin 2000). This paper explores the hypothesis that there are particular parts of the score that are differentially affected in order to convey different emotions. For example, when instructed to play a piece either in a 'sad' or a 'slow' manner performers will in both cases play slower, but we hypothesize that the slowing down will be global in the case of 'slow', but in the case of 'sad' it will be stronger in 'zones of interest' that are apt to convey emotional affect.

In the present study, five highly trained musicians were asked to play five musical excerpts. Since we were interested to see whether generalizations can be made across instruments, we recorded three violins, one trombone, and one cello. Three excerpts were taken from well-known works in the classical repertoire, two were composed for this study. Each participant was instructed to play each excerpt with three different expressive styles: sad, happy and neutral. In a second phase, the participants were asked to re-record all five excerpts with two new expressive styles: neutral fast and neutral slow. Two zones of interest were annotated intuitively: an expressive zone, at the end of the excerpt at the strongest cadence point, and an unexpressive one, which consisted in three or four notes within the excerpt where little manipulation was expected. A range of acoustic variables including duration was automatically extracted. The acoustic data was normalized for duration prescribed by the score, such that only the variation in duration introduced by the performer could be observed. The hypothesis was that participants would lengthen the final part of each expressive zone in the sad interpretation and shorten it in the happy one.

We found that performers lengthened the expressive zone given any instruction, but that lengthening was significantly greater in the sad interpretation compared to either neutral, happy, fast or slow, and also compared to the 'unexpressive' region. Since none of the other interpretations showed any significant change in the expressive zone of interest, it stands to reason that other acoustic parameters are used to encode them. We are currently exploring which, if any, acoustic measures encode happiness, and which other parts of the score or other acoustic variables were modulated for sadness.

Preliminary results of a perceptual experiment wherein participants were asked to rate the recorded excerpts as happy or sad suggest a main effect of excerpt such that some were overall rated more happy or sad. This was likely due to the prominence of major and minor tonalities respectively. Furthermore, certain performers were more successful at conveying affect. This is an exciting result as it indicates that certain performance strategies for conveying affect are more effective than others. Further perception tests are under way to test whether the zone of interest was indeed driving the perception of emotion, as hypothesized.

Thinking Gesturally

Anthony Gritten
Middlesex University

Why is the concept of gesture so useful and popular? Besides acting, at least at first, as a welcome antidote to what-came-before (variously constructed into many a story of redemption-by-gesture), how has the veritable explosion of work on musical gesture managed to become a potent force for unification, integration, and synthesis across many sub-disciplines of Musicology, across methodological, technological, and aesthetic divides? What gives anybody the right to suggest, probably quite reasonably in March 2010, that the future of musical gesture studies is bright?

Scholarship on musical gesture has generally focussed on two interrelated constellations of concepts (sometimes in tension). One is the musical body, the other is music cognition. The musical body has emerged as a key theme on the back of two major developments: first, Musicology working through its positivist and Modernist heritages; secondly, the assimilation of the new discipline of Performance Studies, which has been slowly working its way beyond its initial theoretical grounding in Chopin and the (midi-)piano. The theme of music cognition has developed in leaps and bounds on the back of the general scholarly embrace of a technological infrastructure and the analytical opportunities afforded by digital technology. Considered (perhaps somewhat problematically) as a whole, scholarship on musical gesture continues to strive for better ways to negotiate a fundamental distinction that cuts across virtually all intellectual traditions and methodological biases: between acoustic sound and musical gesture. Scholarship seems to be heading towards a general understanding of why the concept of gesture is important in all forms and cultures of music making, tied up as it is with issues of agency and intention, and with concepts of creativity (whether or not the product is physically tangible).

As the above summary should make clear, the concept of gesture has been useful because it has contributed to a constellation of interrelated concepts that has included, *inter alia*, voice, authenticity, freedom, subjectivity, expression, autonomy, organicism, and interiority. Of particular note is the extraordinary quality of achievements that has been afforded to scholarship on musical gesture on the back of committed, intense, and sometimes apparently irreversible investments in this constellation of concepts, alongside equally imaginative insights set in motion by broadly deconstructive critiques of these same concepts. While it is noted here that there has been an essentially Modern quality to this approach to musical gesture, it is worth pausing to consider whether and how these satellite concepts orbiting around musical gesture continue to provide an appropriate image and model for musical practice in our digital, distributed, debt-laden, diasporic world of the first decade of the twenty-first century and beyond. Sooner or later, if scholars don't ask the question then creative artists will ask it: Does gesture still matter?

One way to appreciate the space opened up by this question is to ask a further question: What would music without gesture sound like? If gesture is that which affords us music (rather than merely sound), then what happens in its absence? At first glance, what we

would be left with would be not much more than a representation of the formalist note-spinning or displaced compositional theory common to the scholarship of earlier eras, or perhaps the emaciated figure of the performer portrayed in Stravinsky's aesthetics. In addition to this degree zero of musical practice, music-without-gesture would also be ergonomically efficient, quietly productive, and neatly contained within the categories invented for it, and that would have certain advantages. However, even these categories are associated with gesture; gestures are often described or modelled as efficient, productive, and contained. Indeed, we are so used to talking of gesture, whether as part of a neat theoretical construct or simply as an everyday part of what-music-just-is, that it has become something of a metaphor we live by, and has infiltrated almost everything we say about music, whether hermeneutic, aesthetic, analytical, cognitive, psychological, sociological, or anthropological. In part this may be because many of our definitions of gesture necessarily remain tantalisingly general: a gesture is "any energetic shaping through time that may be interpreted as significant" (Robert Hatten). It is also because gesture has developed its own grammar: we have gestures-as-nouns (the musical gesture in bar 5) and gestures-as-verbs (the violin gestures at the piano in an ascending motive), and we have developed multiple theoretical discourses that map gesture onto all other musical parameters (usually in terms of supervenience), that effectively rewrite earlier theories of music in gestural terms (hence the allusion above to narratives of redemption-by-gesture), and that – most importantly – are assimilated by all talk of what human expression, effort, or work is achieved by music.

There is another reason why gesture does still matter, and why it is hard to imagine music – indeed life in general – without gesture. This is that there is something in the concept of gesture that resists reduction to this-or-that, to relationships between parameters, properties, and qualities, to counting. There is something in gesture that works differently to the concepts mentioned above, that works both at a more fundamental level of activity (action and decision making) and at a more detailed level (the singularity of owned expressive presence). This moment of musical gesture is what makes music a human activity.

This talk will, on the back of an account of the triumphs of scholarship on musical gesture alluded to above, unpack this moment of musical gesture, moving from the constellation mentioned above towards a constellation of concepts that includes looseness, inefficiency, expressivity, resonance, mobility, resistance, energy, intensity, and presence. It will be argued that gesture is the site and vehicle for the transfer of energy between domains (Hatten's "energetic shaping"), and that, as such, gesture should be thought of as the entropic loophole of music making: that event through which, at which point, and by means of which music happens, and in consequence of which we are afforded and enjoy those multifarious activities that we describe as 'musical', whether compositional, performative, perceptual, critical, or all of the above. This moment of energetic shaping, of entropic affordance (or however it is categorised theoretically), deserves the appellation, 'thinking gesturally', for it is what draws us into music in the first place.