Lars Graugaard
Departement of Software and
Media Technology
Aalborg University Esbjerg
Niels Bohrs Vej 6
DK-6700 Esbjerg
Denmark
lag@cs.aaue.dk

Applying a Performer's Physical Gestures to Sound Synthesis in Real-Time

Abstract

Motivation and strategies for affecting electronic music through physical gestures are presented and discussed. Examples of such usage in practise are reported, and the results and future possibilities are discussed.

Introduction

Musical expression has strong ties with physical gestures for different purposes and in different contexts. A conductor communicates musical parameters through gestures, and chambermusic performed without a conductor relies heavily on information passed on among the performers through gestures for coordination of rhythm, dynamics, and phrasing. A player involuntarily elicits gestures as a result of the combination of the music's expressivity and its performance needs, and the audience processes that as a complementary component to the listening experience.

This paper presents motivations and strategies for including a performer's gestures as part of the music performance, at times as the primary link between the performer and the sounding music. The particular case of one performer and an interactive electronic accompaniment is the basis of the studies.

Gesture, music, and expression

Connections between human gesture, expression, and music are manifold and tangled, and a general practise for its use has yet to come about.

A substantial number of systems for detecting movement and transforming it into sound has been developed for artistic purposes (Radio Baton, Very Nervous System), entertainment purposes (Personal Orchestra, Home Conducting, Musini, EyeToy), and empowering and rehabilitating purposes (SoundBeam), though not all have been put into production (DJammer).

Several reports have been made on interpreting human gesture for its expressive qualities (Choi 1998). Antonio Camurri has conducted much work since the late 1980's, mostly in the area of dance-music intermodalities through recognition of non-symbolic, expressive data from human movement and gesture (Camurri et al. 1999). The Kansei concept of information processing forms the background of a large body of this work (Hasimoto 1997). This, combined with Rudolf Laban's effort space (Laban 1963) has permitted the description of a kinematics-space that may function as a link between expressive content of different communication channels, such as music and human gesture and dance, etc. (Canazza et al. 2001). Anders Friberg describes a performance system where human sound and gestures are combined in influencing the output of an interactive system (Friberg 2004) and controlling expressive musical details of a score through a fuzzy logic system (Friberg 2005). A set of full-body gestures and patterns, such as jumping etc., is categorized where the users influence the system by performing the gestures at varying intensities.

Musical expression is commonly described through narrative analysis, but this does not contain descriptive and predictive powers sufficient to translate into a machine tool. Leman and Camurri lay out a three-way connection between signal analysis, expression experience, and the associated narrative (Leman 2004). The authors mention the possibility that action expressiveness could be related to the mind's so-called mirror-neurons, by which huminids encode audio-visual actions, later to be discharged during action or observation of the same action. This means that emotion is stored and recognized as precise entities, important in social interaction, survival, a.o.

Connections between the emotion content of a musician's movement patterns and the emotional expression of the music performed have been investigated, where relational sets of movement/emotion cues for expression decoding were found (Dahl et al. 2004).

Enacting performers

Musicians' presence on the concert stage has been used on many occasions to enhance musical content and in attempts at redirecting current artistic aims. In recent years music theatre has come about as a particular sub-cross-genre of opera and regular concert form with theatre and mostly executed by musicians. In particular, opera can be understood as an embodied art involving human action and not reducable to its acoustic dimension only (Heile 2006). Music theatre also holds this property, and the Argentinian/German composer Mauricio Kagel has been foremost in establishing its prominent elements. Many other composers have contributed important work to the practise of utilizing the scenic presence of performers, such as Karlheinz Stockhausen, George Crumb, John Cage, Helmuth Lachenmann, and György Ligeti. The aims have been very varied, from representing a musical mini-drama or providing the music's proper atmosphere, to rebelling against the traditional concert form in subtle and not-so-subtle ways. In short, it is fair to say that the presence of the performer on stage today can be used freely as an integral part of the performance, since the tradition to do so has been firmly established.

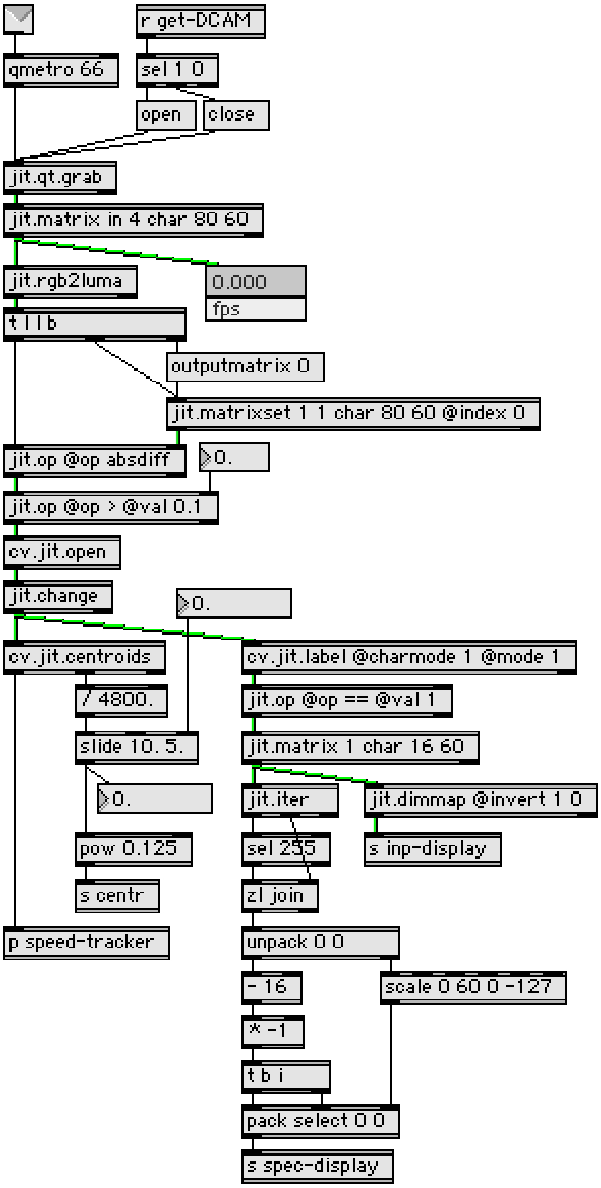

Figure 1*; Gestures You Made – code example of video analysis*

Sound-shaping performers

Quite a lot of work has been done interpreting dancers' movements for expressive content through video detection, for subsequent mapping onto musical parameters. David Rokeby's *Very Nervous System* from 1990 is one of the first real-time usages, and much simpler yet effective approaches are found in Christian Möller's *Electro-clips*. Today, installations commonly use human movement data, but the fact that they often use museum guests and similar non-expert users moving about in public space means that the data analysed is much less expressive and sophisticated, implying other strategies and aims for its use (Rebelo et al. 2005).

Figure 2*; Gestures You Made – music example*

CASE 1: Gestures You Made

My composition *Gestures You Made* for oboe and interactive computer is mainly interactive through sound analysis (Graugaard 2005/06). But towards the end of the piece the player stops playing at moments, and instead affects the sound synthesis by assuming body positions and making physical gestures inside a 'videospace' adjacent the performance space. The result can be likened to an externalized inner dialogue of the performer, who comments on own playing through physical gestures, for then to comment on these gestures through playing. The link between the two affect channels is the parallelism between expression in sound and in human gesture. The parallelism is underscored by the fact that the two channels affect the same sound, which at times is done simultaneously through playing and physical gesture. It is straight-forward that playing in interactive music can produce and affect electronic sound, but it is interesting to note that the production and affect of sound through physical gestures are just as readily accepted as a valid connection in the audience's perception, once the initial surprise has ended.

The performer's movement is detected by measuring the difference between subsequent video frames. *XY*-position of the on-pixels from the motion detection having an off-pixel immediately above – the upper part of the movement mass – is used to affect the spectral envelope of the sound synthesis (fig. 1).

The sound synthesis uses a custom-made frequency modulation method based on (Jensen 1999) consisting of 32 components spaced at *n*f(0)* with the high-level attributes modulation strength, randomization bandwidth, and component correlation. The spectral envelope is detailed as a list of amplitudes of the 32 sound synthesis components. The fundamental *f(0)* of the instrument is mapped to the fundamental *f(0)* of the sound synthesis according to the equation

$$y = \left(\frac{x-a}{b-a}\right)^e * (d-c) + c \qquad \textbf{(1)}$$

where a and b are input range, c and d are output range, and e is curvature factor. The input range is updated as the instrument restricts its range for a specified time, resulting in self-adjustment of optimal expressive range. The instrument's running envelope is mapped onto values of strength, randomization, and correlation of the sound synthesis chosen by trial-and-error, to provide sufficient timbral variation.

The section in which the motion detection occurs is based on improvisatory transitions between a sequence of motivic cells (fig. 2). The cells are based on material from earlier parts of the composition, and the gestural comments are inserted freely by the performer after each completion of 1-3 cells.

The mapping produces an intuitive connection between movement, playing, and sound synthesis, where the player perceives that higher body positions result in more energy in the spectrum (and at the frequencies according to the left-right position), higher notes from the instrument produce higher-pitched sounds, and variation in dynamics produce a varying timbre. The intuitive connection is also readily perceived by the audience, because the result is very much a one-to-one relationship of the overall sonic gesture, in spite of the variations caused by the sound analysis.

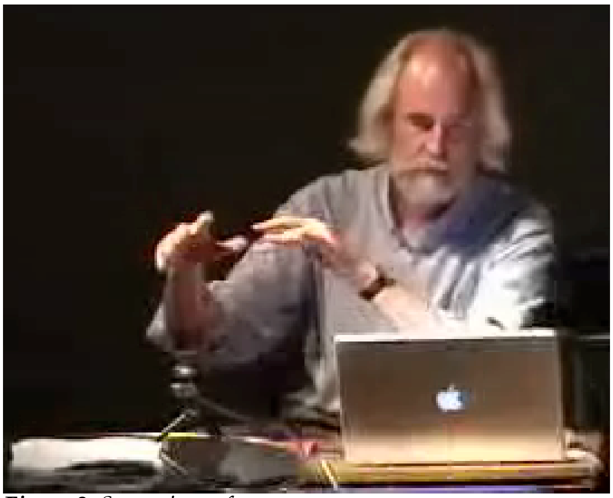

Figure 3*; Stgo – in performance*

CASE 2: Stgo

I am playing music on my laptop in Stgo (Graugaard 2005), and I use hand gestures to affect the sound of the music (fig. 3, 4). Laptop performance has developed its own performance practise of an apparently very inactive performer participation, bordering on the enigmatic (Cascone 2003). Informing laptop music with body gestures may take laptop performance out of this attitude towards the concert situation – or it may make it even more enigmatic and ritualistic – but the artistic reason for me to include video tracking was that it would make it possible to instill human expressiveness into procedural electronic music in real-time. The same motion detection technique is used for measuring the performer's hand movements. Since no sound input is involved it was important to find other means to track relevant movement data. Movement speed seemed to be a relevant factor, and displacement speed of the centroid of the motion mass was chosen to determine *f(0)* of the sound synthesis, so that more movement from the performer would produce higher values of *f(0).* Short amplitude envelopes are applied to the sound, and mass of the movement determines the width of these sound grains so that a bigger mass produces longer grain durations. *XY*-position of on-pixels having an off-pixel above – the upper part of the mass of movement – is used to affect the spectral envelope of the sound synthesis (fig. 1).

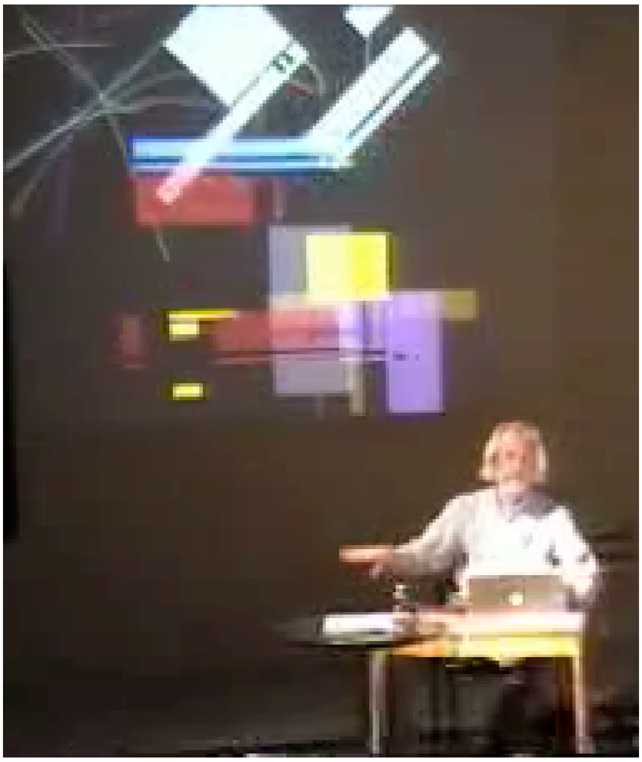

Figure 4*; Stgo – in performance, visuals by Ricardo Vega*

The sensation of shaping the sound is strong, certainly because the movements and the resulting sound is easily related, but also very much because the hands are our primary shaping tool. We shape physical objects by means of them, and the imaginary shaping of sound in a similar fashion is directly translated and readily felt. The energy of the sound spectrum is easily shapeable by hand movements in the *XY*-field, and rapid hand movements produce higher-pitched sounds, while movement with open hands facing the camera – or movements on the *Z*-axis – produce longer/shorter sound grains. Consequently, the audience readily perceives the hands' shaping of the sound even though not all relations in the sound shaping are directly traceable.

Conclusions

Informing sound synthesis with human gestures in real-time is a very potent performance enhancement technique. It adds a dimension to the performance where the visual presence of the performer extends into the sound in a very direct and unambiguous way and enhances the music's presence to envelop a physical presence. This is in effect a multi-modal interactive music system with a noteworthy musical potential, because an informative channel is added to the synthesis process or high-level proceedings that is not easily established through other means.

Tracking categories of hand positions and patterns of hand gestures would make good sense, particularly in the case of laptop performance. This could make it possible to signal to the computer what to do by a simple sign language, for instance switching between modalities for further detailed affecting.

A future step would be to support affect description, but few results with extraction and use of clear, subjective qualities in real-time affective systems have so far been presented. Paradigms supporting a broad palette of affect descriptors will have to be developed before such work can be undertaken.

The hand gesture interface could be made as an independent application with mapping of hand gestures into a normalized, multidimensional space, then to be mapped onto the sound synthesis or high-level music descriptors of choice. Experiments have been successfully carried out here (Modler et al. 2003) where simple actions were controlled, such as soundfile playback and volume. (Mulder et al. 1997) propose a shape manipulation task called sound sculpting, where shape description parameters are mapped to timbral parameters in order to reduce the cognitive load of simultaneous multidimensional control tasks, such as music-making.

Some difficulties were encountered with webcams due to their 15 frames/second data output limit. This meant that techniques relying on fast and precise tracking would become imprecise, such as displacement speed. The advantage of using cheap webcams is that the techniques become readily accessible and it seems that other techniques must be explored that will not be hampered by the low frame rate and frame size of webcams.

References

Camurri, A., Trocca, R., Volpe, G., 1999: "Interactive systems design: A KANSEI-based approach", *Proceedings of NIME2002*; Dublin, Ireland.

Canazza, S., De Poli, G., Rodá, A., Vidolin, A., Zanon, P., 2001: "Kinematics-energy space for expressive interaction in music performance", In Proceedings of MOSART, Workshop on current research directions in Computer Music; pp. 35-40; Barcelona, Spain.

Cascone, K., 2000: The Aesthetics of Failure: ""Post-Digital" Tendencies in Contemporary Computer Music", *Computer Music Journal*, 24:4, pp. 12-18, Winter 2000. Massachusetts Institute of Technology.

Choi, I., 1998: "From Motion to Emotion: Synthesis of Interactivity with Gestural Primitives", *Emotional and Intelligent: The tangled Knot of Cognition*, AAAI Fall Symposium, Orlando FL. (October), pp. 22-25.

Dahl, S. and Friberg, A., 2004: "Expressiveness of musician's body movements in performances on marimba", In A. Camurri, G. Volpe (eds.) *Gesture-based Communication in Human – Computer Interaction*, LNAI 2915 (pp.479-486), Berlin, Springer Verlag.

Friberg, A., 2004: "A fuzzy analyzer of emotional expression in music performance and body motion", *Proceedings of Music and Music Science*, Stockholm, October 28-30 2004.

Friberg, A., 2005: "Home conducting – control the overall musical expression with gestures", *Proceedings of the International Computer Music Conference 2005*, pp. 479-482; Barcelona, Spain September 5-9 2005.

Graugaard, L., 2005: "Stgo", Premiered December 6th 2005 by Lars Graugaard; Cech Santiago (Chile).

Graugaard, L., 2005/06: "Gestures You Made", Commissioned and premiered February 12th 2006 by Eydís Franzdottír; Dark Music Days 2006, Reykjavik (Iceland).

Hashimoto, S., 1997: "KANSEI as the Third Target of Information Processing and Related Topics in Japan", In A. Camurri (Ed.): *Proceedings of the International Workshop on KANSEI: The Technology of Emotion*; pp. 101-104, Gothenburg, Sweden

Heile, B., 2006: "Recent Approaches to Experimentral Music Theatre and Contemporary Opera", *Music and Letters*, Vol. 87 No. 1 pp.72-81; 2006, Oxford University Press.

Jensen, K., 1999: *Timbre Models of Musical Sounds*. PhD Dissertation, Report no. 99/7, Department of Computer Science, University of Copenhagen; ISSN 0107-8283.

Laban, R., 1963: *Modern Educational Dance*. Macdonald & Evans Ltd. London; ISBN 0712113819.

Leman, M., Camurri, A., 2004: "Musical Content Processing for Expressive Gesture Applications in Interactive Multimedia", *Proceedings of the*

Conference on Interdisciplinary Musicology (CIM04); Graz, Austria.

Modler, P., Myatt, T., Saup, M., 2003: "An Experimental Set of Hand Gestures for Expressive Control of Musical Parameters in Realtime", *Proceedings of the 2003 Conference on New Interfaces for Musical Expression* (NIME-03), Montreal, Canada.

Mulder, A., Fels, S., Mase, K., 1997: "Empty-handed Gesture Analysis in Max/FTS", *Proceedings of the Kansei - The Technology of Emotion Workshop*; Genova - Italy.

Rebelo, P., Alcorn, M., Wilson, P., 2005: "A stethoscope for imaginary sounds: interactive sound in a health care environment", *Proceedings of the International Computer Music Conference 2005*, pp. 829-832; Barcelona, Spain September 5-9 2005.