Gesture and Emotion in Interactive Music: Artistic and Technological Challenges

Lars Graugaard

PhD on Published Work

Oxford Brookes University

March 2006

In partial fulfilment of the PhD degree

INCLUDED PARTS **4**

PAPERS 4

WORKS 4

ABSTRACT **6**

ACKNOWLEDGEMENTS **7**

GLOSSARY **8**

1. INTRODUCTION **9**

1.1 MUSIC, COMPUTERS, AND INTERACTION 9

1.2 MUSIC, EXPRESSION, AND EMOTION 12

1.3 TRACKING EXPRESSION AND INTENT 13

1.3.1 SOUND TRACKING *15*

1.3.2 HUMAN GESTURE TRACKING *16*

1.3.3 TRACKING PHYSIOLOGICAL SIGNALS *17*

1.4 THE SOUND, GESTURE, AND PHYSIOLOGY AMALGAM 18

2. THE INTERACTION INCENTIVE **21**

2.1 THE INTERACTING MUSIC SYSTEM 21

2.2 AN INTERACTIVE NARRATIVE OR AN INHABITED WORLD? 22

2.3 INTERACTION, EXPRESSION, AND EMOTION 23

2.4 FURTHER CONSIDERATIONS 24

3. BALANCING THE REQUIREMENTS **26**

3.1 (GRAUGAARD 2004): FORMAL CONSEQUENCES AND POSSIBILITIES 26

3.1.1 The interaction time-space in GUITAR *26*

3.2 (GRAUGAARD 2005A): UTILITARIAN AND EXPRESSIVE ACCOMPANIMENT 27

3.2.1 Notated and electronic score in la quintrala *28*

3.2.2 Audio analysis and pitch-set comparison *30*

3.2.3 Output triggering *30*

3.2.4 Expressivity and presence enhancement *32*

3.2.5 Compositional consequences *32*

3.3 DISCUSSION 32

4. VISUAL GESTURE DETECTION IN MUSIC **33**

4.21 (GRAUGAARD 2006A): USING HUMAN GESTURES FOR AFFECTING MUSIC 33

4.1.1 Motivation *33*

4.1.2 Gesture analysis and sound synthesis *34*

4.1.3 Cases *34*

4.1.3.1 Stgo 35

4.1.3.2 Gestures You Made 35

4.2 DISCUSSION 35

5. PHYSIOLIGICAL SIGNAL DETECTION AND SOUND SYNTHESIS **37**

5.1 (GRAUGAARD 2005B):

SENSING EMOTION AND INTENTION THROUGH PHYSIOLOGICAL SIGNALS 37

5.1.1 Data extraction and analysis 37

5.1.2 System 38

5.1.3 Formal and intuitive 38

5.1.4 Sound and video synthesis 39

5.1.5 A user cycle 41

5.2 DISCUSSION 41

6. CONCLUSIONS 42

6.1 MAIN CONTRIBUTIONS 42

6.2 SPECIFICS 42

6.3 CIRCUMSCRIBING DIGITAL MUSIC SYSTEMS 43

6.4 THE ARTISTIC POTENTIAL 44

7. REFERENCES 46

APPENDIX I. INCLUDED PAPERS 60

INCLUDED PARTS

The dissertation consists of a critical appraisal and the following parts. The papers are identified by their roman numeral in the reference section.

PAPERS

(Graugaard 2004)	**(I)** Graugaard, L.: Open and Closed Form in Interactive Music. In U. Kock Wiil (ed.) *Computer Music Modeling and Retrieval, Second International Symposium CMMR 2004*; Esbjerg, Denmark, May 26-29 2004; Revised Papers; pp. 149-157; Springer Verlag Berlin Heidelberg; ISBN 3-540-244858-1
(Graugaard 2005a)	**(II)** Graugaard, L.: Unifying Performer and Accompaniment. In R. Kronland-Martinet, T. Voinier, and S. Ystad (Eds.): *Computer Music Modeling and Retrieval 2005*, LNCS 3902, pp. 169-184, Springer-Verlag, Berlin Heidelberg 2006
(Graugaard 2005b)	**(III)** Graugaard, L.: The SoundGlove II: using sEMG Data for Intuitive Audio and Video Affecting in Real-time. In *Proceedings of The Third Annual Conference in Computer Game Design and Technology*; 8th-9th November 2005, Liverpool, UK

WORKS

(Graugaard 2001)	Graugaard, L.: *GUITAR*. Composition for guitar and interactive, computer generated accompaniment. Premiered March 3rd 2001 at Carl-Nielsen Academy of Music, Denmark. Score, CD recording, programme application. (**D**)
(Graugaard 2003-04)	Graugaard, L.: *La Quintrala*. Opera for five singers and interactive, computer generated accompaniment. Commissioned by Den Anden Opera (Denmark), Oper an der Leine (Germany), and Norrbotten Opera (Sweden). Premiered September 2nd 2004 at Den Anden Opera, Copenhagen, Denmark. Score, DVD recording, programme application. (**D**, **2**)

(Graugaard 2004-05) Graugaard, L.: *The SoundGlove II*. Installation for sensor glove, 8 channel souind and video. Commissioned by Museum of Modern Art, Roskilde (Denmark). CD recording, programme application with performance instructions. (**D, II**)

(Graugaard 2005c) Graugaard, L.: *Stgo*. Premiered December 6th 2005 by Lars Graugaard; Cech Santiago, Chile. Commissioend by Cech. CD recording, programme application with performance instructions. (**D**)

(Graugaard 2006c) Graugaard, L.: *Gestures You Made*. Composition for oboe and interactive, computer generated accompaniment. Commissioned by Eydís Franzdottír. Premiered February 12th 2006 by Eydís Franzdottír; Dark Music Days 2006, Reykjavik, Iceland. (**D**)

ABSTRACT

This dissertation presents a new and expanded context for interactive music based on Moore’s model for computer music (Moore 1990) and contextualises its findings using Lesaffre’s taxonomy for musical feature extraction and analysis (Lesaffre et al. 2003). In doing so, the dissertation examines music as an expressive art-form where musically significant data is present not only in the audio signal but also in human gestures and in physiological data. The dissertation shows the model’s foundation in human perception of music as a performed art, and points to the relevance and feasibility of including expression and emotion as a high-level signal processing means for bridging man and machine. The resulting model is multi-level (physical, sensorial, perceptual, formal, expressive) and multi-modal (sound, human gesture, physiological) which makes it applicable to purely musical contexts, as well as intermodal contexts where music is combined with visual and/or physiological data.

The model implies evaluating an interactive music system as a musical instrument design. Several properties are examined during the course of the dissertation and models based on acoustic music instruments have been avoided due to the expanded feature set of interactive music system. A narrowing down of the properties is attempted in the dissertation’s conclusion together with a preliminary model circumscription. In particular it is pointed out that high-level features of real-time analysis, data storage and processing, and synthesis makes the system a musical instrument, and that the capability of real-time data storage and processing distinguishes the digital system as an unprecedented instrument, qualitatively different from all previous acoustic music instrument. It is considered that a digital system’s particular form of sound synthesis only qualifies it as being of a category parallel to the acoustic instruments categories.

The model is the result of the author's experiences with practical work with interactive systems developed 2001-06 for a body of commissioned works. The systems and their underlying procedures were conceived and developed addressing needs inherent to the artistic ambitions of each work, and have all been thoroughly tested in many performances. The papers forming part of the dissertation describe the artistic and technological problems and their solutions. The solutions are readily expandable to similar problems in other contexts, and they all relate to general issues of their particular applicative area.

ACKNOWLEDGEMENTS

I would like to thank Dai Griffiths and Paul Dibley for guiding me through the requirements of the dissertation, Jens Arnspang for giving me the possibility to do it, Robert Rowe for putting it all into context, and Cort Lippe for setting the whole thing off.

My thanks also go to Barry Moon, Marc Leman, Peter Ingwersen, Kristoffer Jensen, Andrea Valente, and Hanne Schneller for helping me in many different ways with specific questions and issues during the course of the work. I also want to thank the many excellent musicians and singers who have performed the compositions included in the dissertation as well as the persons and institutions who have commissioned them.

Finally I'd like to thank Asunción Claro – to whom this dissertation is dedicated – for making it all worthwhile.

Cover art by Asunción Claro.

GLOSSARY

HMI	Human-machine interaction
NURBS	Non-uniform rational b-spline surface
PDS	Primary data stream
SDS	Secondary data stream
HLA	High-level attributes

1. INTRODUCTION

The computer has emerged as an instrument capable of interacting in real-time with a music performer and produce independent musical expression. The context of interactive music is the interaction between a music performer – a singer or an instrumentalist – and a digital system. Real-time analysis of qualitative aspects embedded in the musical performance data is most often used, but other types of performance data acquired in a variety of ways have also come into use as input to interactive systems. The content is analysed according to a variety of criteria at the physical, sensorial, perceptual, formal, and expressive level. The resulting artworks are vastly different, and can take the form of new digital instruments, adaptive hyper-instruments, on-stage real-time multimodal environments, interactive dance/music systems, and interactive museum exhibits to name a few.

The present dissertation is based on such works by the author:

- GUITAR (2001), commissioned by Carl-Nielsen-Academy of Music (Denmark),
- La Quintrala (2003-04), commissioned by Den Anden Opera (Denmark), Oper an der Leine (Germany), and Norrbotten Opera (Sweden),
- The SoundGlove II (2004-05), commissioned by Museum of Contemporary Art, Roskilde (Denmark),
- Stgo (2005), commissioned by Cech (Chile), and
- Gestures You Made (2005-06), commissioned by Eydís Franzdottír (Reykjavik, Iceland).

1.1 MUSIC, COMPUTERS, AND INTERACTION

Analysing and synthesising music have been attempted since the early days of computers, and scientific research and musical endeavours into sound and music have been intertwined from the start (Gabor 1947, Xenakis 1971). The musical qualities of a piece of computer music originally had to be embedded by the composer at the moment of composing in independent time. The music was stored in fixed form containing the musical and expressive qualities rendered at the moment of committing to media. Subsequent playback does not affect this content, and there is no understanding of human expression present at the reproduction stage. Computer music is highly interdisciplinary (Moore 1990) and closely related to developments in computer hard- and software, digital signal analysis, artificial intelligence, psychoacoustics, perception, and music cognition

(Fig. 1-1).

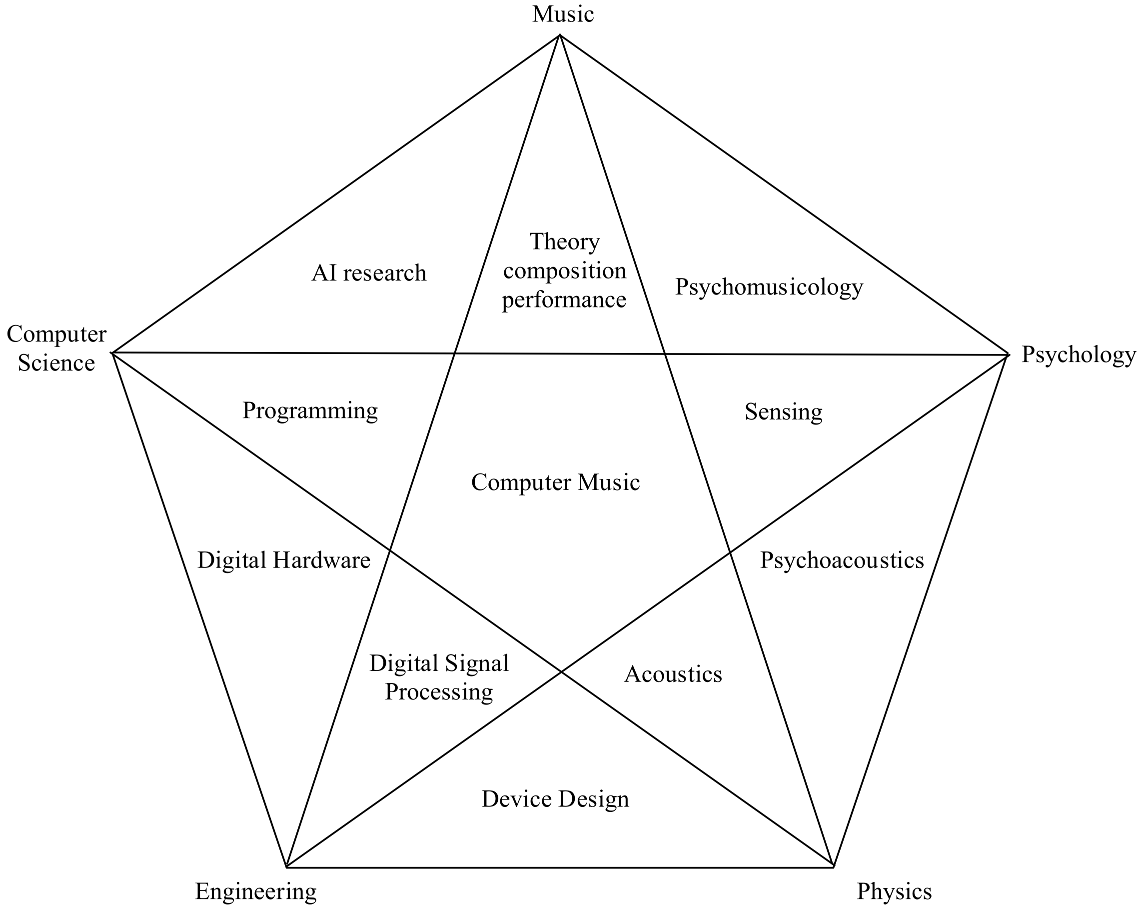

Fig. 1-1 *Moore's disciplinary context of computer music (Moore 1990).*

The more recent concept of interactive music requires that sound analysis and synthesis take place simultaneously and in real-time on a computer in response to and/or affected by a musical performance. The compositional task of the composer includes a programming task of preparing for such interaction, while observing that the interaction takes place in accordance with the compositional arguments of the work. This permits the music rendering to be affected by some aspect of the performance, typically perceptual aspects of the sound, or high-level musical features.

The term *interactive* describes the functionality of the computer rather than the music itself, since music is in its nature interactive, being a performing art. Interactive music consists of a set of performance instructions (in notated or verbal form or some combination thereof) and a digital component which makes it up for a digital instruction set, or a digital 'score'. All relevant musical information is no longer exclusively contained in the performance instructions and in the performer's understanding of the ramifications of their execution, because the computer takes independent part in the music performance, capable of affecting its low, mid, and high levels (Graugaard 2004). The interdependency of these two instruction sets defines the nature and content of the musical work, its limitations in execution, and its characteristics in performance.

As these areas have developed, so has the composer-programmers' possibility to explore and apply unique and distinct musical propositions different from a purely acoustical music composition. Subsequent areas that consequently have been drawn into the disciplinary context of interactive music are interactive storytelling and emotion analysis and synthesis. If Moore's diagram for the disciplinary context of computer music is modified accordingly and targeted at interactive music, a complex web of highly sophisticated areas going into the making of interactive music appears (Fig. 1-2). As a consequence, human-machine interaction (HMI) in the area of music has opened up for artistic experiments at the fringe or well outside the traditional areas of musical thinking and practice of conservatories and music departments.

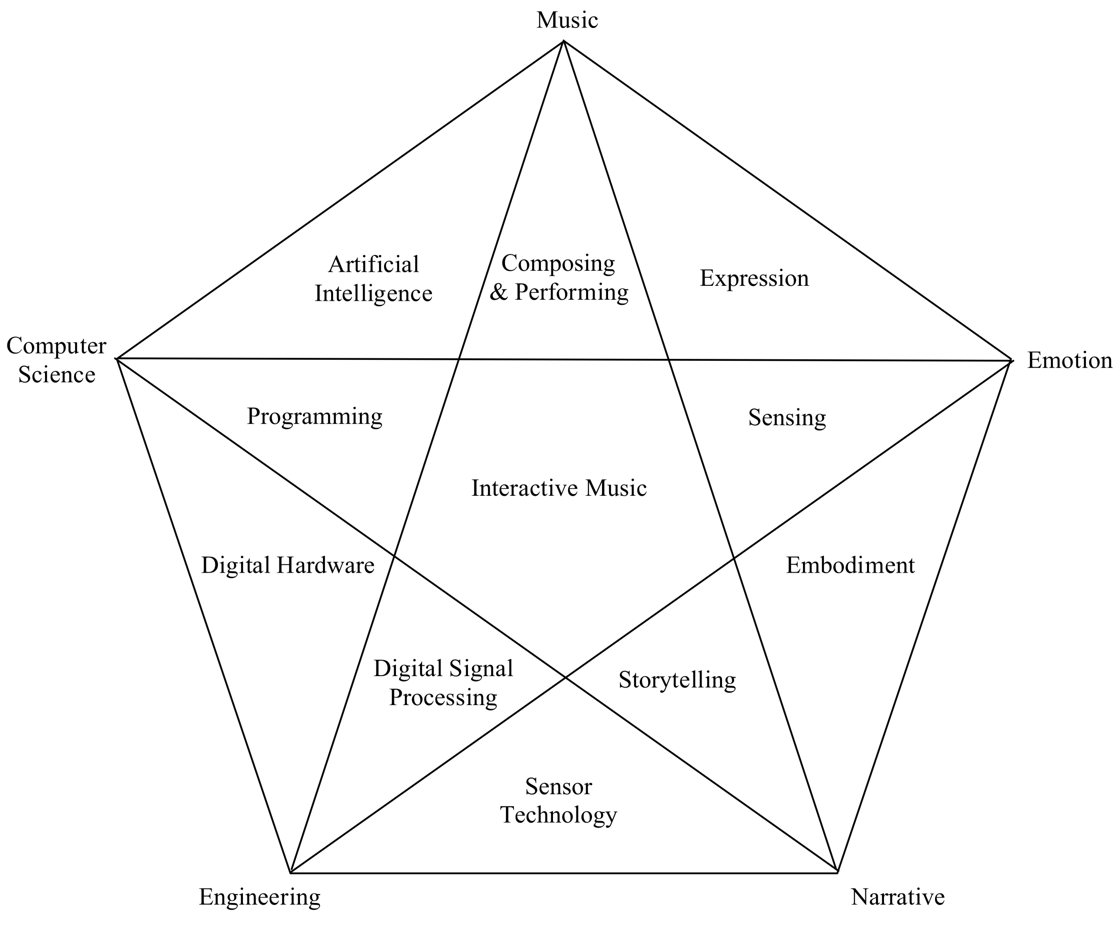

Fig. 1-2 *Disciplinary context of interactive music.*

1.2 MUSIC, EXPRESSION, AND EMOTION

Traditional listening models based on notes grouped into melodies, rhythms, chords, and harmonic progressions are only applicable to a small group of expert listeners such as musicians, and there is strong evidence that non-musicians do not hear music in these terms (Martin et al. 1998). But emotional skills are an essential part of human intelligence; they modulate human communication and are fundamental to human activities including the making and consumption of music (Scherer et al. 2001). The significance of the emotional aspects in musical content is readily appreciable by scholar and layman alike and it seems obvious to capacitate computers with the ability to recognise and respond to human emotion in the field of interactive music performance.

Concepts from affective computing (Picard 1995) are therefore applicable in interactive music, together with models of human information processing including, but not limited to emotion (Hashimoto 1997). It is, however, important to clarify what we mean when we suggest that a computer may recognise and react to emotion. A computer does not understand the background for human emotions, it only recognises emotional states to the extent they are manifested on signal level. Likewise, humans would not understand computer generated synthetic emotion as we understand our emotions (that is, what triggers, enhances, and calms them), but only as we perceive them being manifested by the computer in a sensible manner. We try to understand, and by doing so we place the incoming emotions in our context. The process is inward and similar to what happens with human-human emotions where a received statement is interpreted before it is reacted to. A system's ability to analyse and synthesise emotion in real-time can be a powerful tool for the composer-programmer to use in the decision-making processes during performance, in performance assistance tasks for presence and expression enhancement, and in the computer's independent, graduated response to the performer. The computer may base its response on some combination of emotion readings, internal settings, and low and mid level performance analysis.

The strength of emotion and the ease with which humans detect basic emotions in music have contributed to establishing sets of detectable expression categories through a limited number of high-level music parameters (Sundberg 1993, Gabrielsson et al. 1996, Camurri 2000, Bresin et al. 2000, Juslin 2000, Canazza et al. 2001, Zanon et al. 2003, Canazza et al. 2004). In (Camurri et al. 2000), the authors report on an abstract space representing emotional content of music. The emotions were derived through analysis of perceptual content of a musical statement played with different expressive intentions and the detection of emotions were sufficiently solid to be independent of the instrument on which the music was performed. The emotion space was then used for synthesising expression in a mechanical performance of the same musical statement, where emotion descriptors were mapped to high-level music expression parameters for microvariations in timing, dynamics, and pitch. The emotions were successfully recognised by test subjects. However, not all emotion recognition techniques are applicable to real-time signal-based performance/composition, since many of them rely on high-level music representation input for their processing, rather than addressing the emotion content of the audio signal (Leman 2002).

The Kansei concept of data processing has been important in providing a conceptual framework for much work on analysing emotion content of human efforts (Hashimoto 1997). Kansei, while including emotion, is slightly different from the concept of emotion, as it extends the definition of 'human expression data processing' to include also the weaker categories of feelings, intuition, and empathy. Other expressive channels such as human gesture and body movement are also contained within the Kansei concept. Hashimoto considers the force inherent to a human gesture the most important and complete conveyer of human emotion for such systems, and the interrelation between emotion, music, human movement and gesture has been researched with the aim of creating multimedia systems aimed at art, entertainment and rehabilitation (Camurri et al. 1999, 2005b, Suzuki et al. 2004).

In conclusion, digitally informed music has reached a point where it has become possible for a system to analyse a performance for emotion content and apply the information in real-time to aspects of its response. This development places the computer as a musical instrument in its own right and with its own characteristics, because no existing musical instrument displays this property.

1.3 TRACKING EXPRESSION AND INTENT

Tracking is the maintaining of a constant data ratio between an input data stream and a secondary data stream for alignment purposes. Parameterisation is the process of defining a system or a part of a system – or setting the conditions for its operation – as a set of numerical values, while mapping relates each element of a given set of parameters with one or more elements of a second set. The algorithms for data extraction of the secondary data stream are crucial in getting usable and truthful data for further processing in interactive music.

Parameterisation and mapping hold important aspects of composing interactive music, because they imply an interpretative layer controlling the system's modalities and their responsiveness towards the performer. Data storage and retrieval make it easily possible to affect musical interaction on small, medium and larger time-scales. The musical aspect the affection is able to modulate depends on the methods chosen but in general terms a system conceivably can access all concept levels of the mentioned time-scales. Using Lesaffre's taxonomy for musical feature extraction and analysis (Lesaffre et al. 2003) we can relate the time-scales to musical qualitative aspects and we see that all aspects of musical meaningful information can be affected by a digital music system on the physical, sensorial, perceptual, formal, and expressive concept levels (Fig. 1-3). The challenge lies in developing the methods and algorithms for extracting relevant data and up to the present comparatively less work has been carried out on the higher concept levels of form and expression.

The importance of parameterisation and mapping for the attributes of the interactive performance space of the human-computer complex is central to this dissertation, because it is the composer's primary work-field for making dynamic, interpretative affect-connections between analysis and synthesis. Many, including (Lippe 1996, Choi 1997, Tanaka 2001, Friberg 2004) provide insight into this area, where choices of parameterisation and mapping are capable of trivialising a potentially compelling artistic argument, since dynamic parameterisation and mapping modulate the intellectual, musical, and emotional impact of a performance in no insignificant way. The task is to devise mappings of the input parameters to sound synthesis in such a way that the relationship between action and resulting sound is intuitively comprehensible for the user, yet not trivial (Mulder et al. 1997).

<table>
<tr><th colspan="2">STRUCT</th><th colspan="2">CONCEPT LEVEL</th><th colspan="5">MUSICAL CONTENT FEATURES</th></tr>
<tr><td rowspan="5">CONTEXTUAL</td><td rowspan="3">global beyond 3 sec</td><td>HIGH II</td><td>EXPRESSIVE</td><td colspan="5">cognition | emotion | affect = syntactic+semantic concepts</td></tr>
<tr><td rowspan="2">HIGH I</td><td rowspan="2">FORMAL</td><td>melody</td><td>harmony</td><td>rhythm</td><td>source</td><td>dynamics</td></tr>
<tr><td>key
profile</td><td>tonality
cadence</td><td>rhythmic patterns
tempo</td><td>instrument
voice</td><td>trajectory
articulation</td></tr>
<tr><td rowspan="2">global < 3 sec</td><td rowspan="2">MID</td><td rowspan="2">PERCEPTUAL</td><td>successive intervallic pattern</td><td>simultane intervallic pattern</td><td>beat
IOI</td><td>spectral envelope</td><td>dynamic range
sound level</td></tr>
<tr><td colspan="2">pitch</td><td>time</td><td>timbre</td><td>loudness</td></tr>
<tr><td rowspan="3">NON-CONTEXTUAL</td><td>local + spatial</td><td>LOW II</td><td>SENSORIAL</td><td colspan="2" rowspan="2">periodicity pitch
pitch deviations
fundamental frequency</td><td rowspan="2">note duration
onset
offset</td><td rowspan="2">roughness
spectral flux
spectral centroid</td><td rowspan="2">neural energy
peak</td></tr>
<tr><td rowspan="2">local + temporal</td><td rowspan="2">LOW I</td><td rowspan="2">PHYSICAL</td></tr>
<tr><td colspan="2">frequency</td><td>duration</td><td>spectrum</td><td>intensity</td></tr>
</table>

Fig. 1-3*; Lessafre's taxonomy for musical feature extraction and analysis.*

Tracking applicable to interactive music is not restricted to the audio signal, since other perception data contains information that exposes the emotion content of the music performed or the emotion intent of the performer. Examples of such data are a performer's full-body gestures detected with a camera for content processing, and physiological data containing emotion state information detected by sensors.

1.3.1 SOUND TRACKING

The primary source for tracking musical expression and intent is the audio signal. Mechanical actions of the instrument producing the sound are available, but the drawback of mechanical tracking is that no information of the sound produced by the action is retained. A microphone is the sensor device generally used in sound tracking in interactive music because it gives access to the sound as it is being detected by human hearing. Processing of the audio signal typically aim at extracting perceptually relevant data such as pitch, loudness, and note onset, as well as more generalised features derived from the sound's spectral envelope such as spectral centroid, noisiness, and other statistically inferred data.

Sound tracking is relevant to the dissertation's papers (Graugaard 2004b, Graugaard 2005a, and Graugaard 2005b).

1.3.2 HUMAN GESTURE TRACKING

Music performance has strong ties with complimentary physical gestures for conveying musical expression. A conductor makes use of expressive gestures and in chamber music important information is passed among the performers through gestures for co-ordination of rhythm, dynamics, and phrasing. It is a natural step to analyse a performer's gestures for expressive cues and emotion content (Zannos et al. 1997, Camurri et al. 1999 and 2002, Friberg et al. 2002, Friberg 2004). Correlations between the perceived emotion content of a musician's movements in performance and the analysed emotion content of the music performed was studied in (Dahl et al. 2004), where it was found that basic emotions such as happiness, sadness, and anger were easily detected by the test subjects. Such is the strength of our conditioned link between gesture and sound that Denis Smalley is able to convincingly base his concept of gestural surrogates on the perceived link between a musical gesture and an underlying human or instrumental gesture (Smalley 1997). A third-order surrogacy is, according to Smalley, a gesture inferred or imagined in the sound where there is uncertainty of the source or reality of the gesture. It is tempting to modify the argumentation slightly and point to an observed local or full-body gesture with its coupled sonic response where the spectator may feel a strong bond between the gesture's emotional intent, even though there is no obvious reason as to why the details of the sound are exactly as they are.

The origins of music cannot be easily separated from those of dance – stomping, clapping hands, snapping fingers and enjoying the movement of one's body is part of the earliest form of religious feelings – as archetypal forms were laid down by early humanoids through stylised movements accompanied by 'body music'. This complementary nature of dance and music is as vivid as ever, and real-time systems have been made to extract expression data from full-body human movements in music and dance, and relate the data to aspects of the accompanying music in multimedia or other types of interactive systems (Camurri 2002). In (Friberg 2004) a performance system is described where sound and human full-body gestures combine to influence through fuzzy logic the musical and visuals output of an interactive system. Users affect the system by vocal utterances and performing a set of gestures at varying intensities.

Human gesture tracking opens up to the field of installations, and human gesture tracking of non-expert users – such as museum guests – is very often a pivotal point in music-based installations (Paine 2002), where depth of data is sought through a variety of adaptable analysis-synthesis methodologies inserted between the human presence and nature of behaviour on one side, and the physical space on the other. The sonic and visual response of the technology must attach to the communicative channels of feelings, intuition, and empathy in order to be intuitively experienced by the inhabitant of the installation. The relation between the inhabitant and the work is highly non-linear and its success is judged by the tension space that is generated through the directness/obliqueness of the mapping onto the sound/vision space and its resolution over time. This relation's resolution is ultimately an artistic matter, since a music-based installation creates a dynamic field of expectations that are fulfilled and negated over time and thereby articulate the time-space aggregate of the artistic work.

Human gesture tracking is relevant to the dissertation's papers (Graugaard 2005b).

1.3.3 TRACKING PHYSIOLOGICAL SIGNALS

Most interfaces aimed at expression detection analyse an audio or video stream for perceptual features. But physiological data extracted from involuntary body activity such as muscle tension, skin conductivity, blood pressure, and heart rate can also be used for this purpose, as they contain information regarding emotional states and expressive behaviours. Music is mediated through perception modalities and emotions but only recently have attempts been made to understand the emotional state behind physiological signals in the context of music. On the other hand, emotions and their physiological manifestation do not have a straightforward link, and humans do not display similar emotional response to the same stimuli nor report a given emotion similarly in different contexts. A detailed automatic emotion classification related to music from physiological features is therefore not an easy task, but distinct musical events have been found to cause significant variation in emotion rating inside a narrow time frame as well as some constant inter-individual physiological response-clusters related to music's harmonic and dynamic change (Grewe et al. 2005).

Lang's well-known 2D arousal/valence model (Lang 1995) covers a wide range of emotions, but high-resolution automatic emotion sensing depends on careful selection of physiological signals and reliability evaluation models (Wang et al. 2004). Some emotion categories correlate to certain physiological signals, where for instance galvanic skin response (skin conductivity level) and heart rate correlate to emotion intensity. Optimisation of the physiological data and affective state correlate can be enhanced by different protocols and algorithms, and by using several physiological signals in parallel (Healey et al. 1998, Rani et al. 2003, Wang et al. 2004). Reducing the number of classes on the arousal/valence continuum gives more explicit results, and a gratification valence can be used in substitution for some of the basic emotions or as complementary data for obtaining more depth in the information stream (Herbelin et al. 2004).

Physiological signals as emotion identifiers relate to the idea of non-linguistic description that unites perception with its sensimotoric basis (Leman 2006b). This is a significant connection which facilitates the design of interactive music systems for non-expert users who do not have a vocabulary to describe what they do, but may utilise their emotional and otherwise musical experience when interacting with the system. The demarcation of the artistic work also facilitates optimising the interpretation of intentions inherent to the user's action in such cases, because the result of the analysis is placed in and manifested through the predefined action-response space of the artistic work. The challenges can hereby be turned to benefit in the process of artistic creation, where the composer-creator makes sense of the normalised input in a predefined, dynamic musical space.

Physiological signal tracking is relevant to the dissertation's paper (Graugaard 2005b).

1.4 THE SOUND, GESTURE, AND PHYSIOLOGY AMALGAM

In more general terms, expression data acquired through the channels of sound, vision and human physiology can be related to strategies for storytelling as expressive, interactive narratives, because the model is of the time domain and no longer a exclusively musical one. This opens up to contexts other than purely musical, and when engaging a non-expert user, the result is no longer primarily a performance and the result is the establishing of an inhabited world, connected to its 'population' through a variety of sensor streams picked up through microphones, cameras, and other data-capturing devices (Fig. 1-3). A music performance can also be thought of in similar terms, that is, as the exploration of a pre-existing interactive narrative-based multi-sensorial space where the performance becomes the sounding manifestation of the performer's exploration of that world through reading a musical score, a (directed) improvisation, an oral explanation, or any combination of these. This approach characterises much existing work in digital art and entertainment (Camurri et al. 2005, Leman 2006a), and such interactive works based on music are characterised by a great many action-reaction patterns where some may be quite different from those a musician is trained at dealing with.

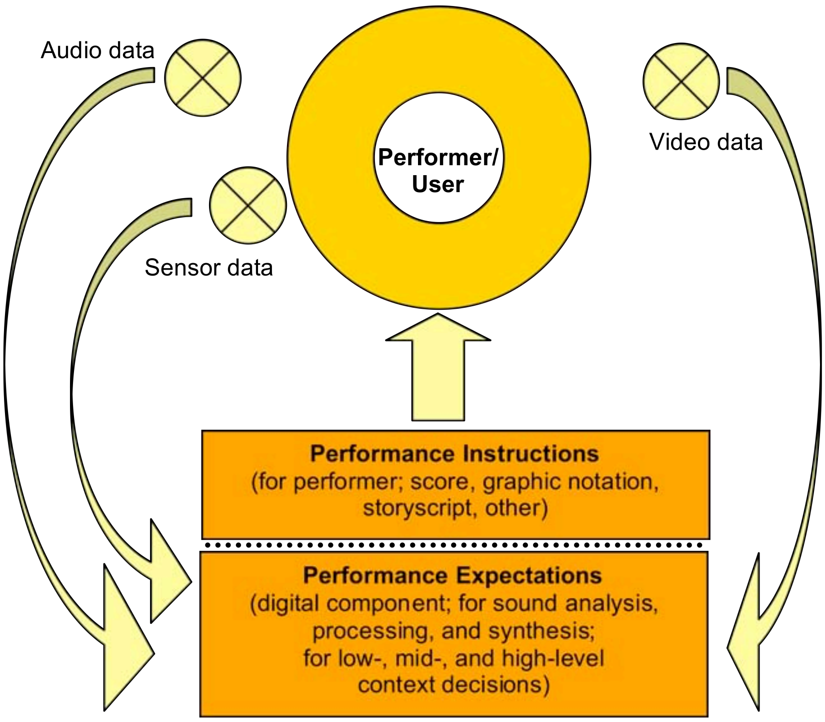

Fig. 1-4 *Interactive music's multisensorial context*

When the expressive musical actions manifest themselves through other means than those well-established in the traditional musical art-forms, then the generalised thinking – the music theory – falls short of covering all relevant aspects of the musical work. There have been significant advances in the fields of video and audio processing, pattern recognition, computer vision, and affective computing, but only little theoretical work has been done so far to reconcile traditional music theory with interactive music and affective, cross-modal art-forms. Attempts to apply ideas from music theory and statistical signal processing have not yet led to successful musical multimedia systems. Advances therefore depend on other disciplines and multimodal context-sensitive HMI is likely to become the single most widespread research topic of the artificial intelligence research community, because advances in this area will change how mass-market consumers interact with technology. Artists and performers will use computing in new ways, and an example of this is the work done to unify approaches to identification of emotion factors in movement and in music. For an overview see (Leman 2006a). A central work described in (Camurri et al. 1999) identifies an 'effort space' using Rudolf Laban's definition of four effort factors of time, space, weight, and flow, for identifying movements by means of time-space-weight vectors.

Even though the realisation of a robust, multimodal, adaptive, context-sensitive analyser of human non-verbal affective state still is far away, human emotion recognition in music has had some initial success. Such recognition is most likely to be accurate when it combines multiple modalities with pre-defined information about the performer's context and goal together with low-level feature extraction and high-level reasoning. Emotion recognition detects basic emotional states as they are manifested on signal level, but it is tempting to apply synthetic emotion to sound synthesis. Synthetic and artificial emotions hinge on the ability of the computer to compute in ways humans cannot, and their validity as 'emotions' lies in our reaction: if we can recognise it as a valid manifestation of some (unnamed) emotional state, then it is. The computer is a tool, and the fact that we may give it aspects of emotions does not mean that it ceases being a tool. Artists have explored 'glitch' as a way to humanise technology (Cascone 2000), but it could be questioned if technology failure is an analogy to human emotion. Instead it may be regarded as an exploration for new aesthetics, which in that case makes it a human endeavour in a very old tradition and not a manifestation of independent machine emotion.

2. THE INTERACTION INCENTIVE

The tendency in interactive music has so far been to fine-tune the actions of a digital accompaniment for the needs of a particular musical composition. A piece may rely more on the composer's musical and programming skills in creating expressive music, than on a digital system capable of analysing high-level expressive data for high-level response, with appropriate supportive low-level actions. This is a perfectly viable approach for making instrument-computer compositions in artistic terms and is a logical extension of the tradition of fixed form or sequencer based accompaniment. The system is at the most only partly interactive and operates somewhere between actual interaction and simulation of such interaction. This may not affect the end result negatively – if the system mimics interactive capabilities well, then the music is perceived to have interactive qualities, and thus satisfies the need for flexible, musical expression. But it also means that the system does not possess any interactive attributes to explore during composition or bring into play during performance.

2.1 THE INTERACTING MUSIC SYSTEM

A computer's sound production and sound class are different from acoustic instruments, and this places a digital music system in an instrument category of its own, alongside the acoustic instrument families. But the computer combines this with a number of unique and shapeable features, executable in real-time. Among these are:

- Data processing and storage
- Displacement in time
- Procedural data processing
- Dynamic adaptation

This makes it possible for the system to sense and actuate in accordance with human sound perception. Data storage makes it possible to displace information over time, and its processing makes available innumerable methods for extracting, generating, applying, combining, and modelling data. The speed with which this can be done makes multi-level processing straight-forward, hereby creating a rich context for data handling among levels and modalities. The relation between a performer and an interacting computer is therefore vastly different from the relationship of a soloist and a human accompanist because it includes so many possibilities of musically relevant data processing that are uniquely different from those of a human accompanist. It is not suggested that a computer is superior (which it very often is not), but we merely state that its range to some degree overlaps and to a large degree lies outside that of a human accompanist.

2.2 AN INTERACTIVE NARRATIVE OR AN INHABITED WORLD?

If we consider that an interactive computer in its memory carries the sonic counterpart of the performance score with the necessary input nodes for its reaction to the performance and its own action set, we can then interpret this as a synthetic world ready to be scrutinised by the performer. The performer explores the computer's feature set when performing, while the computer adapts to the input and evolves its own actions according to the progression of the notated score. The initiative fundamentally lies with the performer, even though the computer may have a significant independent part. This can be likened to a symbiotic relation among two entities where an environment – the interactive computer – is being explored by an independent organism – the performer – in real-time.

Such concepts from interactive, narrative storytelling in interactive multimedia are tempting to apply to the new and wider scope of interactive music, because they deal with organised time structures carrying some form of meaning explored in real-time by a user. But there are fundamental differences between music and narrative and storytelling – first of all because music is not directly comparable to language, since its units do not carry absolute meaning – making a cautious approach necessary. Storytelling has several definitions but resembles musical structuring if defined as the art of developing coherent event structures in time in order to provoke determined effects in the audience (Zagalo et al. 2004). Use of the term 'narrative' also varies considerably, from constricted interpretations involving strong time-space continuity to very broad understandings of the formal, rhetorical, or thematic coherence of some construction in the time domain. Not many reports relating storytelling and narrative to interactive music are available to us. Interpretations of the narrative or expressive/emotional elements of music are found in (Davidson et al. 2002), while (Berry et al. 2003) describe what they call an interactive musical narrative when a user navigates, listens, and responds to the creatures and plants inhabiting their artificial intelligence world *Gakki-mon*. A systematic approach to handling the non-verbal expressiveness of musical gestures in interactive multimedia involving reaching back to their narrated form is found in (Leman et al. 2004). High-level narrative descriptions of expressiveness interact with a gesture-based representation of expression for recognition and action, which in turn interacts at signal level. The authors use the connections to establish links between the multimodal system and our ability to talk about expressiveness, but they do not give mention to strategies for their use in structures resembling interactive narrative storytelling. Mention to time based narrative structures is given in (Leman et al. 2006), but only in passing, and then connected to the modifications such structures may cause on mapping and parameterisation of representations at the levels of gesture and signal.

2.3 INTERACTION, EXPRESSION, AND EMOTION

Neurological studies (Damasio 1994) have shown that humans' efficiency in making decisions reduce significantly without emotion, and Picard's central argument for affective computing – that a system cannot be truly intelligent without emotion – consequently becomes a central issue for interactive music (Picard et al. 2001). All living intelligent systems have emotion in some form. Humans have the most sophisticated emotion system of all, and the capability of artistic expression, such as music, for expressing emotions beyond verbal articulation is the ultimate testimony to this. Emotionally grounded musical aggregates are rewarding listening experiences through their emotional content, and not by structural qualities on the symbolic plane alone, because what satisfies our musicality are the emotions expressed (and their unfolding over time), rather than 'factual' aspects void of emotions.

Interactive music performance systems will therefore benefit from aspects of emotionlike mechanisms. Such mechanisms may be different from what humans possess, because it may not always make sense to attempt a precise equality between human and machine emotion attributes when we do not know the totality of what they contain, and therefore do not know if we need more to justify our categorisation. We can better try to reduce the highly composite activity of music making to a number of functional requirements in a particular work. We can even imagine them to be different from those of humans, and we should not be troubled by the thought of their exhibiting appreciable, musical qualities. Categorisation of emotions is justified because music performance establishes a situation of apparent and oblique connections between logical precision and intuitive expression. The fundamental feature requirement is clarity of interaction, and the success criteria are the ability to transgress the interaction mechanism and establish channels of non-verbal, affective communication.

2.4 FURTHER CONSIDERATIONS

The properties that justify digital systems as containing autonomous musical and artistic qualities are many, as previously stated. Some of these areas have been developed extensively, but many new insights will come about, altering artistic and musical vision. We may become able to build a synthetic character that can emulate a living human performer, and we may make a computer that is intelligent – depending on the definition of such intelligence – and we may attempt to model human emotions. But then: what should it do? I believe we can aim at giving the computer the ability to adapt to emotion feedback in a way that enhances, furthers and – possibly even – redefines aspects of musical composition, and other art-forms with a significant sound element. For example, a musical goal may not aim at combining several interactions into a single, conclusive result. A musical goal can reside in an articulate richness of different modes of interaction that contribute to independent yet simultaneous musical demands. The concurrent access to these articulation layers enhances a performer's expressive potential where the musical gesture may expand through complementarity, rather than fusing towards one task or into a single statement. Such reasoning has consequences for musical form, detail, and expression and establishes the particular qualities and feature set of interactive music, and digitally informed art.

But what has been solved when we equip an interactive composition with some form and degree of emotional intelligence? It seems that all essential compositional problems remain unresolved by this fact, because the computer does not have any *a priori* schemata for suitable reaction to affective data. These reactions rely on the nature of the system, and uniquely so, because an emotion-set is unique to a composition, just as musical aspects of the composition are unique to the composition. We use the same material of scales, chords, and rhythms but their specific use is unique to a piece, just as we have a given range of emotions whose particular constellation and development in the course of a composition is unique and uniquely defines the composition.

All possible components of an affective system will not need to be present in such a system. A set of main functional categories is:

- Regulating formal processes,
- Guiding and influencing presence, attention, and selection,
- Displaying meaningfulness,
- Supporting with intelligent decision-making, and
- Dealing decisively and flexibly with complex, unpredictable inputs to a resource-limited system.

Maybe it is some day possible to create a truly intelligent synthetic music performer, but it is not very likely to happen in the immediately foreseeable future. Emotions may not even be the aspect that we want to emphasise when distinguishing between performer and computer. The proper fusion of the two into a relational whole may reside on another level where neither the performer nor the computer feels and acts independently and where new expressivity of a nature beyond mutually self-sufficient endeavour is revealed.

3. BALANCING THE REQUIREMENTS

Ramifications for composing and performing music when an interactive computer is utilised are manifold and multi-layered and relate significantly to issues of functionality, form, and expressivity. Problems and solutions encountered are of a combined technical, practical and artistic nature, and due balance between these aspects must be observed. This balance must achieve agreement among diverging features so they mutually reinforce each other, and this will in turn determine central characteristics of the interactive composition.

3.1 (Graugaard 2004): FORMAL CONSEQUENCES AND POSSIBILITIES

The unique relation between performer and computer generated accompaniment in the composition *GUITAR* (Graugaard 2001) for guitar and interactive computer is described. The relation establishes a perception-based multi-parameter environment, which the performer explores. Open-form notation is used to reinforce the interactive aspect of the environment, and large-scale form connections is established to reinforce cohesion, by using music material for synthesis purposes extracted from the performance in real-time. The roles of computer and performer is redefined to those of *world* and *inhabitant*, and the audience become witness to the inhabitant's exploration of the world, and how they affect each other during performance.

GUITAR was premiered March 3rd 2001 at the Carl-Nielsen Academy of Music, Denmark.

3.1.1 THE INTERACTION TIME-SPACE IN *GUITAR*

Interaction in music takes place in the time continuum from local triggering to global implicit affect (Lippe 1996) in a mapping continuum from one-to-one to one-to-many/many-to-one (Wanderley 2000) and in a spatial continuum from signal level to gesture level. Lesaffre's taxonomy can place the interaction in *GUITAR* in relation to musical content features. The overall formal layout is inherent to the composition but a central artistic argument is that detailed instantiation is determined by performance aspects captured in real-time, their effect uncontrollable by the performer. Interaction takes place as immediate low- and mid-level performance enhancing action-reaction sets while high-level interaction shapes the evolution of an open form section and cross-references among distinct sections up to several minutes apart. This falls within Lesaffre's taxonomy except that *GUITAR* also affects the macro-level of form properties. The result is a mesh of influences where interaction not always takes place in an obvious way albeit of much consequence to the outcome of a single performance.

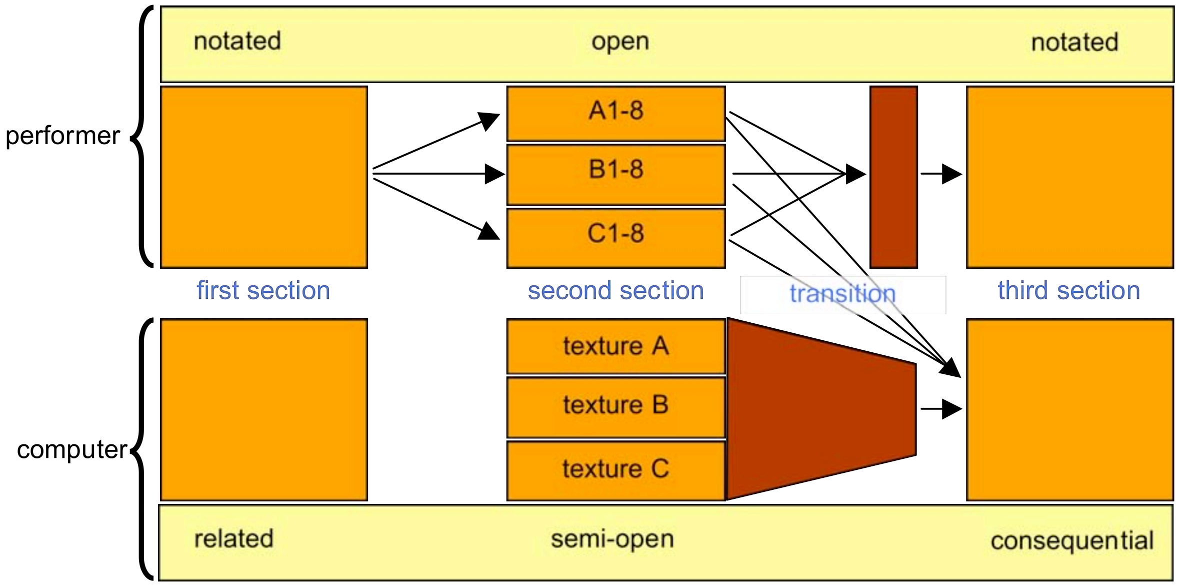

Fig. 3-1*; GUITAR, overall layout; from (Graugaard 2004b).*

GUITAR's formal layout is in three sections (Fig. 3-1) where references in the third section are made to the first and second sections. The second section is in open form notation and choices by the performer in this section influence the computer accompaniment in the third section. The aim was to let characteristics of the open form section influence the third section in such a way that performance-unique coherence could be achieved. The particular relation remains hidden to the audience as they would need to listen to another performance to notice the difference. This fact is not of importance to the piece because the point is not to distinguish between performances but to utilise the particularities of a given performance to create coherence between the second and third sections and consequence inside the third section with respect to properties of the second section.

3.2 (Graugaard 2005a): UTILITARIAN AND EXPRESSIVE ACCOMPANIMENT

These papers describe a singular system for accompanying singers, used in the opera *La Quintrala* (Graugaard 2003-04). The system had to unite three independent and partly conflicting purposes without limiting any of them:

- provide pitches for entry notes and tonal orientation,
- inject the singers' expressivity into the digital accompaniment, and
- securing flexible means for compositional coherence.

A procedure was devised whereby dynamic comparison between input and a predefined chords could be done, and data output could be triggered by different means. The system had to work in real-time and had to adapt for a variable number of singers.

La Quintrala was premiered September 2nd 2004 at Den Anden Opera, Copenhagen, Denmark.

3.2.1 NOTATED AND ELECTRONIC SCORE IN *LA QUINTRALA*

The notated score of *La Quintrala* contains the parts for the singers as well as the accompanying chord as pitch-sets (Fig. 3-2). The electronic score contains all sound analysis and synthesis methods and all procedural methods required by the composition as well as the chord sequence from the notated score. The singers know the accompaniment will contain pitches of a given chord – and possibly others – but the timbre of the pitches would not be indicated. The electronic score is mechanically advanced by a technician tapping a space bar during performance and each event adjusts the application's internal set-up and step through the chords as a sequence of pitch-sets. Mechanical score advancing was decided upon because automatic score-following systems were not found to be sufficiently reliable to be used unsupervised.

Fig. 3-2*; score example from La Quintrala, third staff is supporting pitch-sets; from (Graugaard 2005a).*

An important feature of the system is that input comparison and its mapping to the harmonic descriptors are implemented independently of sound synthesis. The advantage is that a uniform and consistent way of communicating basic harmonic content to the singers can be maintained throughout the opera. It can in this respect be likened to a piano reduction of the orchestra score of a traditional opera, albeit with much less information since all information of rhythm and dynamics in the accompaniment was a property of the sound synthesis or varying procedural methods and therefore kept out. The notation of the vocal parts is done in traditional fashion for best reading for the singers – unless when more unconventional open-form notation was used for artistic purposes – together with information of character, tempo, and dynamics.

3.2.2 AUDIO ANALYSIS AND PITCH-SET COMPARISON

Analysis of the singing extracts the four most prominent frequency components from the audio signal (Jehan et al. 2001). These components are compared to a stored pitch-set corresponding to the chord defined in the composition for supporting the singing at that particular moment. The analysis result is passed on to the comparisor (Fig. 3-3) to get the closest match which then would be passed on to the sound synthesis stage. Closest match can be dynamically defined according to the compositional context and is decided by a dissonance and a harmonicity factor. The dissonance factor is derived from a standard spectrum transformation approach where the analysis result is placed on a continuum between no correlation and complete pitch-set alignment. The alignment is determined by a dissonance factor in the range (0.-1) while a harmonicity factor in the range (0.-1.) determines the best match according to the ratio between the input and the possible candidates of the pitch-set.

3.2.3 OUTPUT TRIGGERING

Means are provided to control the comparer's output density as well as range and assigned duration. The controls were introduced in order to make the system able to incorporate and respond to expressive aspects of the singing. The mapping stage is better suited for this purpose since it provides a global and unified means of controlling affection possibilities. Appropriate parameter values are ultimately determined by artistic considerations because the specific nature of the affect depends on compositional needs relating to dramaturgical and scenic developments at any particular moment. The consequence for the singers is that their musical intensities and the intentions inherent in their singing is projected onto the accompanying sound canvas, hereby supporting their musicality and enhancing their stage presence.

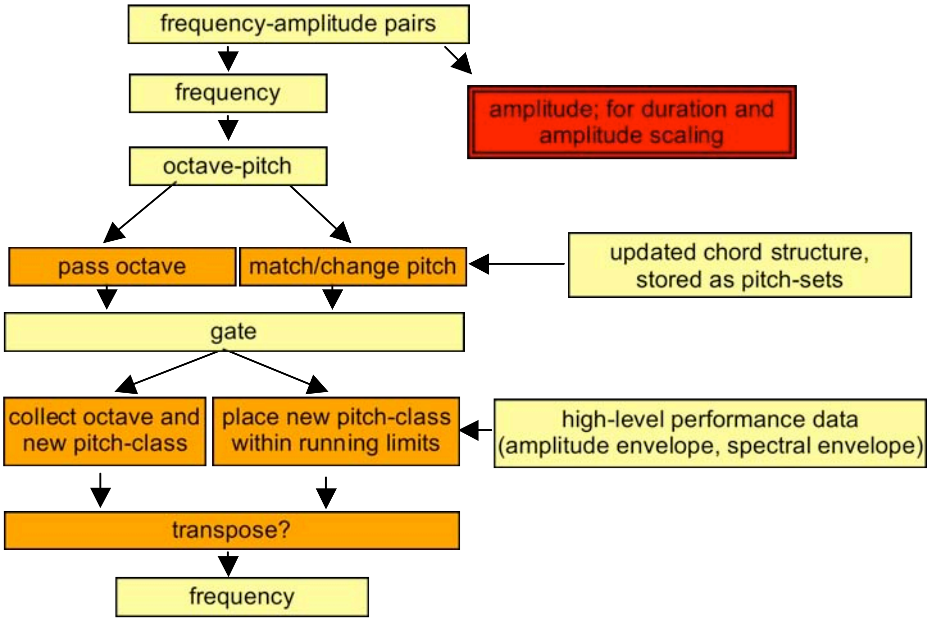

Fig. 3-3*; input and pitch-set comparison; from (Graugaard 2005a).*

An important matter is the triggering of individual outputs from the comparer. Triggering can take place from four different places, individually or in combination, and each with appropriate parameters (Fig. 3-4). The system hereby responds to compositional choices stored at system level, to expressive parameters from the singers, and to real-time interpretative layers between singers and sound synthesis activated *ad hoc*.

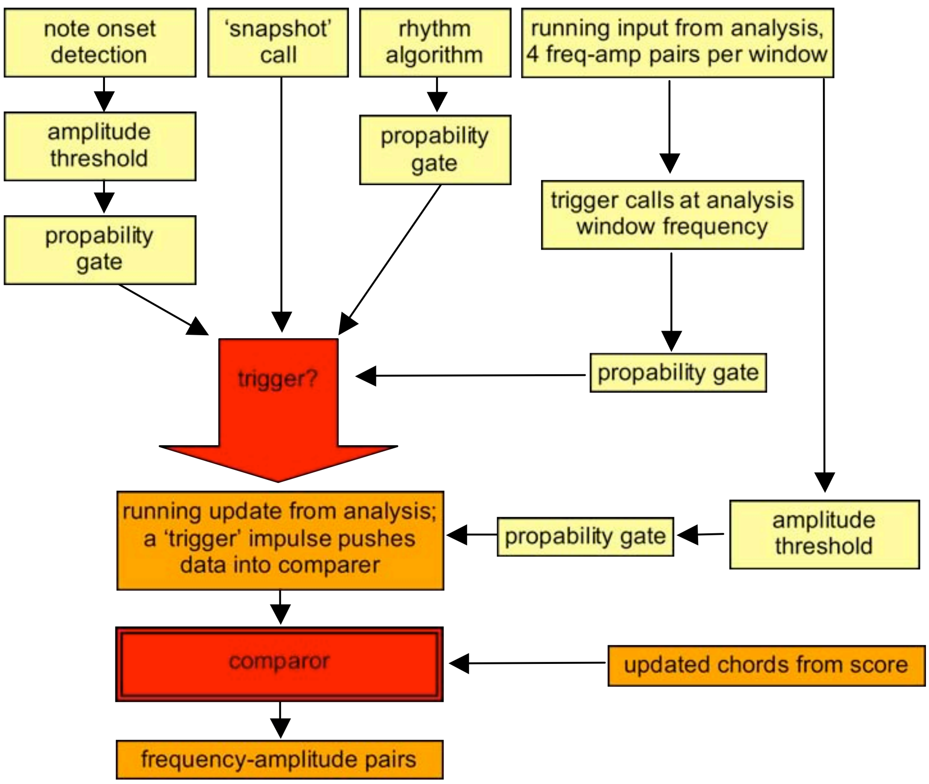

Fig. 3-4*; input-output flow for comparison of spectral peaks; from (Graugaard 2005a).*

3.2.4 EXPRESSIVITY AND PRESENCE ENHANCEMENT

The interactivity continuum in *La Quintrala* is between direct triggering and implied relations and very often with different processes located at different points on this continuum. The singers experience in an intuitive fashion that their musical efforts are supported by the sound synthesis and that their expressive intentions are augmented, juxtaposed, or commented on by the accompanying sound canvas. They do not necessarily feel that they control the accompaniment but always feel that it enhances their presence, both musically and dramatically.

3.2.5 COMPOSITIONAL CONSEQUENCES

The system was able to reconcile the at times opposing requirements of the singers and the composer. The system would enable the accompaniment to correlate with the singers without restricting them in their timing or in performing extra pitch content. It always adhered the synthesis output to both the singing and the chords specified in the score independent of the procedural and synthesis methods chosen. When open-form techniques were used it would be able to smoothly evolve from one compositional approach to the subsequent, at times very different, approach.

3.3 DISCUSSION

Functionality for high-level decisions of form can be built into an interaction system since not all musical functionality needs reside in the area of low and medium-level interaction. A system's combination of low, medium and high-level interaction addresses the musical context at the physical, sensorial, perceptual, formal, and expressive levels with the corresponding musical features in melody, harmony, rhythm, timbre, and dynamics. This means that interaction can take place on all musical levels, but it also implies that the effect of interaction will be felt in very different ways according to the characteristics and properties of these levels, in a given piece of music and the potential for the system seems extensive. A combined network addressing performer requirements, medium-level parameter modulation, and means for compositional correlation with top-level control greatly facilitates the control of musical interaction.

4. VISUAL GESTURE DETECTION IN MUSIC

The body plays a central role in musical activities. Physical gestures can be optimised for manipulating a musical instrument for music performance, and physical gesture controllers can translate gestures into sound. Many of such controllers have been developed for musical purposes (Wanderley 2000).

4.21 (Graugaard 2006a): USING HUMAN GESTURES FOR AFFECTING MUSIC

We detect human gestures for mapping onto musically expressive qualities in the compositions *Stgo* (Graugaard 2005) and *Gestures You Made* (Graugaard 2005-06). The mapping in *Stgo* injects real-time expressive qualities into laptop music through hand gestures, while the mapping in *Gestures You Made* uses our particular duality concept of sound *shaping* and sound *colouring* through analysing a performer's playing and full-body movement.

Stgo was premiered December 6th 2005 in Santiago, Chile, and *Gestures You Made* was premiered February 12th 2006 in Reykjavik, Iceland.

4.1.1 MOTIVATION

Music is a sonic art-form, which conveys structures and emotions. Emotions are expressed by humans in a variety of ways including physical gestures such as full-body movements and hand gestures. Music is furthermore a performed art where the presence of a human performer normally is required. The visual presence of the performer is an integrated part of music performance where the audience watches the performer's physical elicitations while performing. Interactive music is a hybrid system where a computer system interacts in real-time with a human performer, mostly an instrumentalist or a singer. Since music expresses emotions, it seems viable to use other data channels than sound for conveying emotions to an interacting musical data system. With the advent of inexpensive webcams for detecting vision data – and efficient algorithms for extracting statistical information of this data – computer vision has become one of such new channels for exploration.

The applications described in (Graugaard 2006a) concerns analysis of a musician's full-body gestures (Graugaard 2006c) and a laptop performer's use of hand gestures (Graugaard 2005c).

4.1.2 GESTURE ANALYSIS AND SOUND SYNTHESIS

Video analysis is done by computing the difference between consecutive video frames. The robustness of this well-known motion detection technique is very convenient in a performance situation where stage set-up cannot easily be optimised for more sophisticated techniques. Webcams have a drawback in that they typically output only 15 video frames per second, which prevents the use of techniques requiring speed and precision. The performer's movement is detected, and *XY*-position of the on-pixels from the motion detection having an off-pixel immediately above – the upper border of the movement mass – is identified as the upper border of the movement. Motion mass and the displacement speed of the motion mass' centroid is also calculated.

The sound synthesis method is a variation on the timbre model for musical sounds described in (Jensen 1999). It is essentially sound synthesis by frequency modulation containing 32 individual components spaced at $n*f(0)$ with the exposed high-level attributes modulation strength, randomisation band-width, and component correlation. The sound can be continuous (affected by the aforementioned parameters) or amplitude envelopes can be applied to make regular or irregular punctuations of the output.

4.1.3 CASES

The system has been explored in two separate compositions under very different circumstances, one using an instrumental performer's full-body movements, the other using a laptop performer's hand gestures. They share the idea that gesture should expressively affect sound and that the gesture-sound complex should be conceived as an audio-visual whole by performer and audience alike.

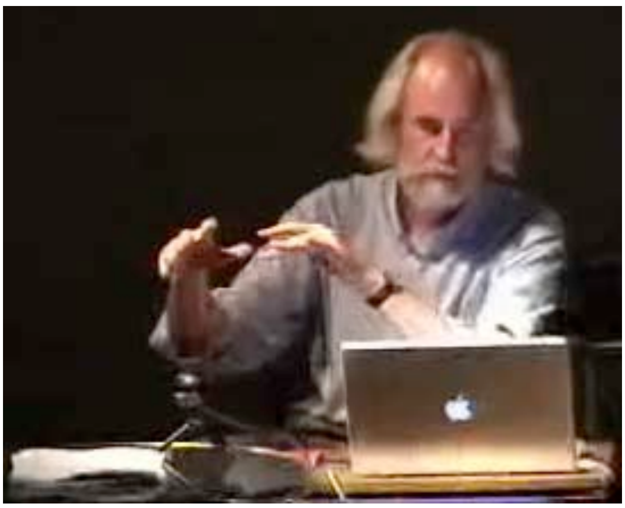

Fig. 4-4*; Stgo in performance; from (Graugaard 2006a).*

4.1.3.1 STGO

The role of the performer in laptop music is almost enigmatic (Cascone 2000). Informing laptop music with human gestures may take laptop performance out of this condition – or it may make it even more enigmatic and ritualistic – but the artistic reason for including video tracking in the composition *Stgo* for player and interactive/algorithmic computer system (Graugaard 2005) was that it would make it possible to instil human expressiveness into electronic music in real-time. Hand gestures are chosen to affect sound and music (Fig. 4-4). The upper border of the movement mass affects the relative amplitude of the 32 synthesis components, while displacement speed of the motion-mass centroid determines *f(0)* of the sound synthesis, and the volume of the movement-mass determines the length of the applied amplitude envelopes. This produces a conceptual mapping where the upper contour of the hand gestures determines timbre, faster hand movements produce higher fundamentals, and larger motion mass – due to increased proximity or predominantly openhand movements – produce longer grains.

4.1.3.2 GESTURES YOU MADE

The same motion detection technique is used in *Gestures You Made* for oboe and interactive computer (Graugaard 2005-06) where full-body movements are detected for upper-border limits applied to relative sound component amplitude. But sound is also incorporated into the synthesis through regular analysis of the instrument with the purpose of conceptually dividing sound synthesis into complementary parts of shaping and colouring the sound. The synthesis output is in essence continuous and the fundamental frequency *f(0)* of the instrument analysis is mapped to *f(0)* of the sound synthesis while the running amplitude is mapped to HLA in settings arrived at through experimentation during rehearsal. Gesture is hereby tied to relative component amplitude and playing is tied to synthesis spectrum: the sound is *shaped* by the full-body gestures and *coloured* by the playing.

4.2 DISCUSSION

Gestures are a natural way for humans to convey intentions and emotions. The challenge for a digital system utilising gestures consists in creating an intuitive environment that permits direct, natural and immediate transmission of intentions and emotions between performer and system. This engagement should eventually lead the user to perceive the outcome of the actions as a direct and natural response, where the interface layer takes on a subordinate role relative to the expression sought. The user becomes the judge of success through the easy engagement with the interface, and this success criteria have to some degree been met in the two cases described above. An immediate reaction by the system of a musical nature reasonably resembling the intention of the user was recognised, but it may be that the system reaction seemed satisfactory due to the abstractness of the sonic environment combined with the musical predetermination of the composition.

5. PHYSIOLIGICAL SIGNAL DETECTION AND SOUND SYNTHESIS

Physiological signals contain relevant information for indicating a person's emotional state and direction of wilful intent. We may interpret the information as the person's expressive reaction to stimuli in a real-time system. When the stimuli are part of the system, an action-reaction loop is established where the user informs the system and reacts to the system's interpretation of the input, hereby transforming stimuli and affecting action in real-time. We then facilitate sound synthesis modulation by inserting a normalised space between the user reaction and the system stimuli.

5.1 (Graugaard 2005b): SENSING EMOTION AND INTENTION THROUGH PHYSIOLOGICAL SIGNALS

The unique artistic reasoning behind the installation *SoundGlove II* is to combine a user's conscious and subconscious self into one coherent expression for the user to interact with, represented in abstract audio and video. We analyse the muscles controlling the user's hand and finger movements through a unique system of a primary data stream (PDS) and a secondary data stream (SDS). The analysis result is applied to sound and video synthesis in real-time through two mapping spaces. The installation evolves over time, influenced by summarised user actions through 'scenarios'.

SoundGlove II was at public display September 30th – December 18th 2005 at the Danish Museum of Contemporary Art.

5.1.1 DATA EXTRACTION AND ANALYSIS

The sensor glove in *SoundGlove II* is placed on a pedestal in the middle of a room and sound is diffused in eight channels in a cubic set-up centred around the glove while visuals are projected on one of the four walls demarcating the installation space. It makes use of a skin electromyograph (sEMG) to extract physiological data from the user (Fig. 5-1). The signal is bipolar in the frequency range 20-1.600 Herz and the primary data output is a measurement of the muscular tension represented by the signal's amplitude envelope. Further data is extracted through standard audio analysis methods as perceptual data such as fundamental frequency, spectral centroid, and noisiness.

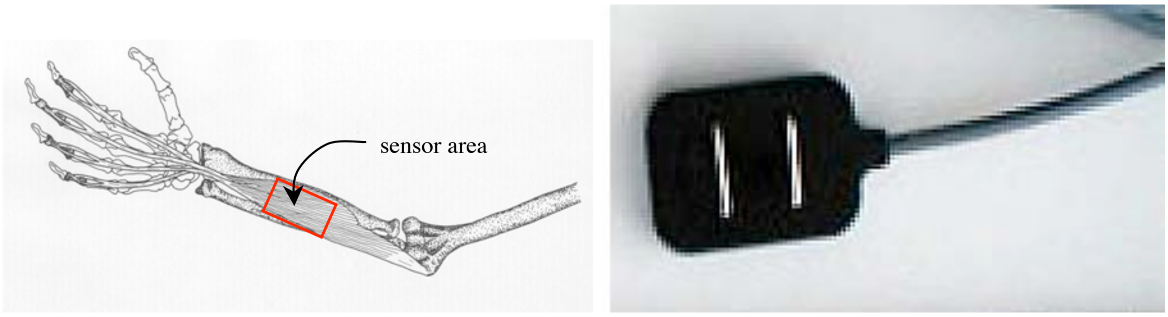

***Fig. 5-1**; relevant sensor area and the sEMG sensor; from (Graugaard 2005b).*

Analysis of the input is divided into a PDS and a SDS. This is done in order to extract more information from the single channel data stream from the sEMG with the aim to extrapolate intention behind the physical gesture. The PDS is the running amplitude envelope interpreted as the energy presented to the system. Three quality parameters *noise value*, *harmonicity value*, and *frequency value* constitute the SDS as a multidimensional space around the PDS. This improves data resolution and enhances the flexibility and responsiveness of the system even though it is not possible to extract true emotion data with only one physiological signal sensor. To do this the sEMG would have to be combined with another physiological data channel such as galvanic skin conductivity or heart rate, or another strategically placed EMG sensor. This was not possible to due to restrictions in the installation interface, but the method of a combined PDS and SDS was sufficient to deduce relative emotional states or intentions that would function as intent direction indicators.

5.1.2 SYSTEM

Analysis data is used to affect sound and video in real-time in the short and medium time span and to affect the evolution of the system over longer and very long time spans. A 'composition commonology space' and a 'visuals qualitative space' are defined whereby the analysis data is parsed to influence the appropriate parameters of sound and visuals. No direct binding sound-visuals are defined, instead the two spaces correlate through their individual relation to the shared physiological data which affects them in complementary ways. This complementarity correlates with the user action and our perception's pronounced intermodal nature establishes a perceptual correspondence between sound and visuals.

5.1.3 FORMAL AND INTUITIVE

Users are intended to get the feeling that they are able to influence the system by their passive presence and conscious actions but the exact nature of the influence was intended to remain somewhat oblique. This decision underscores the nature of the relation between the conscious and subconscious, which is distinguished by the fact that the unconscious is out of reach of the conscious. On the other hand, the relationship must not become so weak that users do not get sufficient feedback to establish an intuitive communication channel with the system preventing them from experiencing a relationship between their presence and actions on the one hand and the response of the sound and the visuals on the other. Some actions are therefore intended for low-level triggering while other data influences musical and graphic phrases in a duration of 3-8 seconds.

A profile of each user's actions is stored in computer memory as user related 'scenarios' and the system keeps track of the combined weighted scenarios which are employed to affect the system evolution. *SoundGlove II* had a predefined life span of three months and the scenarios would influence the evolution of this life span. The installation can in this context be considered as a 'world' and the users as a (sequence of) 'inhabitants', influencing – and being influenced by – its evolution.

5.1.4 SOUND AND VIDEO SYNTHESIS

The sound is algorithmically generated and consists primarily of timbre modulated time-dependent events, or pulsations, further described in section 5.2. The pulsations are defined in a set of layers as probabilistics, periods, and amplitudes, and they can take the form of rhythms depending on these parameters and a set of high-level parameters related to the aforementioned system life span. The layers are defined as frequency ranges with free movement over the entire frequency range. The waveforms contained in the events is a set of synthetic and ecological continuous noises but the system allows for any kind of audio signal fed into it, pre-recorded or live.

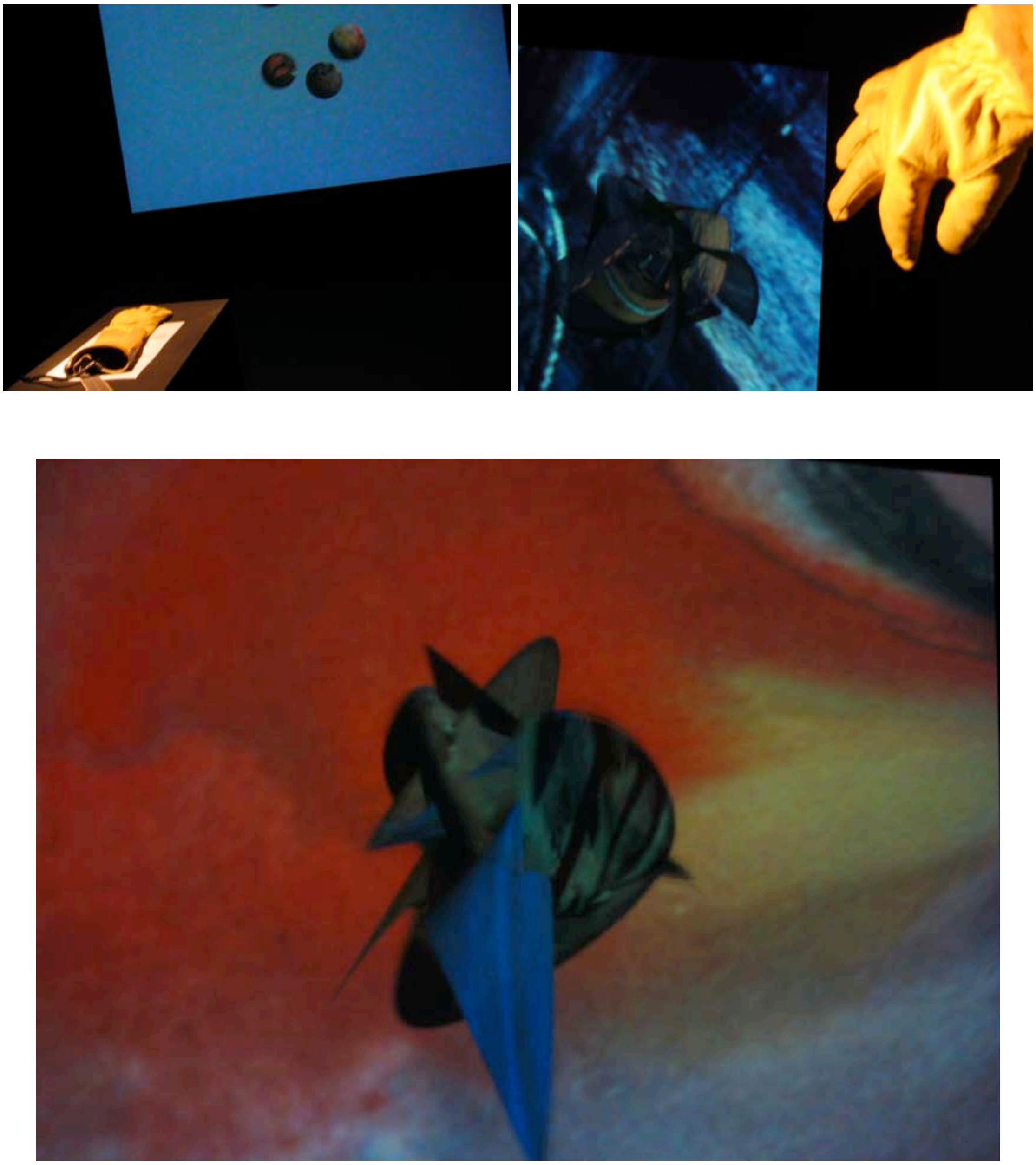

Fig. 5-2*; SoundGlove II in passive and active mode, screenshot of texturised OpenGL; from (Graugaard 2005b).*

The visuals consist of texturized OpenGL objects. The principal object is the sphere, of which several are combined to form the environment. These objects' variable textures are determined by a short-span user analysis taking place during the first 5-10 seconds after a new user is registered by the system. The user then influences a complex graphic element which represents the immediate user characteristics in its shape, primarily through the number of control points, their position, and movement. A non-uniform rational b-spline surface (NURBS) was chosen for its computational efficiency in representing a complex shape. The textures are an ordered set of gouache paintings and the initial analysis determines the appropriate texture for the NURBS and the complementary textures for the spheres (Fig. 5-2).

5.1.5 A USER CYCLE

The audio-visual representation has an active and a passive state. While the active state of the visual component uses the elements mentioned in the previous subsection, the passive visual state represents the system as a world seen from afar. The foreground consists of small spheres referring to the scenario of the previous user and the background is a slowly morphing tiled surface. Upon registering the presence of a new user the visual component descends onto this user's world where the background is the texturized surface in front of which the texturized and real-time animated NURBS gradually appears. The user is then able to move around in the world and by doing so affects the NURBS. The music component likewise classifies a new user during the first 5-10 seconds and makes a choice among several parameter settings for the sound synthesis methods and low- and mid-level settings. These parameters are then further affected by the user input in accordance with the music component's interpretation of the PDS and SDS through the compositional commonologies space. The passive mode is an attenuated and frozen setting based on the previous user's scenario.

5.2 DISCUSSION

Physiological data can be used in a nonexpert system such as a museum installation for visitors to explore. An intuitive feedback loop to the user is important where depth-of-expression can be achieved without any user learning required. Such a system may take into account the impact from its users when in use for an extended duration, hereby providing variety in accordance with its surroundings. More sensors would enhance flexibility of *SoundGlove II* for use in a variety of musical situations. 3D tracking seems a promising combination to the sEMG sensor and fitting a GSR sensor in the glove's fingers would also provide relevant data. Mapping techniques with fuzzy logic could be used for mapping to specific emotion categories, and improved break-down into PDS and SDS data-streams could provide more data depth. When more accuracy is achieved it will be possible to use physiological data analysis to navigate with reasonable accuracy, even though it will be a navigation tool that will be more responsive to – involuntary – emotion response than to conscious control. The software component could have user configurable settings.

6. CONCLUSIONS

The dissertation presented artistic work that unites a compositional argument or its procedures with specially crafted interactive systems. The opera *La Quintrala* is inseparable from the software environment designed to align functionally and artistically the singers with the digital accompaniment, both for practical and artistic purposes. The installation *SoundGlove II* explores a physiological signal and attempts to extract a profile of the user, define a sonic and visual world, and allow the user to explore this world. The compositions *Stgo* and *Gestures You Made* integrate human gesture with sound synthesis and music rendering and fuse the performer's stage presence into the musical work.

6.1 MAIN CONTRIBUTIONS

The dissertation's main contribution was to identify an enlarged disciplinary context for interactive music based on recent research from affective computing. In doing so, it puts forward

- An inclusive understanding of the computer as a musical instrument possessing specific interaction functionalities different from acoustic instruments.
- An identification of vision and human gesture as modalities influential in music perception.
- Connections between high-level human emotion and musical expression as a feasible tool in creation.
- The fundamentally uncategorisable nature of such inclusive interactive music systems' appearance.

The contribution was placed in relation to Lesaffre's taxonomy for musical feature extraction and analysis (Lesaffre et al. 2003), hereby relating to all features of music as a perceptual manifestation.

6.2 SPECIFICS

The dissertation showed that a digital music system only can be classified as a musical instrument when it possesses real-time sensing and actuating properties at a level comparable to the music-mediated interaction between a performer and an acoustic instrument. The system's means of sound production places it in an instrument category of its own, next to the acousitic instrument groups. Its ability to store and algorithmically process data sets it apart from all acoustic music instruments. The system's interactivity is shapeable and dynamic, and is the composer-programmer's extension of the musical composition into the digital system. Distinction between the interactive system and its artistic output is difficult – if at all possible – when the system contains a unique set of properties determined by the intentions of the output.

Signal analysis in an interactive music system is not restricted to audio signals only, but can be other input data that carry human emotion manifestations related to musical expression such as human local and full-body gesture, and physiological signals. It was shown in the dissertation that this is not alien to music performance, because other modalities besides sound may be highly influential in music performance and consumption.

Emotion recognition was pointed to as a relevant high-level link between performer and a digital system. Aspects of a person's emotional state are present in low-level data of sound, gesture, and physiology. Such data can be analysed into basic emotion categories or in a arousal/valence space, and can be combined with regular sound and vision analysis in creating expressive intermodal artworks. The dissertation pointed out that digital musical expression based on synthetic emotions is feasible, and that synthetic emotions need not be restricted to human emotion categories as long as their manifestation in sound or otherwise is intuitively appreciable by humans.

6.3 CIRCUMSCRIBING DIGITAL MUSIC SYSTEMS

The attributes of the computer in interactive music identifies it as inhabiting a new instrument category. But the instantiation of one such instrument is an instantiation of one possibility in a digital system space which should be judged on its ability to instantiate many and vastly different such instruments. Furthermore, a particular digital system commonly achieves sophistication at the expense of generalisation. This indicates that the methods available to us are still rudimentary, since so much fine-tuning of the system is required in order to have the machine participate with sensing and actuating at a level that satisfies our appreciation of its functionality. The expressive capabilities of the digital system are consequently best displayed in the particular piece for which it was intended. It is conceivable that a system designed for one piece with some extension to its parameters and the control options could be usefull in other pieces but they will have to be of a nature similar to the original. It doesn't seem possible to create a generalised instrument covering any major field in interactive music and speculating about which and how many categories of systems needed to cover the entire field seems premature as long as we have no way of knowing if we have covered the field and exhausted its possibilities.

In general terms a musical instrument is of lesser interest to a listener (except to specialists) because it is the music produced with it that is its aim and purpose. If we appreciate the violin we do so for the music that can be performed on it, and not for the violin's sound out of context. Granted, we may very well appreciate its sound but what makes the sound come alive and express emotions is the music performed with it. Out of context the violin is just a – highly sophisticated – sound-source. Likewise, we may design a sophisticated interactive music system but it is ultimately judged on the music we perform with it. We must still allow further work to be done with digital music systems so that we slowly can fathom their nature, before we undertake to identify their combined nature.

One aspect of this nature is the unique way in which an interactive digital system becomes a digital 'score' and an integral part of the piece it is intended for performing. This, given the unprecedented aspects of digital music systems, means that they are very likely to change what we perceive as being 'musical', and what the subject areas relevant to the performance of such 'music' is. This is indeed what have characterised the evolution of the acoustic musical instruments, and a fluid denotation must be maintained to allow for these new subject areas to be included into the new context. Satisfactory use of interactive music systems therefore require knowledge of a vast range of subject areas in computer science, engineering, and the arts but the process of solving artistic and technological challenges of interactive music will slowly unravel the identity of the systems needed for its performance. Significantly, (Jordá 2005) puts forward 25 statements approaching different issues when creating musical instruments based on computers, or 'musical computers', rather than attempting to define the utopian 'perfect' digital instrument.

6.4 THE ARTISTIC POTENTIAL

The creative possibilities exposed by the dissertation are multi-layered, multi-directional, and multi-contextual. Human expression is multi-modal. The layering of modalities in a digital music system into a coherent and unified whole therefore integrates human expression. Identification of such digitally extended expression is a fertile ground for developing new characteristics of expression, where human perception in the arts is challenged and furthered by expanded digital systems. The artistic success of the musical works included in the dissertation is appreciated through the combination of artistic concepts founded on tradition and their articulation in this expanded creative scenario. The content is appreciated through the existing perceptual channels and on the prevailing cultural backdrop, but it is furthered by the explicit completeness of the embodied augmented reality that digital music systems are capable of. The increase in recent years in artistic and scientific research and experiments in the field can only be judged to continue and further expand the field and our understanding, of which the present dissertation is an example.

7. REFERENCES

All references quoted in the dissertation are listed here. References labelled (**D**) are quoted in the critical appraisal of the dissertation. The references quoted in the papers constituting this dissertation are labelled with their corresponding roman numbers (**I**, **II**, **III**).

(Adorno 1997)

Adorno, Th. W.: *Aesthetic Theory*. University of Minnesota Press, ISBN 0-8166-799-6. (**I**)

(Agarabi et al. 2004)

Agarabi, M., Bonato, P., De Luca, C. J.: A sEMG-based Method for Assessing the Design of Computer Mice. *The 26'th International Conference of the IEEE Engineering in Medicine and Biology Society*. San Francisco, September 2004; pp. 301-302. (**III**)

(Berry et al. 2003)

Berry, R., Dahlstedt, P.: Artificial Life: Why Should Musicians Bother? *Contemporary Music Review*, 22:3, pp..57-67, Taylor & Francis Ltd 2003, ISSN 0749-4467. (**D**)

(Bresin et al. 2000)

Bresin, R., Friberg A.: Emotional Coloring of Computer-Controlled Music Performance. *Computer Music Journal* 24:4:44-63. (**D**)

(Cai et al. 2004)

Cai, R., Lu, L., Zhang, H-J., and Cai, L-H.: Improve Audio Representation by Using Feature Structure Patterns. *ICASSP04*. (**II**)

(Camurri et al. 1999)

Camurri, A., Hashimoto, S., Ricchetti, M., Suzuki, K., Trocca, R., Volpe, G.: *KANSEI Analysis of Movement in Dance/Music Interactive Systems* (**D**)

(Camurri et al. 2000)

Camurri, A., Dillon, R., Saron, A.: An Experiment on Analysis and Synthesis of Musical Expressivity. *Proceedings of the XIII Colloquium on Music Informatics*. L'Aquila, Italy. (**D**, **III**)

(Camurri et al. 2002)

Camurri, A., Trocca, R., Volpe, G.: Interactive Systems Design: A KANSEI-based Approach. *Proceedings of NIME 2002*. Dublin, Ireland (**D**, **III**, **II**)

(Camurri et al. 2005b)

Camurri, A., Volpe, G., De Poli, G., Leman, M.: Communicating Expressiveness and Affect in Multimodal Interactive Systems. *IEEE Magazine* January-March 2004 pp. 43-53 (**D**)

(Canazza et al. 1999)

Canazza, S., Rodá, A.: Adding Expressiveness in Musical Performance in Real Time. *Proceedings of the AISB*, Symposium on Musical Creativity, pp. 134-139; Edinburgh, Scotland. (**III**)

(Canazza et al. 2001)

Canazza, S., De Poli, G., Rodà, A., Vidolin, A., Zanon, P.: Kinematics-energy Space for Expressive Interaction in Music Performance. *Proceedings of MOSART, Workshop on current research directions in Computer Music*; pp. 35-40; Barcelona, Spain. (**II**)

(Canazza et al. 2002)

Canazza, S., De Poli, G., Rodá, A., Soleni, G., Zanon, P.: Real-time Analysis of Expressive Contents in Piano Performances. *Proceedings of the 2002 International Computer Music Conference*. pp. 414-418. Gothenburg, Sweden. (**D**, **III**)

(Canazza et al. 2004)

Canazza, S.; De Poli, G.; Drioli, C.; Roda, A.; Vidolin, A.: Modeling and Control of Expressiveness in Music Performance. *Proceedings of the IEEE*, Volume 92, Issue 4, April 2004 pp. 686 - 701. (**D**)

(Cano et al. 1999)

Cano, P., Loscos, A., Bonada, J.: Score-performance Matching using HMMs. *Proceedings of the 1999 International Computer Music Conference*. Beijing, China. (**II**)

(Cascone 2000)

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

(Choi 1997)

Choi, I.: Interactivity vs. Control: Human-machine Performance Basis of Emotion. *Proceedings of 1997 'KANSEI - The Technology of Emotion Workshop'*. pp. 24–35. (**D**, **III**)

(Cram 1986)

Cram, J. R.: Clinical EMG for Surface Recordings. *J&J Engineering Volume 1*, Poulsbo, WA. (**III**)

(Creed et al. 2005)

Creed, C., Beale, R.: Using Emotion Simulation to Influence User Attitudes and Behaviour. *Proceedings of the HCI 2005*; Edinburgh, September 2005. (**III**)

(Dahl et al. 2004)

Dahl, S., Friberg, A.: 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. (**D**)

(Dalton 2004)

Dalton, B.: Creativity, Habit, and the Social Products of Creative Action: Revising Joas, Incorporating Bourdieu. *Sociological Theory* 22:4, December 2004; pp. 603-622. (**III**)

(Damasio 1994)

Damasio, A. R.: *Descartes' Error: Emotion, Reason, and the Human Brain*. Gosset/Putnam Press, New York, NY 1994, ISBN 0380726475. (**D**)

(Dannenberg 1984)

Dannenberg, R.: An On-line Algorithm for Real-Time Accompaniment. *Proceedings of the 1984 International Computer Music Conference*. pp. 193-198. Paris, France. (**D**, **III**)

(Dannenberg et al. 1988)

Dannenberg, R. and Mukaino, H.: New Techniques for Enhanced Quality of Computer Accompaniment. *Proceedings of the 1988 International Computer Music Conference*, pp. 243-249. Cologne, Germany. (**II**)

(Davidson et al. 2002)

Davidson, J. W., Correia, J. S.: Body Movement. In R. Parncutt & G. E. McPherson (Eds.): *The science and psychology of music performance. Creative strategies for teaching and learning*, pp. 237-250, Oxford University Press. (**D**)

(De Lio 1984)

De Lio, T.: *Circumscribing the open universe : Essays on John Cage, Morton Feldman, Christian Wolff, Robert Ashley, Alvin Lucier*. University Press of America, ISBN 0819137480 (**I**)

(De Luca et al. 2004)

De Luca, G. N., Johansson, A., Roy, S. H., Cheng, MS., Gilmore, L. D., Bergman, P., De Luca, C. J.: *Sweat Test for Electro-mechanical Stability of the EMG Electrode-skin Interface*. www.delsys.com/library/applications.htm, 8/7 2005. (**III**)

(DelSys Inc. 2005)

DelSys Inc.: *Surface Electromyograph*. www.delsys.com, 8/7 2005 (**III**)

(Forte 1973)

Forte, A.: *The Structure of Atonal Music*. Yale University Press. New Haven and London. ISBN 0300021208. (**II**)

(Freed et al. 2000)

Freed, A., Isvan, O.: Musical Applications of New, Multi-axis Guitar String Sensors. Presented at *International Computer Music Conference 2002*, Berlin. (**III**)

(Freescale Semiconductor 2005)

Freescale Semiconductor: *Altivec Technology*; www.simdtech.org/altivec, 11/8 2005. (**II**)

(Friberg 2004)

Friberg, A.: A fuzzy analyzer of emotional expression in music performance and body motion. In J. Sundberg & B. Brunson (Eds.): *Proceedings of Music and Music Science*. Stockholm, October 28-30 2004 (**D**, **III**)

(Friberg 2005)

Friberg, A.: 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 (**D**)

(Friberg et al. 2002)

Friberg, A., Schoonderwaldt, E., Juslin, P. N., Bresin, R.: Automatic Real-Time Extraction of Musical Expression. *Proceedings of the 2002 International Computer Music Conference*, pp. 365-367; Gothenburg, Sweden. (**D**, **III**)

(Funkhauser et al. 2002)

Funkhauser, T., Jot, J-M, Tsingos, N.: Sounds Good to Me. *SIGGRAPH 2002* Course Notes. (**III**)

(Gabor 1947)

Gabor, D.: Acoustical Quanta and the Theory of Hearing. *Nature* 159 (1044); pp. 591-594. (**D**)

(Gabrielsson et al. 1996)

Gabrielsson, A., Juslin, P.: Emotional Expression in Music Performance. *Psychology of music* 24, pp. 68-91. (**D**)

(Graugaard 2001)

Graugaard, L.: *GUITAR*. Composition for guitar and interactive, computer generated accompaniment. Premiered March 3rd 2001 at Carl-Nielsen Academy of Music, Denmark. (**D**)

(Graugaard 2003-04)

Graugaard, L.: *La Quintrala*. Opera for five singers and interactive, computer generated accompaniment. Premiered September 2nd 2004 at Den Anden Opera, Copenhagen, Denmark. (**D**, **III**)

(Graugaard 2004b)

Graugaard, L.: Open and Closed Form in Interactive Music. In U. Kock Wiil (ed.) *Computer Music Modeling and Retrieval, Second International Symposium CMMR 2004*; Esbjerg, Denmark, May 26-29 2004; Revised Papers; pp. 149-157; Springer Verlag Berlin Heidelberg; ISBN 3-540-244858-1. (**D**)

(Graugaard 2005a)

Graugaard, L.: Unifying Performer and Accompaniment. In R. Kronland-Martinet, T. Voinier, and S. Ystad (Eds.): *Computer Music Modeling and Retrieval, Third International Symposium CMMR 2005*. Springer Verlag Berlin Heidelberg. (**D**)

(Graugaard 2005b)

Graugaard, L.: The SoundGlove II: using sEMG data for intuitive audio and video affecting in real-time. *Proceedings of The Third Annual Conference in Computer Game Design and Technology*; pp. 157-161; Liverpool, UK. (**D**)

(Graugaard 2005c)

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

(Graugaard 2005-06)

Graugaard, L.: *Gestures You Made*. Composition for oboe and interactive, computer generated accompaniment. Premiered February 12th 2006 by Eydís Franzdottír; Dark Music Days 2006, Reykjavik, Iceland. (**D**)

(Graugaard 2006a)

Graugaard, L.: Sound Gestures Affected by Physical Gestures in Real-time. *Proceedings of the International Computer Music Conference 2006*, New Orleans, Louisiana. (**D**)

(Grewe et al. 2005)

Grewe, O., Nagel, F., Altenmüller, E., Kopiez, R.: Psychological and Physiological Correlates of Strong Emotions in Music. *Proceedings of the NWG Conference*, Göttingen (**D**)

(Grubb et al. 1997)

Grubb L., Dannenberg, R.: A Stochastic Method of Tracking a Vocal Performer. *Proceedings of the 1997 International Computer Music Conference*. pp. 301-308. Thessaloniki, Greece. (**II**)

(Hanna et al. 2002)

Hanna, P., Desainte-Catherine, M.: Detection of Sinusoidal Components in Sounds using Statistical Analysis of Intensity Fluctuations. *Proceedings of the 2002 International Computer Music Conference*, pp. 100-103. Gothenburg, Sweden. (**II**)

(Hashimoto 1997)

Hashimoto, S.: 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. (**D**, **III**)

(Healey et al. 1998)

Healey, J., Picard, R. W.: Digital Processing of Affective Signals, MIT Technical Report No. 444. *Proceeding of the ICASSP'98*. (**D**)

(Heinz et al. 1996)

Heinz, M., Knapp, R. B.: Pattern Recognition of the Electromyogram using a Neural Network Approach. *Proceedings of the IEEE International Conference on Neural Networks*. Washington DC, USA (**II**)

(Herbelin et al. 2004)

Herbelin, B., Benzaki, P., Riquier, F., Renault, O.,Thalmann, D.: Using Physiological Measures for Emotional Assessment: A Computer-aided Tool for Cognitive and Behavioural Therapy. *Proceeedings of the 5th International Conference on Disability, Virtual Reality & Associated Techniques*; pp. 307-314, Oxford, UK. (**D**)

(Inoue et al. 1993)

Inoue, W., Hashimoto, S., Ohteru, S.: A Computer Music System for Human Singing. *Proceedings of the 1993 International Computer Music Conference*, pp. 150-153; Tokyo, Japan. (**II**)

(Inoue et al. 1994)

Inoue, W., Hashimoto, S., Ohteru, S.: Adaptive Karaoke System-human Singing Accompaniment Based on Speech Recognition. *Proceedings of the 1994 International Computer Music Conference*, pp. 70-77. Århus, Denmark. (**II**)

(Jehan et al. 2001)

Jehan, T., Schoner, B.: An Audio-Driven Perceptually Meaningful Timbre Synthesizer. *Proceedings of the 2001 International Computer Music Conference*. Havana, Cuba. (**D**, **III**)

(Jensen 1999a)

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

(Jordá 2005)

Jordá, S.: *Digital Lutherie, Crafting Musical Computers for New Musics' Performance and Improvisation*. PD thesis. (**D**)

(Juslin 2000)

Juslin, P. N.: Cue Utilization in Communication of Emotion in Music Performance: Relating Performance to Perception. *Journal of Experimental Psychology: Human Perception and Performance*, 26, 1797-1813. (**D**)

(Katayose 1993)

Katayose, H., Kanamori, T., Kamei, K., Nagashima, Y., Sato, K., Inokuchi, S., Simura, S.: Virtual Performer. *Proceedings of the 1993 International Computer Music Conference*, pp. 138-145. Tokyo, Japan. (**II**)

(Kim et al. 2004)

Kim, S., André, E.: Composing Affective Music with a Generate and Sense Approach. *Proceedings of the 2004 AAAI*. (**III**)

(Knapp et al. 1990)

Knapp, R. B., Lusted, H. S.: A Bioelectric Controller for Computer Music Applications. *Computer Music Journal*, MIT Press, Vol. 14, No. 1, pp. 42-47. (**III**)

(Laban 1963)

Laban, R.: *Modern Educational Dance*. MacDonald & Evans, London, UK; ISBN 0712113819 (**II**)

(Lang 1995)

Lang, P.: The Emotion Probe: Studies of Motivation and Attention. *American Psychologist*, Vol. 50, No. 5, pp. 372-385 (**III**)

(Langer 1951)

Langer, S. K.: *Philosophy in a New Key*. Harvard University Press. New American Library, New York, New York, ISBN 0674665031. (**III**)

(Lawter et al. 1998)

Lawter, J., Moon, B.: Score Following in Open Form Compositions. *Proceedings of the 1998 International Computer Music Conference*, San Francisco, USA. (**II**)

(Lee et al. 2004)

Lee, E., Marrin Nakra, T., Borchers, J.: You're the Conductor: A Realistic Interactive Conducting System for Children. *Proceedings of the 2004 LEE*. (**III**)

(Leman 2002)

Leman, M.: Musical Audio Mining. *Dealing with the Data Flood: Mining Data, Text, and Multimedia*. SST Netherlands Study Center for Technology Trends, Rotterdam. (**D**)

(Leman 2006a)

Leman, M.: Understanding Musical Expressiveness Using Interactive Multimedia Platforms. *Musica Scientiae*, 2006. (**D**)

(Leman 2006b)

Leman, M.: *Embodied Music Cognition and Mediation Technology*. In press (**D**)

(Leman et al. 2004)

Leman, M., Camurri, A.: Musical Content Processing for Expressive Gesture Applications in Interactive Multimedia. *Proceedings of the Conference on Interdisciplinary Musicology*. Graz, Austria. (**D**)

(Lesaffre et al. 2003)

Lesaffre, M., Leman, M., Tanghe, K., De Baets, B., De Meyer, H., Martens, J.-P.: User-Dependent Taxonomy Of Musical Features As A Conceptual Framework For Musical Audio-Mining Technology. *Proceedings of the Stockholm Music Acoustics Conference*. Stockholm, Sweden. (**D**)

(Lippe 1994)

Lippe, C.: Real-Time Granular Sampling using the Ircam Signal Processing Workstation. *In Contemporary Music Review*, Vol. 10, United Kingdom, Harwood Academic Publishers, pp. 149-155 (**I**)

(Lippe 1996)

Lippe, C.: A Look at Performer/Machine Interaction Using Real-Time Systems. *Proceedings of the International Computer Music Conference 1996*, pp. 116-117, Hong Kong. (**D**, **III**)

(Lippe 2002)

Lippe, C.: Real-Time Interaction Among Composers, Performers, and Computer Systems. *Information Processing Society of Japan SIG Notes*, Volume 2002, Number 123, pp. 1-6. (**D**, **III**)

(Lusted et al. 1996)

Lusted, H. S., Knapp, R. B.: Controlling Computers with Neural Signals. *Scientific American* (**III**)

(Marrin Nakra 2000)

Marrin Nakra, T.: Inside the Conductor's Jacket: Analysis, interpretation and musical synthesis of expressive gesture; PhD thesis, Massachusetts Institute of Technology. (**III**)

(Martin et al. 1998)

Martin, K. D., Scheirer, E. D., and Vercoe, B. L.: Music Content Analysis through Models of Audition. *Proceedings of the ACM Multimedia '98 Workshop on Content Processing of Music for Multimedia Applications*, Bristol UK (**D**, **III**)

(Modler et al. 2003)

Modler, P., Myatt, T., Saup, M.: 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*. Montreal, Canada. (**D**)

(Moore 1990)

Moore, F. R.: *Elements of Computer Music*. Prentice Hall, 1990, ISBN 0-13-252552-6 (**D**, **III**)

(Mulder et al. 1997)

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

(Nagashima 2003a)

Nagashima, Y.: Bio-sensing Systems and Bio-feedback Systems for Interactive Media Arts. *Proceedings of NIME 2003*, pp. 48-53. (**III**)

(Nagashima 2003b)

Nagashima, Y.: Combined Force Display System of EMG Sensor for Interactive Performance. *Proceedings of the 2003 International Computer Music Conference*; Singapore. (**III**)

(Nagashima 2004)

Nagashima, Y.: Controlling Scanned Synthesis by Body Operation. *Proceedings of the ICA 2004*, pp. 134-139; Kyoto, Japan. (**III**)

(Orio et al. 2003)

Orio, N., Lemouton, S., Schwarz, D.: Score Following: State of the Art and New Developments. *Proceedings of the 2003 Conference on New Interfaces for Musical Expression*. Montreal, Canada. (**II**)

(Paine 2002)

Paine, G.: Interactivity, Where From Here? *Organised Sound* 7(3): 295–304, Cambridge University Press, United Kingdom. (**D**)

(Penny 2000)

Penny, S.: Agents as Artworks and Agent Design as Artistic Practice. In K. Dautenhahn (ed.): *Human Cognition and Social Agent Technology*, John Benjamins Publishing Company, Amsterdam, ISBN 1-556-19435-8. (**III**)

(Picard 1995)

Picard, R. W.: *Affective Computing*. MIT Media Lab Perceptual Computing Section Technical Report No. 321. (**D**)

(Picard 1997)

Picard, R. W.: *Affective Computing*. The MIT Press, Cambridge, MA, 1997. ISBN 0262161702. (**III**)

(Picard et al. 2001)

Picard, R. W., Vyzas E., Healey, J.: Towards Machine Emotional Intelligence: Analysis of Affective Physiological State. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(10). (**D**)

(Puckette et al. 1992)

Puckette, M., Lippe, C.: Score Following in Practise. *Proceedings of the 1992 International Computer Music Conference*, pp. 182-185. (**II**)

(Puckette 1995)

Puckette, M.: Score Following using the Sung Voice. *Proceedings of the 1995 International Computer Music Conference*. Banff, Canada. (**II**)

(Puckette et al. 1998)

Puckette, M., Apel, T., Zicarelli, D.: Real-time Audio Analysis for Pd and MSP. *Proceedings of the 1998 International Computer Music Conference*, pp. 109-112. San Francisco, USA. (**I**, **III**)

(Rani et al. 2003)

Rani, P., Sarkar, N., Smith, C.: Affect-Sensitive Human-Robot Cooperation, Theory and Experiments. *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 232-238, Taiwan, September 2003. (**D**)

(Roebel et al. 2004)

Roebel, A., Zivanoviz, M., Rodet, X.: Signal Decomposition by means of Classification of Spectral Objects. *Proceedings of the 2004 International Computer Music Conference*, pp. 446-449. Florida, USA. (**II**)

(Rokeby 1998)

Rokeby, D.: The Construction of Experience: Interface as Content. In C. Dodsworth Jr. (ed.): *Digital Illusion: Entertaining the Future with High Technology*, ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, ISBN 0-201-84780-9. (**III**)

(Rowe 2001)

Rowe, R.: *Machine Musicianship*, MIT Press, ISBN 0-262-68149-8. (**D**, **III**)

(Rutherford et al. 2003)

Rutherford, J., Wiggins, G. A.: An Experiment in the Automatic Creations of Music which has Specific Emotional Content. *Proceedings of MOSART*, Workshop on current research directions in Computer Music; pp. 35-40; Barcelona, Spain. (**III**)

(Scherer et al. 2001)

Scherer, K. R., Zentner, M. R.: Emotional Effects of Music: Production Rules. In Juslin, P.N. & Sloboda, J.A. (eds.): *Music and emotion: theory and research*. Oxford; New York, Oxford University Press. (**D**)

(Smalley 1997)

Smalley, D.: Spectromorphology: Explaining Sound-shapes. *Organised Sound* 2(2), pp. 107-126; Cambridge University Press, United Kingdom 1997. (**D**)

(Sundberg 1993)

Sundberg, J., D.: How Can Music Be Expressive? *Speech Communication*13:239–253. (**D**)

(Suzuki et al. 2004)

Suzuki, K., Hashimoto, S.: Robotics Interface for Embodied Interaction via Dance and Musical Performance. *Proceedings of the IEEE*, vol. 92, no. 4, pp. 656-671. (**D**)

(Tanaka et al. 2002)

Tanaka, A., Knapp, R. B.: Multimodal Interaction in Music Using the Electromyogram and Relative Position Sensing. *Proceedings of the NIME 2002*. Medialab Europe. (**D**, **II**)

(Vercoe 1988)

Vercoe, B.: The Synthetic Performer in the Context of Live Performance. *Proceedings of the 1988 International Computer Music Conference*, pp. 199-200. Paris, France. (**D**, **III**)

(Wanderley 2000)

Wanderley, M.: Mapping Strategies. In Wanderly, M., Battier, M. (eds.): *Trends in Gestural Control of Music*; Ircam, Paris, 2000 (**D**)

(Wang et al. 2004)

Wang, H., Prendinger, H., Igarashi, T.: Communicating Emotions in Online Chat Using Physiological Sensors and Animated Text. *Proceedings of the 1st international conference for human-computer interaction*, Vienna, Austria, pp. 1171-1174. **(D**)

(Xenakis 1971)

Xenakis, I.: *Formalized Music*. Bloomington: Indiana University Press. ISBN 0253323789. (**D**)

(Zagalo et al. 2004)

Zagalo, N., Barker, A., Branco, V.: Story Reaction Structures to Emotion Detection. *SRMC'04*, October 15, 2004, New York, New York, USA. (**D**)

(Zannos et al. 1997)

Zannos, I., Modler, P., Naoi, K.: Gesture Controlled Music Performance in a Real-time Network. *Proc. KANSEI - The Technology of Emotion Workshop*, pp. 60–63, 1997. (**D**)

(Zanon et al. 2003)

Zanon, P., & De Poli, G.: Estimation of Parameters in Rule Systems for Expressive Rendering of Musical Performance. *Computer Music Journal*, 27(1), 29-46. (**D**)

(Zicarelli 2004)

Zicarelli, D.: Cycling74 website at *www.cycling74.com*. Visited 3/3-2004. (**I**)

APPENDIX I. INCLUDED PAPERS

(Graugaard 2004b)	Graugaard, L.: Open and Closed Form in Interactive Music. In U. Kock Wiil (ed.) *Computer Music Modeling and Retrieval, Second International Symposium CMMR 2004*; Esbjerg, Denmark, May 26-29 2004; Revised Papers; pp. 149-157; Springer Verlag Berlin Heidelberg; ISBN 3-540-244858-1
(Graugaard 2005a)	Graugaard, L.: Unifying Performer and Accompaniment. In R. Kronland-Martinet, T. Voinier, and S. Ystad (Eds.): *Computer Music Modeling and Retrieval 2005*, LNCS 3902, pp. 169-184, Springer-Verlag, Berlin Heidelberg 2006
(Graugaard 2005b)	Graugaard, L.: The SoundGlove II: using sEMG Data for Intuitive Audio and Video Affecting in Real-time. In *Proceedings of The Third Annual Conference in Computer Game Design and Technology*; pp. 157-161; 8th-9th November 2005, Liverpool, UK

APPENDIX II. FURTHER PAPERS

(Graugaard 2006a)	Graugaard, L.: Sound Synthesis affected by Physical Gestures in Real-time. *Proceedings of the International Computer Music Conference 2006*, New Orleans, Louisiana. (**D**)

APPENDIX III. INCLUDED SOFTWARE

(Graugaard 2001)	Graugaard, L.: *GUITAR*. Composition for guitar and interactive, computer generated accompaniment. Premiered March 3rd 2001 at Carl-Nielsen Academy of Music, Denmark.
(Graugaard 2003-04)	Graugaard, L.: *La Quintrala*. Opera for five singers and interactive, computer generated accompaniment. Premiered September 2nd 2004 at Den Anden Opera, Copenhagen, Denmark.
(Graugaard 2004-05)	Graugaard, L.: *The SoundGlove II*. Commissioned by Museum of Contemporary Art, Roskilde (Denmark), displayed September 30th – December 18th 2005 at the Danish Museum of Contemporary Art, Roskilde, Denmark.
(Graugaard 2005c)	Graugaard, L.: *Stgo*. Premiered December 6th 2005 by Lars Graugaard; Cech Santiago, Chile.
(Graugaard 2005-06)	Graugaard, L.: *Gestures You Made*. Composition for oboe and interactive, computer generated accompaniment. Premiered February 12th 2006 by Eydís Franzdottír; Dark Music Days 2006, Reykjavik, Iceland.

APPENDIX IV. INCLUDED SCORES

(Graugaard 2001) Graugaard, L.: *GUITAR*. Composition for guitar and interactive, computer generated accompaniment. Premiered March 3rd 2001 at Carl-Nielsen Academy of Music, Denmark.

(Graugaard 2003-04) Graugaard, L.: *La Quintrala*. Opera for five singers and interactive, computer generated accompaniment. Premiered September 2nd 2004 at Den Anden Opera, Copenhagen, Denmark.

(Graugaard 2005-06) Graugaard, L.: *Gestures You Made*. Composition for oboe and interactive, computer generated accompaniment. Premiered February 12th 2006 by Eydís Franzdottír; Dark Music Days 2006, Reykjavik, Iceland.

APPENDIX V. INCLUDED RECORDINGS

(Graugaard 2001) Graugaard, L.: *GUITAR*. Composition for guitar and interactive, computer generated accompaniment. Premiered March 3rd 2001 at Carl-Nielsen Academy of Music, Denmark.

(Graugaard 2005-06) Graugaard, L.: *Gestures You Made*. Composition for oboe and interactive, computer generated accompaniment. Premiered February 12th 2006 by Eydís Franzdottír; Dark Music Days 2006, Reykjavik, Iceland.