SIEMPRE

Social Interaction and Entrainment using Music PeRformancE

SIEMPRE

Deliverable D1.1: Research Requirement

Version	*Edited by*	*Changes*
V1.0-1.7	Carolina Labbé, Kim Eliard, Didier Grandjean, UNIGE-CH	
1.1	Tom Cochrane, QUB	Mostly to sections on co-creation and leadership.
1.2	Antonio Camurri and Giovanna Varni, UNIGE	
1.3	Benjamin Knapp, QUB	Measurement of physiological signals
1.4	Cian Doherty, QUB	Quality of experience
1.5	Esteban Maestre, UPF	Audio acquisition and analysis
1.6	Luciano Fadiga and Alessandro D'Ausilio, IIT	
1.7	Donald Glowinski, Antonio Camurri, and Barbara Mazzarino, UNIGE	

TABLE OF CONTENTS

1. INTRODUCTION

This document presents the research requirements of the SIEMPRE project.

Research in SIEMPRE is focused on the development of theoretical models and analysis techniques of non-verbal social signals. SIEMPRE considers ensemble musical performance and audience experience as the ideal setups. SIEMPRE aims at developing novel research theoretical and methodological frameworks, computational models, and algorithms for the analysis of creative communication within groups of people.

The main research challenges studied can be summarized by the following questions:

1. What are the key factors driving the interpersonal synchronization of the participants (e.g., visual and auditory expressive cues, rules and conventions)? In what ways do these vary according to the physical, emotional and social contexts?
2. Can specific roles inside the group be identified (e.g., leadership, hierarchy)? Can general principles be discerned concerning the influence of some individuals over others?
3. What are the factors that determine feelings of group cohesion or a sense of shared meaning? Can the validity of such reports by participants be supported or confirmed by a set of appropriate expressive multimodal features (eMAP – expressive movement audio physiological signals)?
4. How does social context affects individual intrapersonal synchronization of eMAP and vice-versa? In what ways does the emotion of the individual affect the collaborative creative product?
5. Which are the neurophysiological foundations of creative group communication?

In order to draw distinctions between the concepts under study and to come to an agreement concerning the terminology used and thereby facilitate communication between the partners (especially where sharing of experimental protocols and designs are concerned), a conceptual framework was put together and constitutes one of the first achievements of the project. The first version of the framework was proposed by UNIGE-CH for the Theoretical Workshop of December 3rd 2010 at the Swiss Center for Affective Sciences in Geneva. An updated version of that work with added contributions from QUB is presented in the first part of this deliverable. The second part of this deliverable presents the research methodology in SIEMPRE.

2. CONCEPTUAL FRAMEWORK

We draw largely from Juslin, Liljeström, Västfjäll and Lunqvist's 2010 revision of their BRECVEM model in addition to Juslin and Västfjäl's 2008 review on the possible mechanisms for the induction of musical emotions. In addition we also introduce other relevant concepts and theories in the study of emotional processes and music including the Component Process Model (CPM), emotional contagion, the mirror-neuron systems hypothesis, and the notions of leadership in group interactions. Then, we propose a framework within which to explore the links between music and emotion as well as a glossary of the terminology used.

This document is composed of three major parts: the first describes one integrative theory and its conceptual and methodological approaches to the study of emotion in psychology, while the second focuses on music and emotion in particular. The third provides an integrative view of the theories and methodologies viewed as they pertain to the SIEMPRE project.

2.1 Part 1: conceptual and methodological approaches to the study of emotion in psychology

2.1.1 A few definitions

It is important to begin with at least a working definition of emotion. Though differences do exist between theories of emotion, most authors are increasingly converging in their view of emotion as a componential process. This categorization is not trivial, emotion is viewed as a process in the psychological sense of the word, meaning "[a] sequence of events leading to some change or alteration in the state of a dynamic system" (A Dictionary of Psychology, 2009), where the dynamic system is typically understood as being an organism in interaction with its environment. Further, it is a componential process in the sense that there are different components that constitute such a dynamic organism pattern activation, they typically include but are not limited to: action tendencies, bodily responses, and emotional experience (Moors, 2007). So an emotion can be understood as: a sequence of events leading to some change or alterations in the states of some or all of the components of a dynamic system, i.e. the person.

2.1.2 A few approaches

What causes these changes? According to an appraisal (or cognitivist) view of emotion, it is the appraisal of a stimulus or event which is relevant for the goals or concerns of an organism that will act as the trigger for a response, i.e. the emotion process. This initial "relevance" appraisal then leads to the evaluation of the stimulus, situation or event according to different criteria (the order and number of which being what essentially differentiates appraisal models of emotion from each other) and it is the different combinations of these evaluations that will lead to a unique emotional pattern response. Thus, because different people have different goals or concerns, a same stimulus will rarely be evaluated in the same way and this explains why different people will not react in the same way to the same stimulus or why their responses might change over time. However, even though the cognitive component of emotion is essential for "the continuous, recursive subjective evaluation of events for their pertinence, as well as the coping potential of the individual" (Grandjean, Sander & Scherer, 2008), emotion is more than just the accumulation of cognitive responses to a stimulus according to different evaluative criteria. Indeed, Scherer (2005) has suggested defining emotion in the following manner:

> In the framework of the component process model, emotion is defined as *an episode of interrelated, synchronized changes in the states of all or most of the five organismic subsystems in response to the evaluation of an external or internal stimulus event as relevant to major concerns of the organism* (Scherer, 1987, 2001). The components of an emotion episode are the respective states of the five subsystems and the process consists of the coordinated changes over time (p. 697).

Scherer (2005) then suggests the following relations between the different components and subsystems:

Emotion function	Organismic subsystem and major substrata	Emotion component
Evaluation of objects and events	Information processing (CNS)	Cognitive component (appraisal)
System regulation	Support (CNS, NES, ANS)	Neurophysiological component (bodily symptoms)
Preparation and direction of action	Executive (CNS)	Motivational component (action tendencies)

Communication of reaction and behavioral intention	Action (SNS)	Motor expression component (facial and vocal expression)
Monitoring of internal state and organism–environment interaction	Monitor (CNS)	Subjective feeling component (emotional experience)

Note: CNS = central nervous system; NES = neuro-endocrine system; ANS = autonomic nervous system; SNS = somatic nervous system.

The CPM proposes to explain the differentiation between emotional states as being the result of a sequence of specific stimulus evaluation (appraisal) checks. According to this theory, emotion is the result of cognitive evaluations that the individual makes about an (external or internal) event, or situation, which initiate the emotion. More precisely, emotions are defined as a sequence of state changes in most or all of the organism's 5 systems:

- cognitive (appraisal)
- psychophysiological (peripheral responses)
- motivational (actions tendencies, tendency to respond to the event)
- motor (movement, facial expression, vocalization)
- subjective feeling (emotional experience)

The sets of criteria to evaluate the event are called "stimulus evaluation checks" (SECs). The SECs are organized around four main objectives which are further divided into "sub-goals". The major SECs correspond to the most important information needed by the organism for an appropriate response, namely:

- *relevance:* novelty (suddenness, familiarity, predictability), intrinsic pleasantness, and goal/need relevance.
- *implication:* causal attribution, outcome probability, discrepancy from expectation, goal/need conduciveness, urgency.
- *coping:* control, power, adjustment.
- *normative significance:* internal and external standards.

It can then be said that "emotion are defined and operationalized as complex, multicomponential, dynamic processes that require sophisticated measurement of changes in the different components" (Grandjean, Sander & Scherer, 2008). The notion of synchronization between the different organic subsystems is a crucial concept to operationalize the concept of emotions. In Grandjean, Sander & Scherer, 2008, we define synchronization to refer to the following two phenomena: (1) the establishment of some degree of coherence or synchronization of the different components of emotion during an emotional episode, organized as temporal and functional emergent patterns and (2) the synchronization of neuronal assemblies at the central nervous system level, brought about by functional coupling of different close or distant neuronal populations. These two different phenomena are described below. The main assumption of the component synchronization hypothesis is that emergent emotional processes and the concomitant feelings or emotional experience emerge as a function of a multilevel appraisal-driven response synchronization. Scherer (2001a, 2004) has proposed a componential patterning theory which attempts to predict specific changes in the peripheral subsystems that are brought about by concrete patterns of stimulus evaluation check (SEC) results. The central assumption of the componential patterning theory is that the different subsystems are highly interdependent and that changes in one subsystem will tend to elicit related changes in other subsystems in a recursive fashion. In consequence, the result of each consecutive check will differentially and cumulatively affect the state of all other subsystems. Due to the temporal organization of the appraisal processes, their efferent impact on peripheral, motivational, and motor components is likely to induce some degree of synchronization in these different subcomponents. This is functionally meaningful, as all resources of the organism are recruited to deal with the event that elicited the emotion. The central integration and representation of these synchronized patterns defines the feeling component, which has an important regulatory function. Subjective experience serves a monitoring function, integrating all information about the continuous patterns of change in all other components, as well as their coherence, and then building an integrative conscious representation. The feeling component can be described by a Venn diagram in which a set of overlapping circles represent its different aspects. It has been suggested that the degree of synchronization of the different components might be one of the factors that determines the emergence of consciousness in the emotion process (Grandjean et al. 2008). At the cerebral level, the different neuronal assemblies underlying the processes related to the different components must continuously interact to be able to exchange information and to integrate the multiple representations related to the different components.

Several pieces of evidence indicate that the neuronal synchronization of electrical activity between two or more neuronal assemblies is necessary to allow the communication between distant or local neuronal networks. In particular, Fries (2005) has proposed the communication through coherence (CTC) model, which implies that phase coherence underlies neuronal communication: neuronal assemblies have to be in synchronization to exchange information. Based on this model one can predict that central neuronal regions in emotional processes like the amygdala and orbitofrontal areas have to be synchronized during an emotional episode to exchange information and thus contribute to the organization of the emotional reaction and its representation in terms of feeling. Previous findings suggest that neuronal synchronization may indeed be necessary to process emotional information. For example, Luo et al. (2007) report synchronization between the thalamus/hypothalamus and the amygdala in response to facial threat. Distant synchronization between hippocampus and amygdala has also been shown during various stages of fear memory (Narayanan et al. 2007). These empirical findings of neuronal synchronization in the human brain in response to emotional stimuli highlight the importance of the functional coupling between different distant and local neuronal assemblies and suggest continuous cross-talk between different brain regions during the processing of emotional stimuli. The results of these studies can be interpreted as evidence that different neuronal assemblies, representing different levels of processing in the brain, work in conjunction to assess input of high significance for an individual. This suggestion is reminiscent of the assumption of massive parallel processing in neural network models, and is consistent with a recent proposal of a neural network model of emotional consciousness in which emotional coherence through interactions among multiple brain areas needs to be achieved for emotional consciousness to emerge (Thagard and Aubie, 2008). These synchronizations could occur at different levels, including local and distant neuronal synchronies. In this context it can be assumed that local synchronies, in a specific neuronal network, are necessary to achieve preliminary closure and send information to another neuronal network. For example, to process information relative to the state of the body during an emotional episode, the synchronization of neuronal assemblies inside the insula would be necessary. When a stable representation emerges from this neuronal network, the information might be sent to another part of the brain, for example to prefrontal brain areas (inducing a specific body representation in working memory). In this example, local synchronization is required to build a stable representation, and distant neuronal synchronization is required to exchange this information with another functional unit, in this case the body consciousness state in working memory. The local synchronies of the electrical activity of neuronal assemblies are expected to occur at a higher frequency range compared with the distant synchronies, which should occur at a lower frequency range (Fries 2005). One important implication of the CPM is that we have, as scientists, to dissociate the emotional processes and their subcomponents and the integration and/or the representation of the experienced emotion i.e. the feeling.

This view, i.e. the CPM, is in direct contrast with other theories of emotion, such as *basic theories* of emotion and *bi-dimensional* theories of emotion, that do not necessarily specify “a concrete mechanism underlying emotional response patterning, allowing specific hypotheses”, nor do they make specific predictions about the determinants through which emotion may be elicited and differentiated.

"The basic emotion system" for example proposes the existence of core systems (emotion systems) that develops to deal with stereotypical "species-constant problems related to survival in a time-tested, predictable, and automatic fashion" (Matsumoto & Ekman, 2009, p. 69). These systems are hard-wired, universal and very stable, leaving little or no room for modification. The system basically functions by initially scanning the environment and then converting what is perceived into schemas which are later matched to an existing emotion schema database contained within the system. If there is a match, the corresponding emotion may be triggered, with corresponding "expressive behavior, physiology, cognitions, and subjective experience" (Matsumoto & Ekman, 2009, p. 70). If not, no emotion is elicited and one continues to scan the environment. This approach to emotion is highly functional but does not allow for any variance in the emotional response, so whereas componential theories of emotion suggest it is the appraisal process that drives the "response patterning of physiological reactions, motor expression, and action preparation", basic emotion theories suggest that it is the eliciting events that produces "emotion-specific response patterns such as prototypical facial expressions, physiological reactions, and action tendencies" (Grandjean, Sander & Scherer, 2008). Finally, the term *basic* refers to the small number of emotions, or rather "affect programs" as they have been called in the past, that the theory allows. Usually these are anger, fear, joy, sadness, disgust and/or surprise. All other emotions are variations or combinations of these basic emotion families or affect programs, but the criteria for these classifications remains unclear to this day. Recently, Ekman (2004), as a representative of basic emotion theoretician, suggested that emotions can be triggered by an autoappraiser system, meaning that the emotions are triggered by a series of automatic evaluations or appraisals, but the author did not specify the rules of integration or the granularity of a such system.

The third main type of theory of emotion is defined by "dimensional emotion models" or "dimensional feeling models". As the name might imply, these models strongly base themselves on the subjective feeling component of emotion in describing and differentiating the types of emotional responses one can have, but without specifying any mechanism by which these emotions might be aroused in the first place. The term *dimensional* refers to the origin of the model(s), that is "Wundt's (1905) proposal that feelings (which he distinguished from emotions) can be described by the dimensions of pleasantness–unpleasantness, excitement–inhibition, and tension–relaxation, and on Osgood's work on the dimensions of affective meaning [arousal, valence, and potency]" (Grandjean, Sander & Scherer, 2008, p.485). Though more recently, it is only the dimensions of valence (ranging from pleasant to unpleasant) and arousal (ranging from active to passive) that are studied. According to one of the main current proponents of the model, "Russell (2003), ''core affect,'' presumably the primary emotional reaction, consists exclusively of its position in this bidimensional space, which is only later differentiated and enriched by cognitive and linguistic processing" (Grandjean, Sander & Scherer, 2008, p.485). Though this approach to emotion might prove extremely useful for the measurement of subjective feeling and the subjective experience of mood and emotion in general, it lacks explanations as to the eliciting factors of emotion, their differentiation (there is no explanation for the differentiation between "anger" and "fear" for example which are both high in arousal and low in valence) and can therefore be of little use in predicting an individual's reaction since any cognitive appraisal type of process occurs post-hoc and there is no mechanism proposed by the model.

2.1.3 Summary

In sum, while there are several approaches to the study and definition of emotion, mainly *basic emotion theories*, *bi-dimensional theories* and *appraisal theories*, it seems that authors are converging more and more towards the idea of emotion as a multicomponential process involving all major subsystems of an organism regardless of their approach. In regard to *appraisal theories of emotion*, appraisal is not just a component of the emotion response but the eliciting factor driving the pattern response in the organism. How does this relate to musically induced emotions? While it has been reliably demonstrated that the appraisal mechanism can elicit and differentiate emotions in the case of everyday events and stimuli regarded as highly relevant for the organism, music does not appear to warrant an adaptive response quite in the same way as it is not highly relevant in the same sense and does not appear to have important consequences, particularly immediate survival, for the listener. The idea is to dissociate the emotions related to utilitarian emotions and aesthetic emotions even if they can share some common eliciting mechanisms. How does one *appraise* music exactly? Other mechanisms able to account for and predict the responses individuals have to music must be explored. Some of these will be suggested and discussed in the next section.

2.2 Part 2: The Brain bases for communication and emotion sharing in music, language and action.

2.2.1 The Mirror Neuron System

The discovery of mirror neurons, first in monkey (di Pellegrino et al. 1992, Gallese et al. 1996, Rizzolatti et al. 1996) and then in humans (Fadiga et al. 1995) has shed new light on the brain mechanisms possibly involved in action understanding. Although the premotor mirror neurons don't seem directly involved in emotional sharing (others brain regions such as insula and amygdala are probably much more relevant for this), they for the building blocks through which others' actions are perceived and understood. The mirror neuron system (Rizzolatti, Fadiga, Gallese, & Fogassi, 1996) is formed by premotor neurons, originally discovered in macaque monkey, discharging both when the animal acts and when it sees similar actions performed by other individuals. A system, similar to that found in monkeys, has been indirectly shown to exist also in humans by transcranial magnetic stimulation studies of the motor cortex during action observation (Fadiga, Fogassi, Pavesi, & Rizzolatti, 1995). Further investigations have shown that the mirror system can be activated not only by visually perceived actions but also by listening to action-related sounds (Kohler et al., 2002; Ricciardi et al. 2009) and, in humans, by speech listening (Fadiga, Craighero, Buccino, & Rizzolatti, 2002). In addition to these electrophysiological data, in humans, a number of brain imaging studies point all to a network of brain areas responsible for the visuo-motor transformation mechanism underlying action recognition. It is plausible that the motor resonant system formed by mirror neurons is involved in someone else’s action understanding and, at least in humans, imitation.

2.2.2 Mirror neurons in humans for hand actions and speech

The discovery of tri-modal (motor, visual and auditory) mirror neurons in the monkey ventral premotor cortex has encouraged studies of the auditory properties of the human mirror-neuron system (Kohler, E. et al. 2002). This putative mechanism is thought to map the acoustic representation of actions into the motor plans necessary to produce those actions. Action-related sounds were found to activate the inferior frontal gyrus in addition to the superior temporal gyrus (Pizzamiglio et al. 2005). Sounds executed by the hand or the mouth activate premotor areas in a somatotopical manner, in humans (Gazzola, Aziz-Zadeh, Keysers. 2006). Lewis and colleagues found that tool sounds preferentially activated a cortical mirror-like network (Lewis et al. 2005). This network directly overlapped with motor-related cortices activated when participants pantomimed tool manipulations. Warren and colleagues demonstrate that listening to nonverbal vocalizations can automatically engage the preparation of responsive orofacial gestures, an effect that is greater for positive-valence and high-arousal emotions (Warren et al. 2006). Summing up, sound associated to movement such as action-sounds, tool-sounds or vocalizations, activate the same structures necessary for the production of similar sounds. Listening to these sounds enable the study of simple auditory-motor interactions.

2.2.3 Brain centers involved in music processing

Other studies have investigated the brain areas activated in more complex action-related sounds such as music. Musicians are a particular class of experts who master the ability to map sounds onto movements. Experts already proved to be an interesting model of over-learned perceptuo-motor associations. Dancers and sport players observing and evaluating actions, within their field of expertise, indeed demonstrated a higher motor awareness (Calvo-Merino et al., 2006; Aglioti et al., 2008). Experts, by definition, are subjects who decide to train a particular skill extensively, just like professional musicians do for hours a week.

Musicians, in fact, revealed to be subjects of great interest in the study of how specific training can enlarge somatosensory representation of digits used in practicing the instrument (Elbert et al., 1995). Similarly, the motor system undergoes important plastic changes with specific practice in musicians (Pascual-Leone et al., 1995), and auditory representations were found to be enlarged specifically for musical tones (Pantev et al., 1998). Musicians have also been used to investigate both long-term structural and short-term functional brain plastic changes (Schlaug et al., 1995; Rosenkrantz et al., 2007). Sensorimotor plasticity in musicians is the result of repeated co-occurrences of specific actions and the associated sensory effects (Munte et al., 2002).

A growing number of studies have focused on the mechanisms for the integration of audio-motor information (Zatorre, Chen, Penhune. 2007). A number of neuroimaging researches have recently addressed this issue from different perspectives. On one hand, it has been found that motor and premotor activities could be elicited, in experts, by passive listening to known melodies. For instance the activity of motor centers of expert pianists was enhanced while they were listening to piano pieces (Haueisen, Knosche. 2001). Furthermore, several fMRI studies looked for common activations between perception and production of a musical piece (Bangert, Altenmuller. 2003). These studies confirmed the existence of a complex brain network including motor, premotor and supplementary motor areas, the inferior parietal lobule and the superior temporal gyrus. On the other hand, others studies have focused their attention on the role of training in non-experts. An EEG study showed an increased sensorimotor activity in naïve subjects after a short musical training, both during observation of muted piano movements and passive listening (Bangert et al., 2006). Interestingly, a TMS study further demonstrated that already after 30 minutes of practice, the passive listening of a trained piece increased the facilitation of listeners’ primary motor cortex (D’Ausilio et al., 2006). On the same line, it has been shown that, in non-musicians, premotor activity was specifically increased by passive listening to a trained piece, but not to a different combination of the same notes (Lahav, Saltzman, Schlaug. 2007). Additionally, musical imagery research is a particularly interesting domain of study, dealing with the ability of re-enacting musical experience - being it motor, auditory or both. During these tasks, musicians activate a network of areas similar to that outlined for the passive listening and performance of a musical excerpt (Zatorre, Halpern, 2005; Langheim et al. 2002). Therefore, it is possible to delineate a network of brain areas shared between listening, producing and imaging musical excerpt. This network include the STG, the IPL and motor/premotor regions, and is active in both experienced musicians, and also in naïve subjects after a proper training. Such brain circuit shows close similarities with that found in language studies (Gernsbacher, Kaschak 2003) and, intriguingly, often reported in action execution/observation research. Thus, it is likely that listening to musical excerpts (after proper motor training) activates motor representations required for the actual production of those melodies, with a mirror-like mechanism. In fact, musical excerpts might be considered as actions whose motor representation are preferentially triggered by auditory stimuli (D’Ausilio 2007).

2.2.4 Do language music and actions share similar syntax?

Another interesting parallel between language and music is the similar intrinsic complexity of musical and language structures. In fact, there are more homologies between these two domains than might be expected on the basis of dominant theories of musical and linguistic cognition — from sensory mechanisms that encode sound structure to abstract processes involved in integrating words or musical tones into syntactic structures (Patel, 2003). In an elegant study, Maess and colleagues located in the bilateral inferior frontal gyrus the seat of the musical syntax (Maess et al., 2001). Indeed, in several occasions, the predictability of harmonics and the rules underlying music organization has been compared to language syntax (Patel, 2003). By inserting unexpected harmonics, Maess and co-workers created a sort of musical 'syntactic' violation (Maess et al., 2001). Using magnetoencephalography (MEG) they studied the neuronal counterpart of hearing harmonic incongruity and they found an early right anterior negativity (ERAN) usually associated with harmonic violations (Koelsch et al. 2000). A similar fMRI study revealed that the human brain network involved in processing musical information has strict similarities with that for processing language. Broca's and Wernicke's areas, the superior temporal sulcus, Heschl's gyrus, both *plana polaris* and *temporalis*, as well as the anterior superior insular cortices were all found activated when listening to unexpected musical chords (Koelsch et al. 2002). Tillmann and colleagues investigated the neural correlates of processing harmonically related and unrelated musical sounds in classical priming paradigm (Tillmann, Janata, Bharucha, 2003). These behavioral studies showed that the processing of a musical targets is faster and more accurate when it is harmonically related to the preceding stimuli. Moreover, blood oxygen level-dependent (BOLD) signal measured by fMRI in the IFG was stronger for unrelated than for related targets. This results has been interpreted as a proof that the inferior frontal cortex in involved the processing of syntactic relations and in favor of its role in processing and integrating sequential information over time.

Summing up, several EEG/MEG studies have found the emergence of an early right anterior negativity (ERAN, around 200ms) when subjects were presented with structurally irregular chords (Koelsch et al. 2000; Koelsch et al. 2002). More interestingly the ERAN is very similar to another deviance-related negativity such as the early left anterior negativity (ELAN) which reflects the processing of syntactic structures in language (Friederici 2002). The generator of the ERAN was localized in BA44 (Koelsch et al. 2000), in accordance with other fMRI studies using similar paradigms or using paradigms studying the processing of harmonic, melodic or rhythmic structures (Koelsch, 2006).

2.2.5 The Broca's area as supramodal syntactic center

The overall picture about Broca's area in cognition suggests a pivotal role in critical domains such as language, action and music. Broca's area involvement in language production has a long history, solidly built upon 150 years of lesion studies and corroborated by modern neuroimaging and neurophysiological techniques (Gernsbacher & Kaschak, 2003). Recently though, it's role has been extended also to receptive functions, in the context of integrated brain network models (Pulvermuller, 2005). Although the functional connection and relation between productive and receptive mechanisms is an old scientific question (James, 1890), only in recent years a renewed interest fostered substantial scientific advancements. This new interest is partly due to neurophysiological studies on the monkey (Rizzolatti & Craighero, 2004). These studies, describing neuronal mechanisms for matching executed and observed actions, motivated numerous neuroimaging and neurophysiological studies in search for similar mechanisms in humans. Broca's area was found to be at the center of a brain network for the encoding of action goals, either observed or executed (Rizzolatti & Craighero, 2004). Moreover, action representation in Broca's area was also demonstrated to be triggered by its acoustical counterpart, both in the monkey and humans (Kohler et al, 2002; Gazzola et al, 2006; Lahav et al, 2007) and finally Broca's area was found implicated in the encoding of musical syntax much the way it does encode language structures (Koelsch, S. 2006).

Lesion studies have demonstrated that lesions situated in the *pars opercularis* and *triangularis* of the left inferior frontal gyrus lead to an impairment in gesture comprehension (Pazzaglia et al, 2008). These results are also in line with those by Tranel and colleagues (Tranel et al, 2003), demonstrating that left frontal brain damaged patients have difficulties in understanding action details when presented with cards depicting various actions. The basic idea that Broca’s area is involved in action representation (in broad terms) is also supported by the reported deficit of these patients in specifically representing action verbs (Gainotti et al, 1995). Moreover, we tested patients with a lesion centered in Broca’s area with an action sequencing task and our results suggest the intriguing possibility that Broca’s area could represent action’s syntactic rules rather than the basic motor program to execute them (Fazio et al, 2009).

Actions are denoted by a relevant behavioral goal that, in order to be achieved, requires the composition of simpler motor acts. Single motor acts do not necessarily posses a goal that motivates their execution. On the other hand, the same motor act might be part of very different actions, associated to different goals. Typically a goal-directed action might be “to drink” or “to displace”, and reaching for a glass might be associated to both of these goals without evident differences in kinematics. At the same time, the “drinking” action-prototype might be composed by several acts such as “reach” for the glass, “bring” it to mouth and “swallow”, but can also be satisfied by using a complete different set of acts such as the case of drinking from a public fountain. Actions and motor acts are also composed of simpler units representing the spatio-temporal sequence of muscle activations (Grafton & Hamilton, 2007).These action hierarchies resemble the complex structures shown in other domains such as music and language and more interestingly, the manipulation of these structures is associated with the activation of premotor and Broca’s areas, regardless of the domain of study.

Hence, we propose that Broca’s area might be the center of a brain network encoding hierarchical structures regardless of its use in action, language or music. This hypothesis is also in agreement with recent studies demonstrating that patients with lesions of Broca’s area are also impaired in learning the hierarchical/syntactic structure, but not the temporal one, of sequential tasks (Dominey et al, 2003; Sirigu et al, 1998). The shared feature of music, language and action may therefore be their use of hierarchical/syntactical structures and these results support the idea of a supramodal role for BA44. A recent study using event related fMRI succeeded in disentangling hierarchical processes from temporally nested elements. The authors reported that Broca’s area, and its right homologue, control selection and nesting of action segments, integrated in hierarchical behavioral plans, regardless of their temporal structure (Koechlin & Jubault, 2006). In fact, when comparing the processing of hierarchical dependencies to adjacent dependencies, subjects show significantly higher activations in Broca's area and in the adjacent ventral premotor cortex (Bahlmann et al. 2008). These results indicate that Broca's area is part of a neural circuit that is responsible for the processing of hierarchical structures in an artificial grammar.

Although the proposal that Broca's area could encode a supramodal syntax might seem intriguing, several questions remain still to be addressed. On a theoretical level remains to be defined how, and to what extent, syntactic structures in these domains (action, language and music) do share similar mechanisms and how they interact. Anatomically speaking then, we also need more data on the degree of overlap (and/or segregation) between activities associated to syntax encoding in all these different domains. Finally, it remains somehow obscure the degree of innateness and plasticity of such a syntactical representation. If on one hand we might think that syntactical complexity grows with experience, on the other hand it seems that a basic sensitivity to a set of grammatical rules is present already at birth, regardless of the linguistic experience and exposure. Such claim is supported by a recent study on newborns showing a preference for sequence of stimuli with a structural regularity (ABB as opposed to ABC) and that processing of these redundant sequences elicit activities in the left inferior frontal gyrus (Gervain et al, 2008) One possible, and reconciling, interpretation is that the organization of sensory and motor events in terms of hierarchical structures might be a necessary step to allow comprehension and encoding of experience but that, at the same time, the brain is ready for syntax at birth because of its innate capability to deal with (and statistically appreciate) stimuli regularities.

2.2.6 References

Aglioti, S. M. et al. 2008. Action anticipation and motor resonance in elite basketball players. Nat. Neurosci. doi:10.1038/nn.2182.

Bahlmann, J., R. I. Schubotz & A. D. Friederici. 2008. Hierarchical artificial grammar processing engages Broca's area. Neuroimage 42: 525-534.

Bangert, M. & E. O. Altenmuller. 2003. Mapping perception to action in piano practice: a longitudinal DC-EEG study. BMC Neurosci. 4: 26.

Bangert, M. et al. 2006. Shared networks for auditory and motor processing in professional pianists: evidence from fMRI conjunction. Neuroimage 30: 917-926.

Calvo-Merino, B. et al. 2006. Seeing or doing? Influence of visual and motor familiarity in action observation. Curr. Biol. 16: 1905-1910.

D'Ausilio, A. et al. 2006. Cross-modal plasticity of the motor cortex while listening to a rehearsed musical piece. Eur. J. Neurosci. 24: 955-958.

D'Ausilio, A. 2007. The role of the mirror system in mapping complex sounds into actions. J. Neurosci. 27: 5847-5848.

di Pellegrino G, Fadiga L, Fogassi L, Gallese V, Rizzolatti G. (1992). Understanding motor events: a neurophysiological study. Exp Brain Res, 91(1):176-80.

Dominey, P. F. et al. 2003. Neurological basis of language and sequential cognition: evidence from simulation, aphasia, and ERP studies. Brain Lang. 86: 207-225.

Elbert, T. et al. 1995. Increased cortical representation of the fingers of the left hand in string players. Science 270: 305-307.

Fadiga, L. et al. 1995. Motor facilitation during action observation: a magnetic stimulation study. J. Neurophysiol. 73: 2608-2611.

Fadiga L, Craighero L, Buccino G, Rizzolatti G. (2002). Speech listening specifically modulates the excitability of tongue muscles: a TMS study. Eur J Neurosci. 15(2):399-402.

Fazio P, Cantagallo A, Craighero L, D'Ausilio A, Roy AC, Pozzo T, Calzolari F, Granieri E, Fadiga L. (2009). Encoding of human action in Broca's area. Brain. 132(Pt 7):1980-8.

Friederici, A. D. 2002. Towards a neural basis of auditory sentence processing. Trends Cogn. Sci. 6: 78-84.

Gallese V, Fadiga L, Fogassi L, Rizzolatti G. (1996). Action recognition in the premotor cortex. Brain. 119 (Pt 2):593-609.

Gazzola, V., L. Aziz-Zadeh & C. Keysers. 2006.Empathy and the somatotopic auditory mirror system in humans. Curr. Biol. 16: 1824-1829.

Gernsbacher, M. A. & M. P. Kaschak.. 2003. Neuroimaging studies of language production and comprehension. Annu. Rev. Psychol. 54: 91-114.

Gervain, J. et al. 2008. The neonate brain detects speech structure. Proc. Natl. Acad. Sci. USA 105: 14222-14227.

Grafton, S. T. & A. F. Hamilton. 2007. Evidence for a distributed hierarchy of action representation in the brain. Hum. Mov. Sci. 26: 590-616.

Haueisen, J. & T. R. Knosche. 2001. Involuntary motor activity in pianists evoked by music perception. J. Cogn. Neurosci. 13: 786-792.

Koechlin, E. & T. Jubault. 2006. Broca's area and the hierarchical organization of human behavior. Neuron 50: 963-974.

Koelsch, S. 2006. Significance of Broca's area and ventral premotor cortex for music-syntactic processing. Cortex 42: 518-520.

Koelsch, S. et al. 2000. Brain indices of music processing: "nonmusicians" are musical. J. Cogn. Neurosci. 12: 520-541.

Koelsch, S. et al. 2002. Bach speaks: a cortical "language-network" serves the processing of music. Neuroimage 17: 956-966.

Kohler, E. et al. 2002. Hearing sounds, understanding actions: action representation in mirror neurons. Science 297: 846-848.

Lahav, A., E. Saltzman & G. Schlaug. 2007. Action representation of sound: audiomotor recognition network while listening to newly acquired actions. J. Neurosci. 27: 308-314.

Langheim, F. J. et al. 2002. Cortical systems associated with covert music rehearsal. Neuroimage 16: 901-908.

Lewis, J. W. et al. 2005. Distinct cortical pathways for processing tool versus animal sounds. J. Neurosci. 25: 5148-5158.

Maess, B. et al. 2001. Musical syntax is processed in Broca's area: an MEG study. Nat. Neurosci. 4: 540-545.

Munte, T. F. et al. 2002. The musician's brain as a model of neuroplasticity. Nat. Rev. Neurosci. 3: 473-478.

Pantev, C. et al. 1998. Increased auditory cortical representation in musicians. Nature 392: 811-814.

Pascual-Leone, A. et al. 1995. Modulation of muscle responses evoked by transcranial magnetic stimulation during the acquisition of new fine motor skills. J. Neurophysiol. 74: 1037-1045.

Patel, A. D. 2003. Language, music, syntax and the brain. Nat. Neurosci. 6: 674-681.

Pazzaglia, M. et al. 2008. Neural underpinnings of gesture discrimination in patients with limb apraxia. J. Neurosci. 28: 3030-3041.

Pizzamiglio, L. et al. 2005. Separate neural systems for processing action- or non-action-related sounds. Neuroimage 24: 852-861.

Ricciardi E, Bonino D, Sani L, Vecchi T, Guazzelli M, Haxby JV, Fadiga L, Pietrini P. (2009). Do we really need vision? How blind people "see" the actions of others. J Neurosci. 29(31):9719-24.

Rizzolatti G, Fadiga L, Gallese V, Fogassi L. (1996). Premotor cortex and the recognition of motor actions. Brain Res Cogn Brain Res. 3(2):131-41.

Rizzolatti, G. & L. Craighero. 2004. The mirror-neuron system. Annu. Rev. Neurosci. 27: 169-192.

Rosenkranz, K. et al. 2007. Motorcortical excitability and synaptic plasticity is enhanced in professional musicians. J. Neurosci. 27: 5200-5206.

Schlaug, G. et al. 1995. In vivo evidence of structural brain asymmetry in musicians. Science 267: 699-701.

Sirigu, A. et al. 1998. Distinct frontal regions for processing sentence syntax and story grammar. Cortex 34: 771-778.

Tillmann, B., P. Janata & J. J. Bharucha. 2003. Activation of the inferior frontal cortex in musical priming. Brain Res. Cogn. Brain Res. 16: 145-161.

Tranel, D. et al. 2003. Neural correlate of conceptual knowledge for actions. Cogn. Neuropsych. 20: 409-432.

Warren, J. E. et al. 2006. Positive emotions preferentially engage an auditory-motor "mirror" system. J. Neurosci. 26: 13067-13075.

Zatorre, R. J., J. L. Chen & V. B. Penhune. 2007. When the brain plays music: auditory-motor interactions in music perception and production. Nat. Rev. Neurosci. 8: 547-558.

Zatorre, R. J. & A. R. Halpern. 2005. Mental concerts: musical imagery and auditory cortex. Neuron 47: 9-12.

2.3 Part 3: Conceptual and methodological approaches to the study of musical emotions

In the following section we will clarify a few of the terms to be used in the project and the main models of musical emotion on which we will base our work.

2.3.1 A few distinctions

2.3.1.1 Different processes: Recognition vs. Induction

First, we propose to define terms used to distinguish between different processes that can take place while listening to music.

Firstly, and quite representative of the cognitivist view on music and emotion (Krumhansl 1997), is the process of recognizing the emotions expressed or represented *in* the music. This is typically what is referred to as perception of emotion in music, e.g. Scherer and Zentner's "perception" (2001). Secondly, there is the process of experiencing emotions as a result of listening to music. This can be referred to as induction of emotion *by* music or induction of emotion *in the listener*, see "aesthetic emotions" in Scherer and Zentner (2008), and is typical of an *emotivist* view of emotion.

Recognition	Recognition of emotion in music	Music represents or expresses emotion	"perception" in the Scherer & Zentner sense
Induction	Induction of emotion by music	Listener experiences an emotional episode	"aesthetic emotion" in the Scherer & Zentner sense

Both processes involve "perception" of music in the sense defined below.

2.3.1.2 Different levels: Production vs. perception

As this work takes place within the framework of the SIEMPRE project, where research is focused on ensemble musical performance and audience experience, a distinction will be made between two levels: that of perception and production of music.

In psychology, *perception* is usually understood as the "sensory experience that has been interpreted with reference to its presumed external stimulus object or event". What is meant by *perception* is simply the experience of music listening, including the perception of intrinsic properties of the music in the most basic sense as well as the perception of emotional expression in the music. *Production* on the other hand consists of all aspects pertaining to musical performance, and therefore must include perception as well, especially in the context of ensemble performance.

Perception	Focus on the listener/audience	Music listening in the most basic sense. Whether emotions are recognized in the music or not, experienced in response to the music or not. Not to be confused with « perception » in the Scherer and Zentner sense.
Production	Focus on the musician(s)	Music performance by a musician(s). This of course requires music perception by necessity.

A complete separation between these levels would be artificial, since music production obviously requires music perception, but the distinction is useful to us in the sense that they will put more or less emphasis on either listeners or musicians.

2.3.2 The BRECVEM model

In 2008, Juslin and Västfjäll claimed that if there had been a certain stagnation in the field of music and emotion, it was mainly due to the fact that the processes that could induce musical emotions, i.e. emotions induced by music, had been ignored, abandoned, or insufficiently explored. They therefore suggested six psychological processes[1], in addition to cognitive appraisal, that could account for the induction of emotions by music, namely: (I) brain stem reflexes, (II) evaluative conditioning, (III) emotional contagion, (IV) visual imagery, (V) episodic memory, and (VI) musical expectancy. After testing some of the mechanisms and conducting exploratory studies in the general population, the appraisal mechanism was dropped and following the advice of several authors in open peer commentaries, the *rhythmic entrainmen*t mechanism was added (Juslin, Liljeström, Västfjäll, & Lundqvist, 2010). Thus, the updated model (BRECVEM) now contains the seven mechanisms:

(I) **B**rain stem reflexes,

(II) **R**hythmic entrainment,

(III) **E**valuative conditioning,

(IV) emotional **C**ontagion,

(V) **V**isual imagery,

(VI) **E**pisodic memory, and

(VII) **M**usical expectancy.

We will now elaborate on what is meant by each mechanism according to the authors and add to them when necessary.

2.3.2.1 -I- Brain stem reflexes.

In general terms, "these include reflexes regulated at the level of the brain stem, such as pupillary, pharyngeal, and cough reflexes, and the control of respiration; their absence is one criterion of brain death" (Dorland's Medical Dictionary for Health Consumers, 2007). But as a potential psychological process, it "refers to a process whereby an emotion is induced by music because one or more fundamental acoustical characteristics of the music are taken by the brain stem to signal a potentially important and urgent event" (Juslin & Västfjäll, 2008, p. 564). We would like to highlight here that the notion of expectations is very important at this level, for example in the context of the perception of consonance and dissonance in music. The importance of other brain regions, at low level of processing, should be also highlighted given the fact that, for example, the grey nuclei seem very important in rhythmic perception and representation. We will also discuss the importance of the rhythmic perception in the context of mirror neuron systems and the involvement of premotor or motor brain areas in the context of music listening.

2.3.2.2 -II- Rhythmic entrainment.

According to Clayton, Sager and Will (2005) "[entrainment], broadly defined, is a phenomenon in which two or more independent rhythmic processes synchronize with each other".

However, within the context of the BRECVEM model:

"Rhythmic entrainment refers to a process whereby an emotion induced by a piece of music because the powerful, external rhythm of the music interacts with an internal body rhythm of the listener such as heart rate, such that the latter rhythm adjusts towards and eventually 'locks in' to a common periodicity. The adjusted heart rate may then spread to other components of emotion (e.g. feeling) through proprioceptive feedback, thereby producing increased arousal in the listener. There are two components required in rhythmic entrainment (see Clayton et al, 2005). First, there must be (at least) two autonomous rhythmic processes or *oscillators* (i.e. the beats), autonomy means that they should both be able to oscillate at a given frequency, even if they are separated-which excludes resonance from the notion of entrainment. Second, the two oscillators must interact in some way (perceptual relationships of one specific auditory stream). Entrainment is found throughout nature. It occurs in some way or another in all animal species, and humans appear to have an innate propensity to entrain (Clayton et al, 2005). The cooperative and oscillatory activities of brain neurons may form part of the basis for timing in sensory-motor coordination and meter perception (Jones, 2009). Clayton (2009) has proposed that entrainment is particularly noticeable in activities where rhythmic coordination will make physical work more efficient. Entrainment has not been systematically studied with respect to musical emotion. Kneutgen (1970) found that when soothing lullabies were played for infants, their breathing rhythms became synchronized with the musical rhythm. Further, Landreth and Landreth (1974) found changes in heart rate to be directly related to changes in tempo. Harrer and Harrer (1977) reported that music listeners tended to synchronize either their heart rate or their respiration to the music, and that one could 'drive' the pulse with appropriate music. The entrainment-inducing properties of music that produce affect presumably depend on the music having a marked pulse-and preferably one that is relatively close to the 'natural' heart rate or respiration of the listener. Oscillators do not synchronize instantaneously, and the period takes longer to adjust than the phase (Clayton et al, 2005, p.9, p.15). This means that entrainment is a slower induction process.

2.3.2.3 -III- Evaluative conditioning.

"This refers to a process whereby an emotion is induced by a piece of music simply because this stimulus has been paired repeatedly with other positive or negative stimuli." (Juslin & Västfjäll, 2008, p. 564). This process may occur with or without the awareness of the participant and could account for the phenomenon of induction of emotions by music for no apparent reason and of positive reactions to pieces judged to be of poor quality. The observation that music occurs when music listening is not the primary activity seems to support these claims. Finally, evaluative conditioning "seems to depend on unconscious, unintentional, and effortless processes […] which involve subcortical brain regions such as the amygdala and the cerebellum" (Juslin & Västfjäll, 2008, p. 565). In this process, the participant will not always be able to verbalize the phenomenon.

2.3.2.4 -IV- Emotional contagion.

According to Juslin and Västfjäll, emotional contagion "refers to a process whereby an emotion is induced by a piece of music because the listener perceives the emotional expression of the music, and then "mimics" this expression internally, which by means of either peripheral feedback from muscles, or a more direct activation of the relevant emotional representations in the brain, leads to an induction of the same emotion." And "[evidence] that music with a specific emotional expression can give rise to the same emotion in the listener" would seem to lend support to their views. "Because music often features expressive acoustical patterns similar to those that occur in emotional speech […], it has been argued that we become aroused by the voice-like aspects of music via a process in which a neural mechanism responds quickly and automatically to certain stimulus features, which leads us to mimic the perceived emotion internally." (Juslin & Västfjäll, 2008, p. 566). Recently, for example, Curtis and Bharucha (2010) have empirically demonstrated that the minor third in speech communicates sadness as already mentioned in music by several authors. This phenomenon of emotional contagion might partly explain rhythmic entrainment effects produced by tempo for example.

2.3.2.5 -V- Visual imagery.

"This refers to a process whereby an emotion is induced in a listener because he or she conjures up visual images (e.g., of a beautiful landscape) while listening to the music. The emotions experienced are the result of a close interaction between the music and the images. […] Visual imagery is usually defined as an experience that resembles perceptual experience, but that occurs in the absence of relevant sensory stimuli." (Juslin & Västfjäll, 2008, p. 566). Since mental images can be internal triggers of emotion, this process could be interpreted as being a case of indirect induction of emotion through music which acts more as a stimulant, but Juslin and Västfjäll rather suggest that it is the interaction between the visual images and the music perception that produces an emotion in the listener. Perhaps "listeners […] conceptualize the musical structure through a metaphorical nonverbal mapping between the music and so-called image-schemata grounded in bodily experience (Bonde 2006; Lakoff & Johnson 1980); for example, hearing melodic movement as 'upward'." (Juslin & Västfjäll, 2008, p. 566). The metaphor phenomenon seems important in such visual imagery.

This process would also allow the listener to have a great deal of control over his or her emotional experience of the music by feeding it or modifying it at will.

2.3.2.6 -VI- Episodic memory.

Episodic memory is defined as all personally experienced events. "This refers to a process whereby an emotion is induced in a listener because the music evokes a memory of a particular event in the listener's life.[…] When the memory is evoked, so also is the emotion associated with the memory (e.g., Baumgartner 1992). Such emotions can be rather intense, perhaps because the physiological reaction patterns to the original events are stored in memory along with the experiential content, as proposed by Lang (1979)." (Juslin & Västfjäll, 2008, p. 567). The main idea here is that a specific piece of music is able to reactivate a long term episodic memory representation automatically as it is the case in the "Madeleine de Proust" for odors.

As we understand it, this process could be distinguished from *evaluative conditioning* by being closer to the listener's awareness.

2.3.2.7 -VII- Musical expectancy.

"[Refers] to a process whereby an emotion is induced in a listener because a specific feature of the music violates, delays, or confirms the listener's expectations about the continuation of the music. […] [it] refers to those expectancies that involve syntactical relationships between different parts of the musical structure (Narmour 1991; Patel 2003)." (Juslin & Västfjäll 2008 Emotional responses to music, p. 568).

Therefore, this process could be subserved by areas involved in syntactical processing, whether it be of language, meaningful dynamic gesture patterns or music. Juslin and Västfjäll suggest Broca's area as a likely candidate given its role in the integration of syntactical representations. The role of P600 in ERP study comparing processing of variously incongruent sentences and musical chord sequences in terms of syntax in musicians as been shown in several studies (e.g. Patel, Gibson, Ratner, Besson & Holcombe, 1998). Expectancies can be also studied in the context of appraisal theories in which the predictions of following events (here the specific unfolding of a musical piece) are highly predictable in classical music for example (the opposite is true for some contemporary music styles).

2.4 Part 4: Theoretical and Methodological Framework for the study of Musical Emotions in SIEMPRE

In this section we will attempt to draw an integrative view of the theories and methodologies viewed as they pertain to the SIEMPRE project.

2.4.1 The BRECVEM mechanisms in the context of the CPM and GEMS models

Juslin and collaborators (2010) suggested to distinguish the processes by thinking of them as a number of distinct brain functions that have developed gradually and in a specific order during the evolutionary process, from sensations (brain stem reflexes) to syntactical processing (musical expectancy). The table below describes the hypotheses the authors made concerning the nature of each process in that sense:

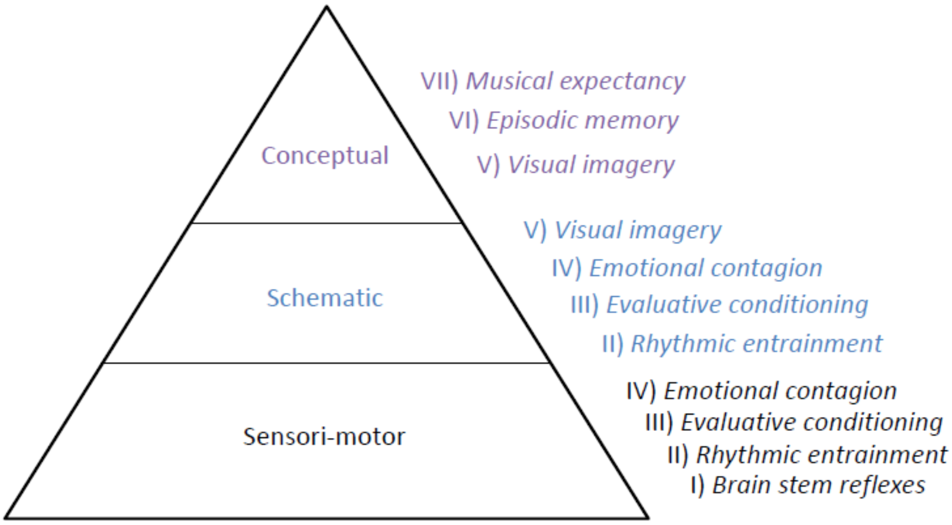
	Characteristic					
Mechanism	Induced affect	Induction speed	Degree of volitional influence	Availability to consciousness	Modularity	Dependence on musical structure
I Brain stem reflex	General arousal, unpleasantness versus pleasantness	High	Low	Low	High	Medium
IIRhythmic entrainment	General arousal, pleasant feelings of communion	Low	Low	Low	High	Medium
IIIEvaluative conditioning	Basic emotions	High	Low	Low	High	Low
IVEmotional contagion	Basic emotions	High	Low	Low	High	Medium
V Visual imagery	All possible emotions	Low	High	High	Low	Medium
VIEpisodic memory	All possible emotions, esp. nostalgia	Low	Medium	High	Low	Low
VIIMusical expectancy	Surprise, awe, pleasure, "thrills", disappointment, hope, anxiety	Low	Low	Medium	Medium	High

Adapted from Juslin et al. (2010, pp. 625-626)

The first column indicates the nature of the induction process and the following ones specify their characteristics.

As can be seen from the table above, the organization of the different levels of processing is somewhat similar to the organization of the levels of processing in the Component Process Model proposed by Scherer (2001), where appraisal occurs both simultaneously and sequentially at different levels: *sensori-motor*, *schematic*, and *conceptual* (Leventhal & Scherer, 1987; Grandjean, Sander, & Scherer, 2008).

Thus, it is not unreasonable to suggest that these processes interact much in the same way at different levels of awareness to induce an emotion and feeling in the listener and it could also explain why reactions to pieces often involve more than one emotion label: “music could induce so-called mixed emotions, because different mechanisms might be activated simultaneously at different levels” (Juslin & Västfjäll, 2008).



In this sense, phenomena such as musical entrainment could be explained by brain stem reflexes and grey nuclei modulations (sensory-motor-like processes) and high level processes, such as appraisal. We suggest that appraisal of the music can also be involved in the elicitation of musical emotions. In addition to *musical expectancy*, some of the CPM checks might be of use here, e.g. relevance appraisal checks: novelty, intrinsic pleasantness (e.g. consonance and dissonance) and goal/need relevance, implications checks: outcome probability.

2.4.2 Summary

We propose that: a) *musical emotions* can be induced by the processes described above in addition to the traditional appraisal processes responsible for the elicitation of *utilitarian emotions*, but that b) they can nevertheless be conceptualized much in the same way as *utilitarian emotions*, i.e. "as dynamic episodes in the life of an organism that involves a process of change in all of its subsystems" (Grandjean, Sander & Scherer, 2008, p. 485). But even if we subscribe to this definition, we still believe that the better way of operationalizing the subcomponent of *subjective feeling* is by the use of the GEMS model proposed by Zentner, Grandjean and Scherer in 2008. Following two studies that allowed them to group the most relevant musical affective terms, they found that these terms fitted best in a 9 dimensional space. According to the authors, the factors corresponded to the following emotional dimensions: Wonder, Transcendence, Tenderness, Nostalgia, Peacefulness, Power, Joyful activation, Tension and Sadness. We propose to use these terms instead of only basic emotion labels or dimensional axes in the SIEMPRE project.

2.4.3 Different foci

The SIEMPRE project will particularly focus on exploring inter- and intra- personal interactions in:

(i) musician-musician scenarios

(ii) conductor-musicians scenarios

(iii) music-listener scenarios

(iv) musician-listener scenarios

Since the overarching strategy of the project is to study the interpersonal processes in live performance and music listening, both production and perception will be studied with the focus depending on the experiment and target population.

Nevertheless, during musician-musician and conductor-musician scenarios we can expect a larger focus on the *production* level, and *perception* for music-listener and musician-listener scenarios (provided the focus is on the audience). We suggest that both *recognition* and *induction* of emotions can be expected though not guaranteed in all four types of scenario, though reliance on self-reported measures will be necessary for anyone studying these processes. We strongly advise anyone doing so to clarify this distinction in their measures by explicitly asking listeners or musicians to state what they either *felt* or *perceived* during a performance.

We focus on four aspects of live performance and listening:

1. *Entrainment*: which creates physical alignment between the individuals;
2. *Emotional contagion*: which creates emotional bonds between them; and
3. *Co-creation*: by which both performers and audience contribute to shaping the overall event.
4. *Leadership*: which explains dynamically the role of the leader musician in a music ensemble, or the role of the conductor of an orchestra section.

2.4.3.1 Entrainment

As it was mentioned earlier (Clayton, Sager & Will, 2005):

> ...the wide range of entrainment phenomena is not based on a single physical process. Rather, the concept of entrainment describes a shared tendency of a wide range of physical and biological systems: namely, the coordination of temporally structured events through interaction (p. 3).

As such, entrainment can be studied at several levels: the behavioral entrainment of several musicians with a conductor and with each other, or the low-level of brainwave synchronizations of neural rhythms with the frequency of a stimulus or a series of stimuli (such as a drum beat at 120bpm = 2Hz). Entrainment can also be studied as an emotion induction mechanism as was discussed earlier (Juslin et al., 2010). Since "[it] is a process that manifests in many ways, some of which involve human agency or cognition" (Clayton, Sager & Will, 2005), we propose that partners specify at what level it is being studied for every experiment.

And since at least a working definition has become necessary, we propose to remain faithful to Clayton's definition of the term which requires that there "be (at least) two autonomous rhythmic processes or oscillators" (Juslin et al., 2010). By *oscillation* we mean the periodic motion of an electrical or mechanical source about an equilibrium position (A Dictionary of physics, 2009) and by autonomy, simply that the oscillators should be able to oscillate even when separated (Juslin et al., 2010). Examples of "endogenous or naturally occurring rhythms within the human body include the heart beat, blood circulation, respiration, locomotion, eyes blinking, secretion of hormones, female menstrual cycles, and many others" including neural oscillations as well (Clayton, Sager & Will, 2005). Outside of the body mechanical oscillators include pendulums and playground swings, and electrical oscillators such as AC current. All of these are able to function separately and have their own source. The condition of autonomy between oscillators will allow us to differentiate the concept from *resonance*, which would be a similar process in the sense that we would also observe a variation of the system when exposed to another (external) periodic force, but this influence is only an increase in the amplitude of the oscillation when "exposed to a periodic force whose frequency is equal or very close to the natural undamped frequency of the system" (The American Heritage Dictionary of the English Language, 2006; A Dictionary of Physics, 2009). So entrainment is more than just *resonance* and it is also more than just *synchrony*, as something that is synchronous is merely something that is "[taking] place at the same time, at the same rate, or with the same period" (A Dictionary of physics, 2009). Indeed, the other component that must be involved in entrainment for it to be entrainment and not just synchrony for example, is the *interaction* of oscillators (Clayton, Sager & Will, 2005). Even though the authors do point out that not all interacting oscillators will entrain as they need to be relatively close in terms of periodicity for the phenomenon to occur and even then strict phase and frequency synchronization is not necessarily observed.

The following are a few definitions of different cases of entrainment that partners might find useful:

- Entrainment to environmental cues (Clayton, Sager, & Will, 2005, p. 5): this kind of entrainment typically depends on environmental cues that are ultimately based on the rotation of the earth, such as day/night cycles, which will affect light/dark and temperature cycles that act as external sources of entrainment for many species at behavioral and physiological levels.

- Asymmetrical entrainment (Clayton, Sager, & Will, 2005, p. 6): this is not so much a type of entrainment as it is a characteristic of cases where one oscillator has no choice but to entrain to an external rhythm, typically environmental cues. It is asymmetrical in the sense that the entrainment can only occur in one direction "the individual cannot influence the entraining rhythm". Here the authors cite circadian rhythms in living organisms as prime examples.

- Mutual entrainment (Clayton, Sager, & Will, 2005, p. 5): this type of entrainment is typically observed between individuals and could thus be understood as a kind of "inter-entrainment". Here the authors cite behavioral examples of crickets chirping and fireflies flashing in synchrony with each other, but one could also suggest the synchrony observed in fish populations.

- Self-entrainment (Clayton, Sager, & Will, 2005, p. 6-7): if the previous type of entrainment could be considered "inter-entrainment" then this kind might be considered "intra-entrainment" by analogy. Here the entraining source is internal and one can observe entrainment at behavioral or physiological levels. In the periods of limb movements in locomotion for example, or when physiological rhythms entrain to each other, like heart rate and respiration rates. More complex types of self-entrainment (at the behavioral level) could include musical self-entrainment such as when one sings and accompanies oneself at the same time.

- Interpersonal entrainment (Clayton, Sager, & Will, 2005, p. 6-7): this concerns entrainment to other individuals, especially at the ultradian level, that is rhythms or cycles that last anywhere between a few seconds or minutes up to a few hours. The authors especially cite the mutual entrainment of organism's "subjective, physiological rhythms" as examples of interpersonal entrainment. This type of entrainment would typically be symmetrical and the authors go on to note that "entrainment may relate phenomenologically to a sense of social belonging, or of one's subjectivity relating to 'something larger': impressions that are frequently linked to musicking, among other activities".

- Frequency or tempo entrainment (Clayton, Sager, & Will, 2005, p. 9): here "the periods of the two oscillators adjust toward a consistent and systematic relationship".

- Phase entrainment, or phase-locking (Clayton, Sager, & Will, 2005, p. 9): here "two processes are phase-locked, focal points occur at the same moment". How is this different from tempo entrainment? In this case it is not just the period of the oscillators that have a consistent relationship, their phase also does. Here the authors point out that "two entrained oscillators have two possible phase-locked states, namely synchrony and anti-synchrony".

What about music? Concerning music, Clayton and collaborators only note that:

> Entrainment to and through music needs to be seen as a particular case of entrainment in social interaction, and its particular qualities explored - as indeed, we need to explore the specific possibilities for entrainment that different musical repertories or performances afford (p. 3).

SIEMPRE background already includes studies in this direction: in (Varni et al 2008; Camurri et al 2010; Varni et al 2010) UNIGE demonstrated how an induced emotion affects changes in mutual, inter-personal and phase entrainment measured in the gesture of a duo of musicians. These proposed approaches individuate novel algorithms and techniques for the real-time measurement of entrainment.

2.4.3.2 Emotional contagion

Emotional contagion has been defined as the "tendency to automatically mimic and synchronize expressions, vocalizations, postures, and movements with those of another person and, consequently, to converge emotionally" (as cited in Hatfield, Cacioppo, & Rapson, 1994). In her writings Hatfield often uses the phrasing "catching someone's emotion" to describe this process which brings to mind the rather passive nature of the phenomenon. Simply put, her view is that it is through the feedback of our mimicking of other's expressions of emotion that we catch other's emotions. We need not be aware of this feedback either. In this context, Lamm and collaborator's "monitoring mechanisms" to verify whose emotion comes from who makes particularly sense (2007). Even though according to Hatfield we only experience, or catch, pale echoes or imitations of people's emotions through this process, i.e. the same emotion but with less intensity.

There have been a number of suggestions to explain the process, essentially centered around the idea that "an action is stored as a sensory feedback representation in our brains [...] This representation is activated when observing somebody perform the action, and will in turn prime the activation of the corresponding motor representation in the observer because of their overlap." (Leiberg & Anders, 2006, p. 421) So when we observe someone in a particular emotional state, a representation of that state will be automatically activated in us, "including its associated autonomic, somatic and motor responses".

Like other emotion contagion researchers, Juslin and collaborators (2010) suggest that the emotional contagion process might be mediated by *mirror neurons*. However, we also suggest that the process might be mediated at an earlier stage by the process of joint attention and eye-gaze to be more specific. Indeed, it has been suggested that eye-contact evolved as a trigger for embodied simulations and that it modulates the presence vs. absence of embodied simulation through activation of the amygdala (Niedenthal, Mermillod, Maringer & Hess, 2010).

De Waal (2008): there are at least two main neural systems mediating empathy = a phylogenetically early emotional contagion system and a more advanced cognitive perspective-taking system. The basic emotional contagion system is thought to support our ability to empathize emotionally and has been linked to the human MNS (mirror neuron system).

2.4.3.3 Joint Musical Performance and Co-creation

Figure 3.1 below summarizes the key stages in the generation of a musical performance. These stages incorporate the features identified by Keller's (2008) analysis of joint musical performance.

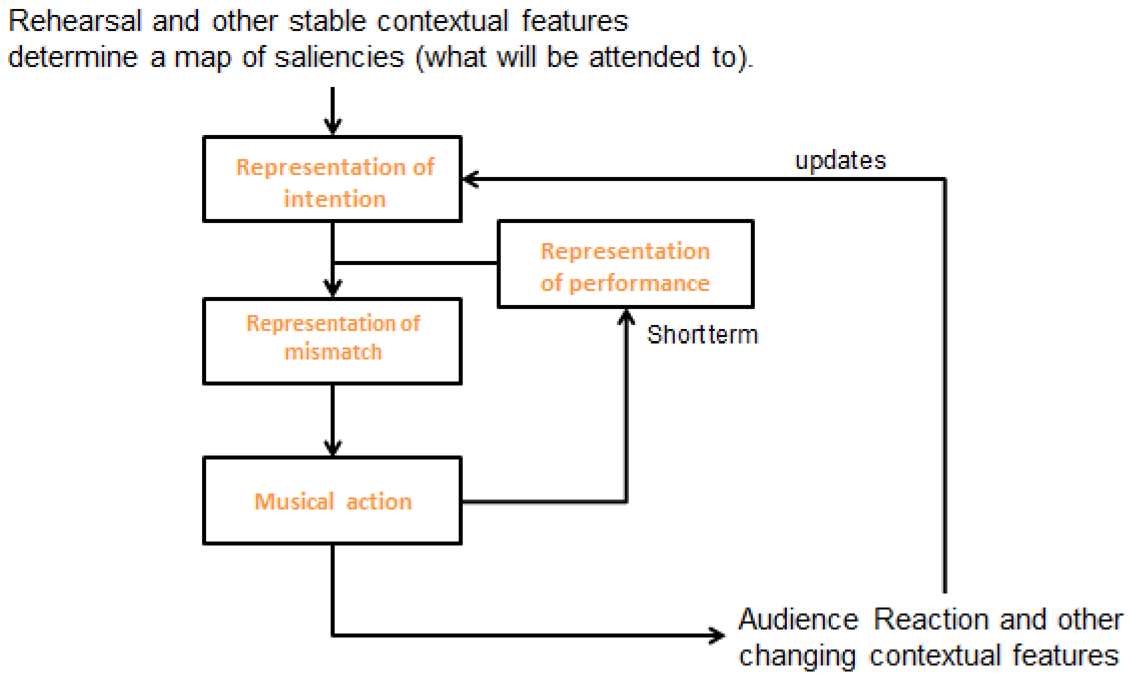

Rehearsal and other stable contextual features determine a map of saliencies (what will be attended to).
Representation of intention
updates
Representation of performance
Representation of mismatch
Shortterm
Musical action
Audience Reaction and other changing contextual features

Fig. 3.1 The generation of a musical performance

i) Stable contextual features and the map of saliences

Musical performances are set within a context of features that motivate the performance in the first place. In particular, the rehearsal process plays a crucial role in solidifying specific performance goals, such as to play in a certain style, or with a certain expressive intent. These goals largely fix what musicians pay attention to during a performance, which strongly influences the phenomena most relevant to their emotional states. As such, they must be recorded as sensitively as possible.

A good way to operationalize an individual's or group's performance goals is with what we call a 'map of saliences'. It is defined as a stable representation of the music, constructed prior to the performance, which specifies what is most important to pay attention to at any given moment. It is in part suggested by a convention-based reading of the score, and in part by the rehearsal process. The main features of this map concern tonal, metrical and expressive hierarchies i.e. the parts of the music that most matter for fixing these properties. Judgements concerning the relative difficulties of various passages will also be included. The group collaboratively constructs a map of saliences for the musical piece as a whole, and the individual musicians derive a map of saliences for their part in particular.

ii) The joint performance

Before making any sound, each individual musician mentally represents the way they want the music to sound. This individual mental representation is largely determined by the map of saliences. However, during the performance itself, the attention of the musicians will also be guided by the various contingencies of performance such as player errors or the audience reaction. In this sense, the intentions of the musicians both as individuals and as a group will be constantly updated during the performance itself. As explored below, this feedback may also be mediated by the leader of the ensemble.

In the immediate context the performance itself, what the musician hears at any given moment will be compared to their musical intention for the following moment. If there is a mismatch between these two states, the musician will act so as to minimize this difference. The mismatch thereby generates the musical action- i.e. direct performance upon the instrument, or actions aimed at influencing the performance of others. These actions then result in musical sounds, the representation of which closes a feedback loop whereby the musical intention for the next moment in time is compared with the current sound. In addition, the sound produced by the musicians also influences short term contextual features such as the reaction of the audience.

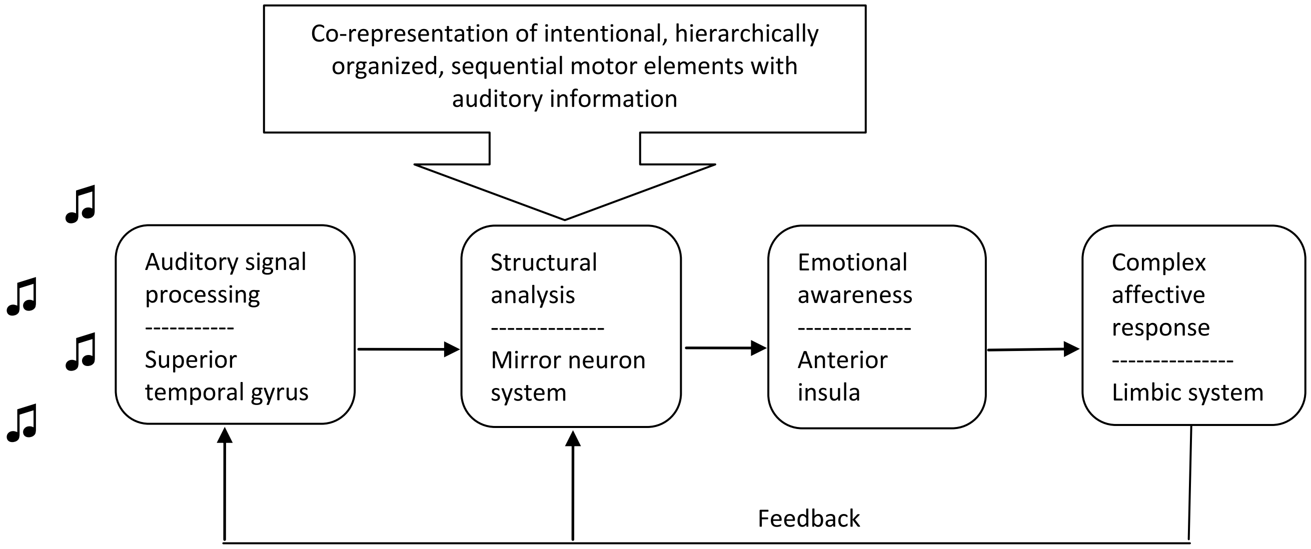
iii) Co-creation and the understanding of the audience

In the SIEMPRE project, co-creation is linked to the notion "*by which both performers and audience contribute to shaping the overall event*". In addition to making direct contributions to the overall quality of sound with say, applause or cheering, it is recognized that the behaviour of the audience influences the behaviour of the musicians. As noted by Koelsch (2010) as well as Cochrane (2009), music leads to increased *social cohesion* of a group, fulfilling the "need to belong", and the motivation to form and maintain interpersonal attachments.

The reaction of the audience is mediated by their understanding of the music. According to the SAME model of affective musical experience, Overy and Molnar-Szakacs (2006, 2009) predict that depending on a listener's level and kind of music training, individuals are able to extract information at different levels of the motor hierarchy: 1) the intention level, 2) the goal level, 3) the kinematic level, and 4) the muscle level. The authors give the example that at one extreme, a professional musician listening to music which they know how to perform is able to access precise information at all levels of the hierarchy, from imagined emotional intentions to specific movements. At the other extreme, a musical novice listening to unfamiliar music from an unknown sound source is not able to access precise information at any level, but may feel beats and rhythms, sub-vocalize, and interpret emotional intention.

2.4.3.4 Mirror neurons as neurophysiological foundation of emotional contagion

Overy & Molnar-Szakacs (2009) specifically point to the importance of mirror neurons (MNS) in the social communication. The brain does not function as an isolated stimulus-response perception-action machine. The brain's functioning is intimately connected with the body and the brain and mind has evolved to interact with other brains and minds trying to predict the behaviors and the feelings of others. The mirror neuron system has the incredible ability to help an individual to understand the meaning and the intention of a communicative signal by evoking a representation of that signal in the perceiver's own brain. By emphasizing the strong power of music, Molnar-Szakacs and Overy (2006) propose that: "[…] *the powerful affective responses that can be provoked by apparently abstract musical sounds are supported by this human mirror neuron system, which may subserve similar computations during the processing of music, action and linguistic information".* In this context, the authors propose the following model:



Model of the possible involvement of the human mirror neuron system in representing meaning and affective responses to music (Molnar-Szakacs & Overy, 2006, p. 237).

2.5 Leadership

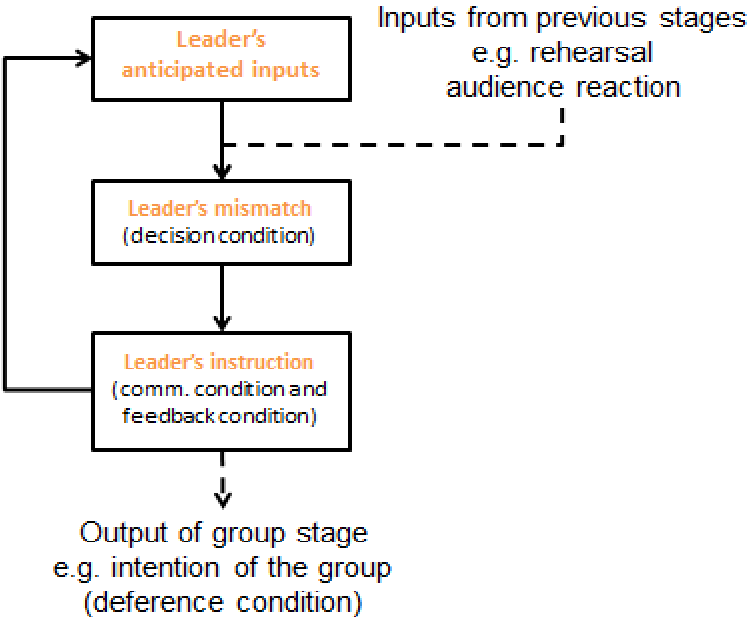
As a general definition of leadership, we propose the following 4 conditions as essential:

1. The leader makes decisions regarding the goal of the group's activities (decision condition).
2. These decisions are communicated to group members via formal conventions and/or informal communications (communication condition).
3. These instructions are then enacted by the group members (deference condition).
4. There are normative consequences for failure or success to follow the leader's instructions (feedback condition).

There are also various optional conditions of leadership such as representing the group as a whole to others, or managing conflict within the group.

2.5.1 Leadership in music ensemble performance

This basic definition of leadership can now be applied to ensemble performance. The group action system depicted in figure 3.1 above can be more or less efficient, coherent, musically effective and so on, and the role of the leader will play an important part in determining this. In particular, the leader acts as a *filter* through which the intentions and behaviours of the group must pass. Referring back to figure 3.1 we may add a leadership node between every stage; in the process of forming intentions, transforming these into actions, representing the performance of the group and determining how this representation adjusts intentions.



As a result of a single figure being responsible for this filtering, the leader also provides a unified focal point both for members of the group and outsiders. He becomes the particularly *salient* feature in the attention of members of the group and outsiders.

2.5.2 Leadership applied to the experimental scenarios

The model of leadership outlined above applies to two of the general experimental scenarios for the SIEMPRE project; that of the small group ensemble (such as a string quartet) and that of the large conducted ensemble. However, the role of the leader in these two scenarios contrast in several ways.

2.5.2.1 Conductor-musician scenario

It is fairly straightforward to explain what makes the conductor of an orchestra its leader. The conductor uses verbal and non-verbal signals to instruct the musicians to behave in certain ways which then they attempt to follow. Most of the work of the conductor is done during rehearsal, where instructions during performance typically refine or reinforce instructions given during rehearsal. Gestural signals by the conductor determine timings, tempo, dynamics and general expressive emphasis both for the group as a whole and individuals within the group.

2.5.2.2 Musician-musician scenario

Meanwhile, for a musician playing one part within an ensemble to lead that group, the question of leadership is much more ambiguous. Most important to note is that the musical leader need not be formally specified. There may be no recognised leader at all. Group members may even battle for leadership. If a leader is recognised, they may be more or less dictatorial in their approach. Leadership can also swap between group members during performance depending on the musical task. For instance, in improvisatory music, a musician who has a clear or novel idea may take the lead. In scored music, the part which plays the greatest role in maintaining group coherence will be a natural candidate for a leader. Leadership can even be split such that one musician is followed with regards to tempo where another is followed with regards to expression. Varni et al. 2010 and Glowinski et al. 2010 propose a model of leadership respectively based on chronemics, i.e. temporal cues and on analysis of group and individual behavioural complexity.

Certain background social factors also influence the probability with which a musician will assume leadership of an ensemble, and as such these factors need to be monitored in any experimental manipulations.

1. Background expertise of the musicians. Greater expertise is generally recognised as enabling leadership.
2. Prior mutual familiarity of musicians. When musicians are already familiar with each other they will have established background norms for recognising musical-leadership, as well as conventions for leadership style impacting on the manner of decision making, communication, feedback and deference.
3. Background personalities of the musicians and their compatibility.
4. Degree to which individual musicians have personal stake in the performance, or whether there is a group reputation to protect.

2.5.2.3 The leader versus the soloist

It is also important that we do not confuse the group leader with the *soloist*. The leader is whichever role is recognised as playing the most important part in maintaining group cohesion-where group cohesion is understood in terms of forming and following the goals of the group e.g. keeping in time or setting a particular musical atmosphere. The case of the soloist meanwhile is when one part is permitted to strike off individually with the group’s support, generally because their part is recognised to be of particular melodic or expressive interest (in other contexts, this figure is known as ‘the talent’). Thus the soloist may have greater expressive freedom for the duration of their solo, including freedom over timing, and will tend to play louder. When this happens, the leader’s role is compromised in ways that will be specified below. In some musical scenarios the roles of leader and soloist may overlap.

The study by Glowinski et al (2010) on the string quartet is a first attempt to distinguish between the musical leader and the musical soloist. This study focuses on music part where the first violin is a clear protagonist with respect to the other musicians (e.g., “*canto accompagnato*, acommon 17-18th musical figure style for string quartet). Experimental manipulation consisting in swapping the musical score of the first and second violin was specifically devised to distinguish between the influence of music part itself and the performance characteristics. Results suggest that a loss in behaviour complexity is a necessary but not sufficient condition for having leadership. The soloist's bodily movements may be less complex because they need not constrain themselves in order to follow the other musicians so much. Meanwhile, the musician-leader's bodily more regular movements should serve to enhance the integration of the ensemble. The second condition is actually that behaviour complexity of the leader contribute to the decrease of the group activity entropy (*drive towards order*).

2.5.2.4 Constraints on the musician-leader

If one musician takes leadership during rehearsal then this more or less follows the same pattern as the conductor. However, during the performance itself, it is clear that due to the additional demands on the musician-leader to play their part, he or she is constrained with respect to the decision condition, the communication condition and the feedback condition. For example, a first violin may make certain gestures regarding timing. He or she may glance periodically at the other musicians to indicate certain expectations that are common knowledge to both, or more generally use facial expressions to signal approval or disapproval, but he or she lacks the resources to do all this to the same degree as the conductor.

As a result, these conditions are compensated in two major ways.

The first major form of compensation is that decision, communication and feedback conditions become partly manifested in the musician-leader's performance. The timings, tempo, dynamics and general expressive emphasis of the musician-leader's performance differ in overall structure in ways that express their dominance:

i) decision condition: asserting how the music *should* sound according to their example. This amounts to playing in a clear and definite style/tempo/dynamic etc. and sticking to it.
ii) communication condition: playing louder, or with emphasis on hierarchically important moments such as timing divisions or exaggerated expressive effects.
iii) feedback condition: here there is the greatest constraint. If there is conflict, the leader can only either re-assert their role in the two above ways with greater emphasis, or give up their leadership in deference to the other by following them instead.

As mentioned above, if there is a soloist, this will constrain the leader further because while the leader can maintain a clear and distinct style they cannot be too assertive if this means conflicting with the soloist. Similarly, if it is the leader who takes on the role of the soloist, then they temporarily give up their capacity to lead since they no longer want the others to follow their example.

The second major form of compensation concerns the deference condition. The responsibility of the members of the group to follow the leader's instructions is enhanced. The increase in this role is proportional to degree of constraints faced by the leader and the way his or her instructions and feedback are manifested in his individual performance. The followers must adjust to the way the leader plays. As such it is partly their responsibility to fulfil the feedback condition by assessing the degree to which their performance coheres with the leader's part.

2.5.3 Experimental applications

Given this theoretical background we are now in a position to understand how modifying the leadership role might impact on the ensemble performance. In general we can manipulate the role of leadership at different levels, then track the effects of this leadership style on emotional synchronization, and the quality of performer and audience experience. Note that in all cases, the role of any soloist must be monitored or controlled so that experimental manipulations are not confounded by the additional influence of a soloist.

High level manipulations
In the following conditions, we dictate broad features of leadership, such as whether there is a leader at all, and the general manner in which decisions are made and communicated:

i) no leader.
ii) leadership battle (secretly instruct two different musicians to assume leadership roles).
iii) democratic unanimity required*.
iv) majority rules*.
v) sensitive leader.
vi) insensitive leader.

(*= manipulations that are only possible during rehearsal scenarios)

Mid-level manipulations
In these conditions we manipulate the specific features of the music, or map of saliences, over which the leader has control:

i) the general goal of the performance*
ii) broad features of the piece, especially expressive content.
iii) musical hierarchical features (tempo, timings, dynamics, timbre etc.)

Low-level manipulations
At this level, we manipulate the 4 fundamental conditions of leadership. It is hypothesized that these low level manipulations are the clearest and most effective way to manipulate the nature of leadership and its effects on group cohesion, coordination, emotional synchronization and performance quality.

i) decision condition. The leader is instructed to either make all decisions ahead of time, or spontaneously during performance.
ii) communication condition. We constrain the way the leader communicates (e.g. visually or only using their musical part).
iii) deference condition. Group members are told whether or not to defer to the leader. Alternatively, we use a scenario in which there is potential conflict between a leader and a soloist.
iv) feedback condition. The leader does not respond to efforts of group members (e.g. using a video performance of the conductor or first violin that the others must follow).

2.5.4 References

Brain stem reflex (2007). In *Dorland's Medical Dictionary for Health Consumers*. Saunders, an imprint of Elsevier, Inc. Retrieved November 3d http://medical-dictionary.thefreedictionary.com/brain+stem+reflexes .

J. K. Burgoon and N. E. Dunbar. Nonverbal expressions of dominance and power in human relationships. In V. Manusov and M. Patterso, editors, The Sage Handbook of Nonverbal Communication, CA: Sage, 2006. Eds. Thousand Oaks.

Camurri A., Varni G., Volpe G. (2010). "Computational Model of Entrainment within Small Groups of People: Toward Novel Approaches to KANSEI information Processing", in Proc. of Intl. Conference on Kansei Engineering and Emotion Research 2010 (KEER2010), Paris (ISBN 978-4-9905104-0-4)

Clayton, Sager & Will (2005) In time with the music: The concept of entrainment and its significance for ethnomusicology

Cochrane, T. (2009). Joint attention to music. *British Journal of Aesthetics*. Vol. 49, No. 1, Jan 2009: 59-73.

De Waal, F.B.M. (2008). Putting the altruism back into altruism: The evolution of empathy. *Annual Review of Psychology, 59*, 279-300.

Fadiga L, Fogassi L, Pavesi G, Rizzolatti G. (1995). Motor facilitation during action observation: a magnetic stimulation study. J Neurophysiol. 73(6):2608-11.

Glowinski, D., Coletta, P., Volpe, G., Camurri, A., Chiorri, C. & Schenone, A. (2010). Multi-scale entropy analysis of dominance in social creative activities. ACM Multimedia Intl Conference, 2010: 1035-1038.

Grandjean, D., Sander, D., & Scherer (2008). Conscious emotional experience emerges as a function of multilevel, appraisal-driven response synchronization. *Consciousness and Cognition*, 17, 484-495.

Hatfield, E., Cacioppo, J. T., & Rapson, R. L. (1994). *Emotional contagion*. Cambridge, UK: Cambridge University Press.

Juslin, P. N. (in press). Music and emotion: Seven questions, seven answers. In I. Deliège & J. Davidson (Eds.), *Music and the mind: Investigating the functions and processes of music.* New York : Oxford University Press.

Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and Brain Sciences, 31*, 559-621.

Juslin, P. N., Liljeström, S., Västfjäll, D., & Lundqvist, L. (2010). How does music evoke emotions? Exploring the underlying mechanisms. In *Handbook of music and emotion: Theory, research, applications*, Series in Affective Science. New York: Oxford University Press.

Keller, P. E. (2008). Joint action in music performance. In F. Morganti, A. Carassa, & G. Riva (Eds.), *Enacting intersubjectivity: A cognitive and social perspective to the study of interactions*. Amsterdam: IOS Press: 205-221.

Koelsch, S. (2010). Towards a neural basis of music-evoked emotions. *Trends in Cognitive Sciences. 14*(3), 131-137.

Krumhansl, C. L. (1997). An exploratory study of musical emotions and psychophysiology. *Canadian journal of experimental psychology, 51(4)*, 336-352.

Lamm, C., Batson, C. D., & Decety, J. (2007). The neural substrate of human empathy: Effects of perspective-taking and cognitive appraisal. *Journal of Cognitive Neuroscience, 19*(1), 42–58.

Leiberg, S., & Anders, S. (2006). The multiple facets of empathy: a survey of theory and evidence. *Progress in Brain Research, 156*, 419-440.

Leventhal, H., & Scherer, K. R. (1987). The relationship of emotion to cognition: A functional approach to a semantic controversy. Cognition and Emotion, 1, 3–28.

Mancas, M., Glowinski, D., Volpe, G., Coletta, P., Camurri, A., "Gesture Saliency: a Context-aware Analysis", in S. Kopp and I. Wachsmuth (Eds.), Gesture in Embodied Communication and Human-Computer Interaction, Lecture Notes in Artificial Intelligence (LNAI), Vol. 5934, p. 146-157, ISBN 978-3-642-12552-2, Springer Verlag Berlin / Heidelberg, 2009

Matsumoto, D., & Ekman, P. (2009). Basic emotions. In D. Sander & K.R. Scherer (Eds.), *Oxford Companion to Emotion and the Affectve Sciences*. Oxford University Press.

Molnar-Szakacs I, Overy K. (2006). Music and mirror neurons: from motion to 'e'motion. *Social Cognitive and Affective Neuroscience, 1*(3), 235-241.

Moors, A. (2007). Can cognitive methods be used to study the unique aspect of emotion: An appraisal theorist's answer. *Cognition & Emotion, 21*(6), 1238-1269.

Oscillation. (2009). In *A Dictionary of Physics*. Ed. Andrew M. Colman. Oxford University Press, 2009. Oxford Reference Online. Oxford University Press.

Overy, K., & Molnar-Szakacs, I. (2009). Being together in time: Musical experience and the mirror neuron system. *Music Perception, 26*, 489-504.

Perception. (2009) In *A Dictionary of Psychology*. Ed. Andrew M. Colman. Oxford University Press 2009. Oxford Reference Online. Oxford University Press. Retrieved August 17th http://www.oxfordreference.com/views/ENTRY.html?subview=Main&entry=t87.e6165

Process. (1999). In A Dictionary of Psychology. Ed. Andrew M. Colman. Oxford University Press, 2009. Oxford Reference Online. Oxford University Press. Retrieved October 26th, 2010, from http://www.oxfordreference.com/views/ENTRY.html?subview=Main&entry=t87.e6674

Resonance. (2009). In *A Dictionary of Physics*. Ed. John Daintith. Oxford University Press, 2009. *Oxford Reference Online*. Oxford University Press. Universite de Geneve. 14 October 2010 http://www.oxfordreference.com/views/ENTRY.html?subview=Main&entry=t83.e2638

Resonance. (2006). In *The American Heritage Dictionary of the English Language, Fourth Edition.* Published by Houghton Mifflin Company. All rights reserved, © 2006 by Houghton Mifflin Company.

Scherer, K. R. (2005). What are emotions? And how can they be measured?". *Social Science Information*., 44(4), 693–727.

Synchronous. (2009). In *A Dictionary of Physics*. Ed. John Daintith. Oxford University Press, 2009. *Oxford Reference Online*. Oxford University Press. Universite de Geneve. 3 November 2010 <http://www.oxfordreference.com/views/ENTRY.html?subview=Main&entry=t83.e3000>

G.Varni, A.Camurri, P.Coletta, G.Volpe (2008) "Emotional Entrainment in Music Performance", *Proc. 8th IEEE Intl Conf on Automatic Face and Gesture Recognition, Sept. 17-19, Amsterdam (ISBN 978-1-4244-2154-1/08).*

G.Varni, G.Volpe, A.Camurri (2010) A System for Real-Time Multimodal Analysis of Nonverbal Affective Social Interaction in User-Centric Media. **IEEE Transactions on Multimedia**, Vol.12, No.6, pp.576-590

Zentner, Grandjean & Scherer 2008 Emotions Evoked By The Sound Of Music

Zentner, M., Grandjean, D., & Scherer K.R. (2008). Emotions evoked by the sound of music: Characterization, Classification and Measurement. *Emotion, 4* (8), 494-521.

2.6 Quality of experience

In studying the emotional and interpersonal communication that makes live performance and listening unique phenomena the SIEMPRE project has identified a number of important mechanisms which will be used as measurement. Physiological signals, bodily movements and analysis of the music are all important in identifying aspects of entrainment, synchrony and co-creation however in order to contextualize them within the overall strategy of understanding why live performances are unique there is a need for subjective measures that assess the quality of experience as well. Quality of experience (QoE) is a broad term that refers to how the performance is experienced by the audience, and in this sense "quality" refers to the nature rather than the positive/negative rating of the performance. By asking participants a number of measures that are designed to assess the QoE we can investigate how the physiological and other quantitative measures we are taking are related to the personal experience of live performance, and will offer an explanatory as well as confirmatory measure of synchrony, entrainment and co-creation. Because of these reasons it is absolutely essential that QoE is considered an important aspect of the SIEMPRE project. There are a number of ways in which we propose to measure QoE that combine to provide an overall picture; continuous rating and retrospective questionnaires by the audience as well as post-performance rating of the audience by judges. This discussion will focus on developing a retrospective questionnaire for QoE in music, the areas which might contribute to our understanding of QoE and the final terms that might be included.

So what are the criteria for inclusion on a questionnaire purporting to measure QoE, given how wide a term it could be? A helpful start is to distinguish QoE from emotional experience. Whilst felt emotions do of course affect the QoE, that is not the core of what QoE represents. For example a performer might communicate a number of things to the audience, only one of which is emotion in the song. But also QoE is not simply communication, there are many intrapersonal and other factors which could affect QoE even after effective communication with the performer or other audience members. For the sake of this study it is also beneficial to prioritise concepts that might have more connection to the other measures being taken (bodily movement, physiology and other qualitative measures). A perfect categorization of which factors most contribute to QoE in music is currently impossible (otherwise there would be no need for this discussion!) but it is hoped through good rationale and testing we can find an acceptable questionnaire for use in the future.

The first place to look for concepts (and thus potential items) for this questionnaire is the psychology of music and emotion. Surprisingly little work has been done on live music performance specifically (as opposed to general music listening) but there are still many theories that can contribute something to a QoE questionnaire. The circumplex model (Russell, 1980) from general emotion theory is used extensively in music emotions research, and both dimensions should be considered for inclusion. The valence factor which represents the positive/negative feeling however might be more useful for emotion theory than for QoE and not as useful outside of its specific model, it will be included but alternatives will also be sought. Arousal on the other hand can be considered very important to understanding QoE and has far better links to the physiology than valence rating which are notoriously hard to differentiate. Meyer's (1956) work on the origins of music and emotion highlight expectancy as the important factor in creating emotions in listeners, and work influenced by this from Krumhansl (1997) cites engagement and tension as important. The latter two factors are definitely for inclusion in a QoE questionnaire as they aren't specific to music itself and reflect the more general experience of music listening, which is very beneficial as there are ways in which the performer might create additional expectancies or engagement beyond that of the music itself. Other theories of emotions in music can also add factors to the QoE questionnaire, particularly basic emotions theory and the GEMS model. Basic emotions theory has been regularly applied to both felt and expressed emotion in music with varying degrees of success and through this there are a few emotions we can choose to ask directly such as happiness, sadness and anger. Other emotions considered "basic" by theorists (fear, surprise, disgust) are considered less prevalent for music than for general emotional life and thus can be omitted from the concept of QoE. The GEMS model developed recently by Zentner, Grandjean & Scherer (2008) is an newer alternative to the previous two models, with the advantage that it was designed especially for measuring felt emotion to music (whereas the other models were taken from the emotion literature). The GEMS model settles on 9 factors over a 40 item questionnaire divided into 3 categories. Given the possible demand characteristics of the QoE questionnaire it would not be advisable to include 9 items just for the GEMS, however it is possible to include the 3 over-reaching categories (sublimity, vitality, unease) so incorporation of aspects of the GEMS is feasible.

Performers can potentially alter a piece of music in many ways during a performance such as tempo, instrumentation etc. Juslin and Timmers (2010) propose a model that categorises expression of emotion in music and the means through which it can be achieved – GERMS (generative rules, emotional expression, random fluctuations, motion principles and stylistic unexpectedness. These are important as large changes will affect audience's QoE not necessarily because of the music itself, but rather because it deviates significantly from the pre-conceived idea of how the song/piece is expected to sound. It is useful to take influence from this model for a number of items that measure whether the performance matched audience's expectations of what the piece or performer should sound like. Anecdotally the importance of this is known by every rock & roll historian from when Bob Dylan "went electric" for the first time at the Newport Folk Festival in 1965, causing him to be booed off stage by his (staunchly folk) fans. Other research conducted by Minassian, Gayford and Sloboda (2003) found that for performers "optimal" performances showed 3 factors; clear intention to communicate, emotionally engagement with the music and belief that the message had been received by the audience. For a QoE questionnaire the factors that can be drawn from this are the intent of the performer to communicate, and the success of that communication. Several items on these will be included on the final questionnaire.

Psychological theories on music aren't the only domain that should guide a discussion on QoE as there has been much work done on the topic by other areas. Ignoring occupational and health psychology (which uses the term quality of experience in an entirely different way) there are many other areas that can contribute to our understanding of QoE in music. One of the key influences for QoE in music will be the concept of flow (Csikszentmihalyi, 1990), first coined to reflect complete immersion in an activity. Flow has been used in affective computing and other activities where immersion is a key concept, but has not been fully applied to a live performance context for the audience. Ten factors are hypothesized for the experience of flow Csikszentmihalyi (1993), and a large amount of these can be included in a QoE questionnaire for music. Examples include distorted sense of time, focus of awareness, concentration, loss of self-awareness, loss of awareness of bodily needs and immediate feedback (for music this could be a feedback loop with the performers or other audience members). In musical terms flow has been related to peak experience (Panzarella, 1980) and strong experiences with music (SEM, Gabrielsson, 2010). Indeed items from the peak experience topic (renewal ecstasy, motor-sensory ecstasy, withdrawal ecstasy and fusion-emotional ecstasy) will also be addressed in QoE for music as they represent an additional means of describing the QoE.

Affective computing in particular is of interest as there has been much work on measuring the quality of experience of a user interacting with a machine. There are two main concepts that can be extracted from this literature and that is flow and presence. Having discussed flow previously the next topic to cover is presence. In affective computing presence refers to virtual environments the subjective experience of being in one place or environment, even when one is physically situated in another. For music this could be conceptualized as the feeling of transporting with the music etc. Personality differences account for a lot of the differences in presence (Sas, Corina and O'Hare, 2003) so it might also be useful to include aspects from the Tellegen absorption scale (TAS) (Tellegen, 1982) which is similar to "openness" in OCEAN, but far more specific and useful for these purposes, or alternatively (and preferably due to its brevity) “willingness to experience presence/suspend disbelief”. This might account not just for differences in presence but also differences in the overall QoE.

There are a number of other factors which can be considered important for determining QoE in music. From emotional literature there is the additional component of potency (Fontaine et al., 2007) and connectedness (with the performer, music and audience). From social psychology and philosophy there are the concepts of power and dominance. There is also the phenomena of frissons (Gabrielsson, 2010), which are strongly associated with strong experiences with music, and some additional items we believe might be influential in detecting QoE in live musical performance (feelings about performer etc.).

Concepts which may be useful for the analysis of Quality of Experience (QoE) are briefly listed below:

Drawn from the emotions in music literature

Arousal Valence	Circumplex
Happy Sad Anger	Basic Emotions Theory
Sublimity Vitality Unease	GEMS
Tension	

Social literature

Power

Dominance

Potency

Connectedness - with the performer
- with other audience members
- with the music

Relationship with performer

Empathy

Awareness of performer

Did the performer show... - emotional engagement with the music?
- a clear intention to communicate with the audience?

Matching to pre-conceptions of song/performance expectation - performer
- music

Enjoyment of expectedness/unexpectedness

Attentional and Flow

Engagement	
Concentration	
Awareness of bodily reaction/needs	Flow
Awareness of surroundings	
Distorted sense of time	
Self-awareness	
Immediate feedback	
Renewal ecstasy	Peak experience
Motor-sensory ecstasy	
Withdrawal ecstasy	
Fusion-emotional ecstasy	

"Presence" (HCI)

Presence in the musical space

Willingness to suspend disbelief

2.6.1 References

Csikszentmihalyi, M. (1990) Flow: The psychology of optimal experience: Steps toward enhancing the quality of life. Harper Collins Publishers

Csikszentmihalyi, M. & Rathunde, K. (1993). "The measurement of flow in everyday life: Towards a theory of emergent motivation". In Jacobs, J.E.. *Developmental perspectives on motivation*. Nebraska symposium on motivation. Lincoln: University of Nebraska Press. p. 60

Fontaine, J., Scherer, K., Roesch, E. & Ellsworth, P. (2007). The World of Emotions Is Not Two-Dimensional. *Psychological Science*, 18 (12), 1050 – 1057

Gabrielsson, A. (2010). Strong experiences with music. In J. A. Sloboda (Ed.), *Handbook of music and emotion: Theory, research, applications.* (pp. 547-574). New York, NY US: Oxford University Press.

Glowinski, D. Camurri, A. "Musique et émotions" in Catherine Pelachaud (eds) (2010) "Systèmes d'Interactions Emotionnelles", Hermes Science éditions, Paris, France.

Juslin, P. N., & Timmers, R. (2010). Expression and communication of emotion in music performance. In J. A. Sloboda (Ed.), *Handbook of music and emotion: Theory, research, applications.* (pp. 453-489). New York, NY US: Oxford University Press.

Krumhansl, C. L. (1997). An exploratory study of musical emotions and psychophysiology. *Canadian Journal of Experimental Psychology-Revue Canadienne De Psychologie Experimentale, 51*(4), 336-353.

Meyer, L.B. (1956) *Emotion and meaning in music*. Chicago: Chicago University Press.

Minassian, C., Gayford, C., & Sloboda, J. A. (2003, March). Optimal experience in musical performance: A survey of young musicians. Paper presented at the Meeting of the Society for Education, Music, and Psychology Research, London.

Panzarella, R (1980) The phenomenology of aesthetic peak experiences, *Journal of Humanistic Psychology*, v20 n1 69-85.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*(6), 1161-1178.

Sas, Corina and O'Hare, G.M.P. (2003) *Presence Equation: An Investigation Into Cognitive Factors Underlying Presence.*Presence: Teleoperators and Virtual Environments, 12 (5). pp. 523-537.

Tellegen, A. (1982). Brief manual for the Multidimensional Personality Questionnaire. Unpublished manuscript, University of Minnesota, Minneapolis

Tellegen, A. (1985) *Structures of mood and personality and their relevance to assessing anxiety, with an emphasis on self-report*, Lawrence Erlbaum Associates Inc.

Zentner, M., Grandjean, D., & Scherer, K. R. (2008). Emotions evoked by the sound of music: Characterization, classification, and measurement. *Emotion, 8*(4), 494-521. doi:10.1037/1528-3542.8.4.494

3. METHODOLOGY - MULTIMODAL ANALYSIS OF NON-VERBAL SOCIAL SIGNALS

3.1 Measuring quality of experience

3.1.1 Self-report

A traditional distinction exists in the literature between emotions felt by listeners and emotions expressed by music (Zentner & Eerola, 2010). The attribution of emotional qualities of music is a complex process allowing humans to represent and explicitly report feelings expressed through music, whereas the induction of emotions is the process of experiencing emotions, i.e. the feelings, as a result of listening to music (Scherer & Zentner, 2001). There is obviously a high degree of subjectivity in the measurement of emotions felt by the listener and the study of emotions expressed through music or represented in music have the advantage of a certain measure of “objectivity” because it is easier to agree on what emotions music expresses, than on what emotions music induces (Campbell, 1942, quoted in Schubert, 2004).

The main difficulty in the measurement of emotions, moods, or feelings related to music resides on the lack of consensus on the definition of emotion (Scherer, 2004). As pointed out by Zentner and Eerola (2010), self-report instruments are derived from a theory or model of emotion. Most studies rely on two traditional models of emotion: the basic emotion theory and the dimensional model of emotion. Indeed, either for the study of emotions felt by the listener or perceived in the music, the majority of studies use the labels from the basic emotion model such as fear, joy, surprise, sadness (Fritz & al., 2009; Juslin, 1997; Juslin, 2000). The dimensional approach, which proposes to conceptualize the emotion in terms of valence and arousal, is also widely used in the study of musical emotions (Chapin, Jantzen, Kelso, Steinberg, & Large, 2010; Schubert, 1999; Nagel, Kopiez, Grewe, & Altenmüller, 2007). There are different types of self-report instruments and methods such as Likert scales, adjective checklist, diary study, forced-choice category, continuous response versions using a computer interface (Zentner & Eerola, 2010). The choice of instrument depends obviously on the type of research questions and what we want to investigate.

3.1.1.1 Continuous scales

As pointed out by Nagel and colleagues (2007), there are different ways of investigating the emotions related to music, such as self report, questionnaires and adjective scales, but all of these approaches are static and therefore unable to demonstrate the complexity of the unfolding of musical emotions. The works of Emery Schubert (1999; 2001; 2004) have been among the first to take into account this characteristic of time and to use continuous measurements. This method allows experimenters to record the judgments of emotions expressed by music in real time and then to follow the changes of perception and attribution over time.

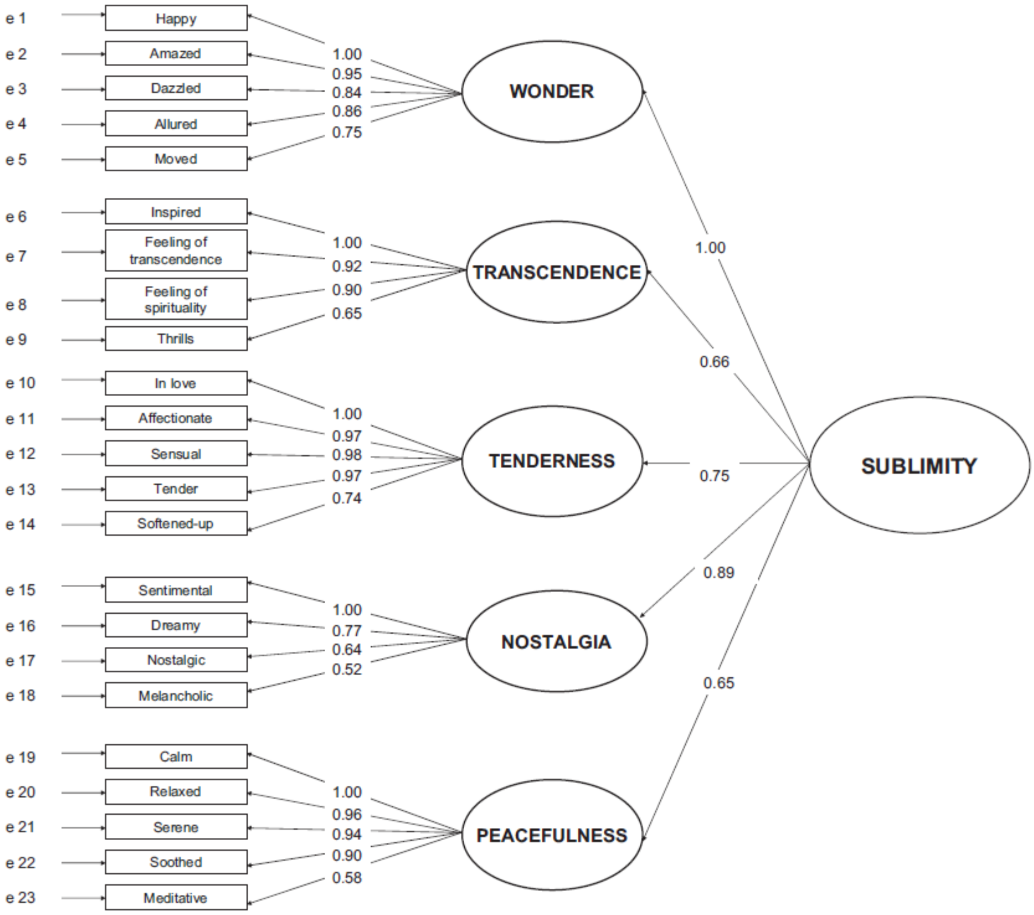
3.1.1.2 Categorical report

Categorical judgments are often used in the literature in order to judge emotions expressed through music or feelings related to musical performances. Even this kind of method is interesting it could be biased to language categorization processes. We demonstrated that continuous judgments and categorical judgment are not the same and the results obtained by these two methods can be slightly different driving different kinds of conclusions or interpretations. In consequence, some results in the literature might be influenced by the method used asking people to assess emotions or other kinds of phenomena (see for example

Péron, Grandjean, et al., 2010). In SIEMPRE studies, we suggest to prefer continuous judgments (online or offline) rather than categorical judgments when it is possible.

3.1.1.3 Geneva Emotional Music Scale (GEMS)

The majority of studies on music and emotion propose to judge musical excerpts in terms of valence and arousal (Vieillard et al., 2008; Chapin, Jantzen, Kelso, Steinberg, & Large, 2010) or in terms of basic emotions (Fritz et al., 2009; Juslin, 2000). However, one might suppose that musical emotions are more complex or subtle and then these approaches might not be the best to understanding emotions related to music. In this context, Zentner, Grandjean and Scherer (2008) propose a new approach for the study of emotions in music. They made a set of experiments enabling them to propose a factorial model of the most relevant emotional terms for the understanding of emotions related to music. These studies gave rise to a nine factorial model of emotions induced by music: the GEMS (Geneva Emotion Music Scale) (Figure 1). In a fourth study, the authors confirmed the nine-dimensional structure of the model and demonstrated that this new framework is more appropriated than the two traditional models of emotion, namely the basic emotion model and the dimensional emotion model. The GEMS model currently represents the most effective attempt for studying the emotions related to music.



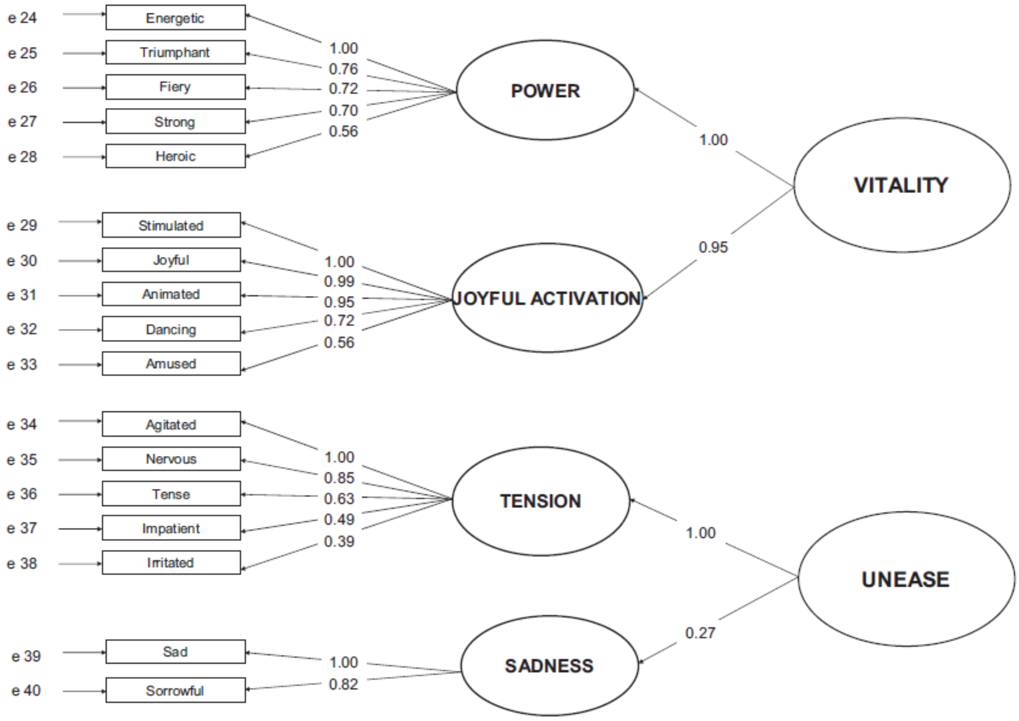


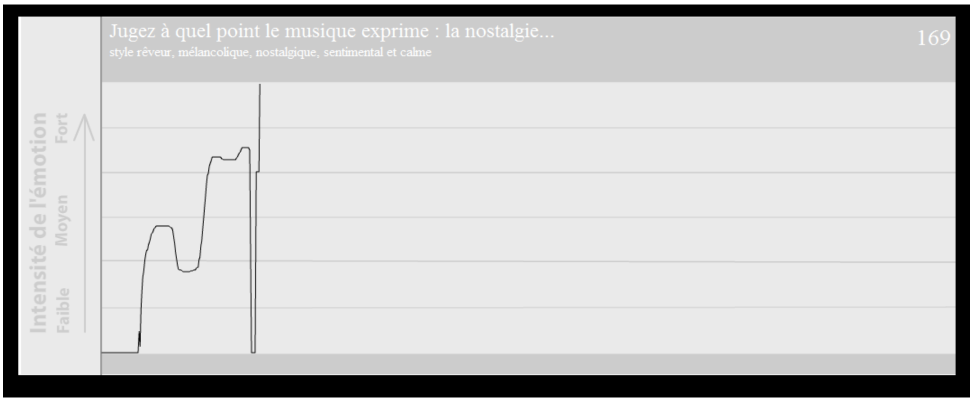
Figure 1. The 9-factorial model of music-induced emotions (Zentner, Grandjean & Scherer, 2008)

3.1.1.4 Dynamic judgements

An obvious feature of music is that it unfolds over time, as emotion does. In order to effectively apprehend the emotions expressed by music, it is preferable, and probably essential, to base the judgments on continuous measurements. For this purpose, we propose to use an approach called "dynamic judgment" (Figure 1).

This new method of dynamic judgment based on the 9 dimensions proposed by the GEMS and using a Flash Interface, allows us to record the dynamic judgments in real-time in a graphical manner. The width of the graph is 1000 pixels (equivalent to 4'16) and the height 300 pixels. Participants have direct visual feedback in the graphic interface of the judgments they were made by moving a cursor up and down as time advances (if necessary the graph-window slides).

Measurements are made every 250 milliseconds. The x-axis represented time, while the y-axis represented the intensity of the emotion expressed by music (e.g. Nostalgia) through a continuous scale marked by three levels of intensity: low, medium, and high.


Jugez à quel point le musique exprime : la nostalgie...

style rêveur, mélancolique, nostalgique, sentimental et calme

Figure 1. Example of the interface of dynamic judgments, here for the dimension of "Nostalgia".

We conducted several experiments enabling us to demonstrate that this method is relevant in a laboratory context as well as during live performance.

Figure 2 and Figure 3 show respectively dynamic judgment in a laboratory context and during a live performance with a professional quartet, “Il Quartetto di Cremona”.

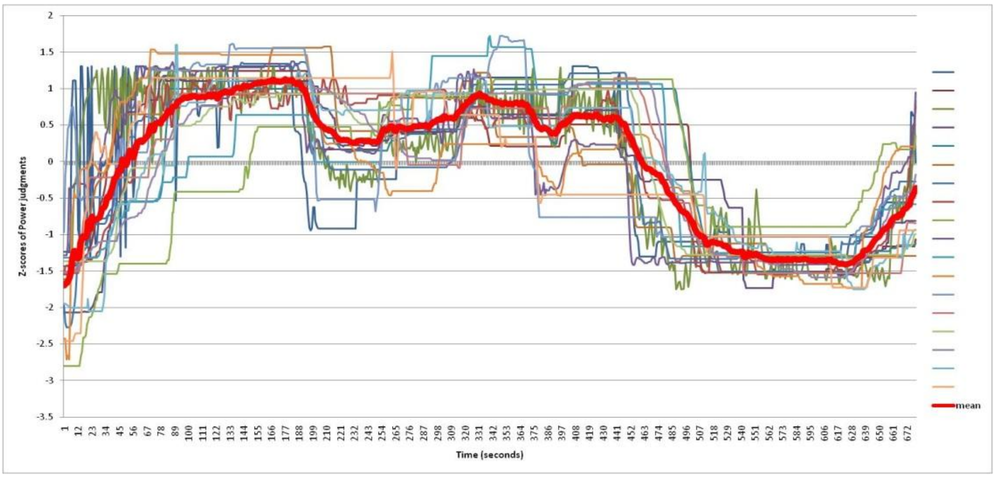


Figure 2. Individual z-scores (N=18) for the dynamic emotional judgment of the 4th movement of the New World Symphony by Dvorak judged on the dimension of « Power », during a laboratory context.

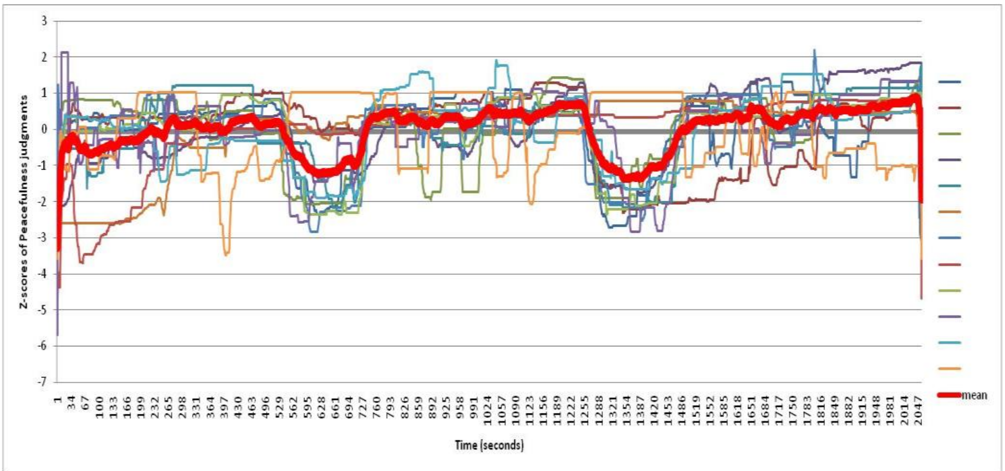


Figure 3. Individual z-scores (N=12) for the dynamic emotional judgment of the 3rd movement of the String Quartet n.3 in A major by Schumann judged on the dimension of « Power », during a live performance.

This new method of dynamic judgment allows us to better understand emotions expressed by music and also to cut the process of attribution of emotional qualities of music. In future studies we will use multiple regression and Granger causality methods (see below) in order to understand how people are able to attribute emotions based on perceptual basic phenomena of acoustical features and musical structures.

3.1.2 References

Chapin, H., Jantzen, K., Scott Kelso, J.A., Steinberg, F., Large, E. (2010). Dynamic Emotional and Neural Responses to Music Depend on Performance Expression and Listener Expression. PLoS ONE 5, (12).

Fritz, T., Jentschke, S., Gosselin, N., Sammler, D., Peretz, I., Turner, R., Friederici, A.D., & Koelsch, S. (2009). Universal Recognition of Three Basic Emotions in Music. *Current Biology*, *19*, 573-576.

Juslin, P. (1997). Emotional communication in music performance: a functionalist perspective and some data. *Music Perception, 14*, 383-418.

Juslin, P. (2000). Cue utilization in communication of emotion in music performance: relating performance to perception*. Journal of Experimental Psychology: Human Perception and Performance*, *26*, 1797-1813

Nagel, F., Kopiez, R., Grewe, O., & Altenmüller, E.(2007). EMuJoy: Software for continuous measurement of perceived emotions in music. *Behavior Research Methods, 39*, 283-290.

Péron, J., Grandjean, D., Le Jeune, F., Sauleau, P., Haegelen, C., Drapier, D., Rouaud, T. , Drapier, S., & Vérin, M. (2010). Recognition of emotional prosody is altered after subthalamic nucleus deep brain stimulation in Parkinson’s disease. *Neuropsychologia, 48*, 1053–1062.

Scherer, K.R. (2004). Which emotions can be induced by music? What are the underlying mechanisms? And how can we measure them? *Journal of New Music Research, 33,* 239-251.

Schubert, E. (1999). Measuring emotion continuously: validity and reliability of the two-dimensional emotion-space. Australian Journal of Psychology, 51, 154-165.

Schubert, E. Continuous measurement of self-report emotional response to music.(2001). In P. Juslin & J. Sloboda (Eds*.), Music and emotion: Theory and research,* pp. 361--392. Oxford, England: Oxford University Press.

Schubert, E. (2004). Modeling Perceived Emotion with Continuous Musical Features. *Music Perception*, 21, 561-585 (2004)

Vieillard, S., Peretz, I., Gosselin, N., Khalfa, S., Gagnon, L., & Bouchard, B.(2008). Happy, sad, scary and peaceful musical excerpts for research on emotions. *Cognition & Emotion, 22,* 720-752.

Zentner, M., Grandjean, D., & Scherer, K.R. (2008). Emotions Evoked by the Sound of Music: Characterization, Classification, and Measurement. *Emotion*, *8*, 494-521

Zentner, M., & Eerola, T. (2010). Self-report measures and models. In P. Juslin & J. Sloboda (Eds.), Music and emotion: Theory and research, pp. 187--221. Oxford, England: Oxford University Press.

3.2 Body movement analysis

This section presents methodological issues and techniques involved in the analysis of body movement both at individual (expressive gesture) and group level (expressive social interaction).

Recent neuroscientific and psychological studies revealed that body postures and movement are important for conveying emotions (see (De Gelder 2006) for an overview). The increasing reliability of motion capture and other vision-based movement tracking and analysis methods has also contributed to the increasing interest in the computational study of expressive body movement. In addition, there is an evidence on the role of body in music and in particular on its communicational aspects that support the identification of suitable behavioral variables for the analysis of social interactions (Dahl et al. 2010, Davidson et al. 2005; Castellano et al 2008; Timmers et al 2006).

3.2.1 Full-body coarse postures

Existing studies on full-body movement use coarse-grained posture features (e.g., leaning forward, slumping back) or low-level physical features of movements (kinematics). Bianchi et al., (2006) formalized a general description of posture based on angles and distances between body joints and used it to create an affective posture recognition system that maps the set of postural features into affective categories using an associative neural network. Mota and Picard, (2003) showed how sequences of postures can be predictive of affective states related to a child's interest level during a learning task on a computer. Natural occurring postures' data from ten children were collected through pressure sensors mounted on a chair. Hidden Markov Models (HMMs) were used to analyze temporal patterns among nine posture sequences to characterize three affective states (high and low interest and behavior of taking a break). A similar study was conducting by Kapoor et al., (2007) to detect pre-frustration behavior using multiple nonverbal channels of information.

3.2.2 Full-body kinematics

Other approaches have exploited the dynamics of gesture referring to few psychological studies reporting that the temporal dynamics play an important role for interpreting emotional displays. Kapur et al. (2005) used full-body skeletal movement data (obtained with the VICON motion capture system on five participants) to distinguish automatically between four basic emotional states (sad, joy, anger, and fear). 3D positions of fourteen body joints were recorded over time to identify the movements performed by actors for each of the selected emotion. The authors showed that very simple statistical measures of motion dynamics (e.g., velocity and acceleration) are sufficient for training successfully automatic classifiers (e.g., SVMs and decision trees classifier). The role of kinematic features has been further established by the recent study of Bernhardt and Robinson (2007). Further developing the motion-captured knocking motion from Pollick et al. (2003), they developed a computational approach to extract affect- related dynamic features. Velocity, acceleration, and jerk measured for each joint composing the skeletal structure of the arm proved successful in the automatic recognition of the expressed affects (neutral, anger, happy, and sad).

3.2.3 Expressive Gesture Analysis

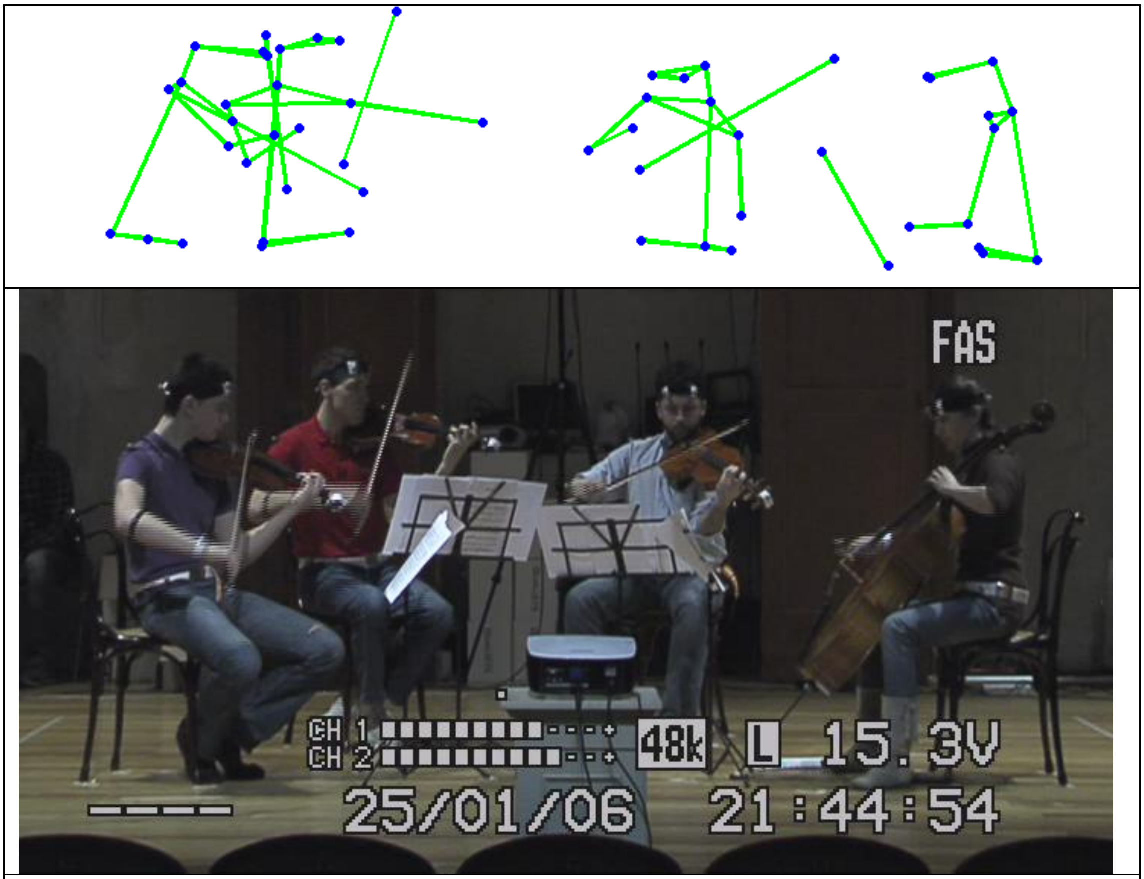
Camurri et al. (2004) developed a qualitative approach to human full-body movement for affect recognition referring to literature in psychology (e.g., De Meijer, 1989; Boone and Cunningham, 1998; Wallbott, 1998) and in humanistic tradition (Theory of Effort choreographer R. Laban, 1971)). Starting from low-level physical measures of the video-tracked whole-body silhouette, they identified motion features such as the overall amount of motion computed with silhouette motion images, the degree of contraction and expansion of the body computed using its bounding region, or the motion fluency computed on the basis of the changes magnitude of the overall amount of motion over time. On the basis of these motion features, they were able to distinguish between four emotional performances of a dance sequence (anger, fear, grief, and joy) by four dancers. Other examples of relevant movement and gesture expressive features include impulsiveness (i.e., whether movement is smooth or rough), contraction (i.e., whether the body is contracted or expanded), directness (i.e., whether a movement is direct or flexible), symmetry, space occupation, fluidity.

Most of these works however attempt to recognize a small set of prototypic expressions of basic emotions like happiness and anger. Moreover, almost all of the existing systems are intended for a single user, whereas social interaction is often neglected. A major research challenge and innovative aspect in SIEMPRE therefore consists of:

(i) analyzing subtler and more significant emotional expressions (e.g., those responsible of emotional contagion) conveyed by full-body and gesture and use them as input for analysis of social interaction (Gatica-Perez, 2009), Vinciarelli et al., 2009).

(ii) developing group cues will also be developed to model the entire group as a collective rather than focusing on each individual separately. This includes for example the dynamics of the polygon, individuated by the heads of the musicians, and of its baricenter.

(iii) developing an entropy-based framework drawing on concepts from the theory of chaotic systems to model and analyze nonlinear dynamics of human actions. Temporal dynamics of human behavior (i.e., timing and duration of behavioral features) can actually be critical for distinguishing between observed behavioral expressions.



Students from the Music Conservatory Niccolò Paganini of Genoa during an experiment at Casa Paganini / UNIGE. Reflective markers of the Qualysis system are placed in their upper-part body joints to extract high-quality behavioral data.

3.2.4 Entropy-based measures for a dynamic analysis of expressive behaviour

The analysis of features dynamics related to human behavior requires specific computational tools: behavioral signals can be analyzed using traditional approaches based on time and frequency domains. However, such measures fail to account for central properties of human movement dynamics: (i) the non- linearity (small perturbations can cause large effects) and (ii) non-stationarity (the statistical properties change with time). Recent entropy-based measures have been developed to address this issue and to extract information from behavioral time series not contained in traditional methods based on mean and variance (Richman & Moorman, 2000).

One of these measure, Sample Entropy, computes the conditional probability that sequences of behavioral signals similar for m points will remain similar when they are extended with one more measure point (m+1). High values of Sample Entropy indicate that the user introduced a change in his movements.

3.3 Behavioral measure of Expressive Social Interactions

A number of features are developed to model and understand social dynamics at work in small group (such as Quartet and Small Orchestra).

3.3.1 Behaviour Saliency

Behavior saliency can be modeled starting from recent results obtained in the field of *computational attention* (Mancas, 2010). In this framework, salient behavior is understood as a behavior capturing the attention of the observer. The saliency index can be computed on any low-level (e.g., speed) and mid-level (e.g., Motion Index) features and indifferently on *n*-number of participants. Motion features can be compared in the spatial context of the current video frame (e.g., one participant with respect to the others in a group) and analyzed on a varying sliding time window (e.g., participant’s movement over time). The Saliency index is based on the so-called self-information. Let us note m_i a message containing an amount of information (e.g., a certain value of a motion index). This message is part of a message set *M* (e.g., the Motion index values over time). The saliency of m_i is defined by :

$$S(m_i) = -\log(p(m_i))$$

where $p(m_i\ )$ is the occurrence likelihood of the message m_i within the message.

3.3.2 Phase Synchronization

The descriptor is computed starting from the measures of the behavior of each user (eiher a raw stream of accelerometer data, or the computation of an higher level index such as Motion Index, Contraction Index, etc.) during a window of a few seconds. Starting from this vector, a Recurrence index is computed, showing how many times the state vector returns close to a previous value.

After normalizing the index on a range [0,1], the autocorrelation between two vectors belonging to different users gives the probability that the state of the system recurs at a given time.

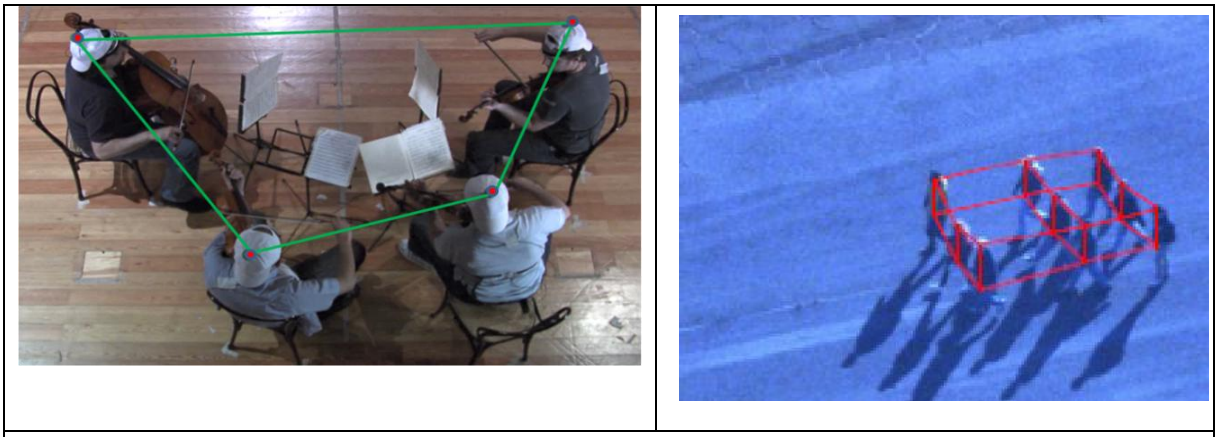
$$\hat{p}_x(\varepsilon,\tau) = RR_\tau(\varepsilon) = \frac{1}{N-\tau}\sum_{i=1}^{N-\tau}\Theta(\varepsilon - \|\vec{x}_i - \vec{x}_{i+\tau}\|)$$

Finally, the phase synchronization index is obtained as the correlation coefficient computed around the normalized generalized autocorrelation functions computed for the users.

3.3.3 Geometric Group Features

Geometric forms can be created to approximate group configurations. In some cases, it can consist in applying existing individual spatial cues (e.g., bounding rectangle or triangle) to a collective situation where the entire group is analyzed instead of a single individual. Usual measure can be performed (e.g., degree of contraction/expansion) to gain insight on the behavior of the group behavior, considered as a single organism. Generic geometric form can be created to adapt to more complex group situation (e.g., polygon relating heads of a string quartet musicians). Additional features can then be devised to adapt to the specificity of group studies. For example the stability of the Center-of-Gravity or polygon area may inform on the group cohesion (e.g., participants are coordinating/compensating movement of the other participants in order to keep the polygon area most stable). Rigidity of formation is also a candidate feature to distinguish between form configuration (see work by Khan et al, 2004).

Body movement is a central component of emotional expression. In addition, there is evidence on the role of body in music and in particular on its communicational aspects that support the identification of suitable behavioral variables for the analysis of gesture expressivity and expressive social interactions.



Example of geometric group features accounting for group dynamics (e.g., cohesion of the group identified the dynamics of the polygon bounding the heads of the musicians).

3.3.4 Leadership (Dominance)

Behavioral analysis of leadership is currently approached through the related concept of Dominance. Dominance is “the set of expressive, relationally based communicative acts by which power is exerted and influence achieved” (Burgoon and Dunbar 2006), by means of forcefulness, monopolizing or involvement.

Two strategies can be devised to measure dominance:

(i) by looking at how, starting from a state where all the members in a group are synchronized, a divergence occurs and then disappears; dominance is seen here as a feature related to the direction of synchronization;

(ii) by analyzing how some member in the group succeed in controlling and regulating their own behavior and the group activity; dominance is seen here as a feature related to behavior complexity.

3.3.4.1 Dominance as a feature related to the direction of synchronization

The direction of the synchronization reveals if the non-synchronized member of the group was a “master” that forced the other members to follow him in the new pattern, or a “slave” that wasn’t fully following the leader. In the former case, a user acts as an external force driving the other/s to reach synchronization; in the latter case all the users combine to bring about synchronization and no driver can be detected.

To compute it starting from two vectors containing the measure of an affective descriptor, the number of recurrence occurring in a short window of time are counted.

$$c^{\tau}(y|x) = \sum_{j=1}^{m_y} \sum_{i=0}^{m_x} J_{ji}^{\tau} \qquad c^{\tau}(x|y) = \sum_{i=1}^{m_x} \sum_{j=0}^{m_y} J_{ij}^{\tau}$$

with:

$$J_{ij}^{\tau} = \begin{cases} 1 & if\ 0 < t_i^x - t_j^y < \tau \\ \frac{1}{2} & if\ t_i^x = t_j^y \\ 0 & otherwise \end{cases} \qquad J_{ji}^{\tau} = \begin{cases} 1 & if\ 0 < t_j^y - t_i^x < \tau \\ \frac{1}{2} & if\ t_i^x = t_j^y \\ 0 & otherwise \end{cases}$$

The relative delay between the two series is then given by

$$q_{\tau} = \frac{c^{\tau}(y|x) - c^{\tau}(x|y)}{\sqrt{m_x m_y}}$$

where qт is between -1 and 1; the absolute value indicates how much leadership is strong, whereas the sign indicates who is the leader. See (Varni et al 2010) for more details.

3.3.4.2 Dominance as a feature related to behavior complexity

Analysis of dominance is here based on the theoretical framework of multi-scale entropy (MSE), a non-linear technique to quantify the behavior *complexity*, i.e., the information expressed by the body movement dynamics over multiple time scales (Costa et al. 2005). Recent studies suggest that the dominant person is the one which behavior complexity is :

(i) relatively low with respect to the complexity of others in the group
(ii) highly correlated with the group activity

The dominant person (*leader*) appears as the one able to “integrate” others’ activity and to decrease the total entropy of the group.

Considering a time series, the computation of the *Complexity Index* (CI) comprises three distinct processes deriving from the Multi-Scale Entropy method :

(i) a coarse-graining procedure to represent the system’s dynamics at different time scales;
(ii) the quantification of the degree of irregularity of each coarse-grained time series through the application of Sample Entropy (SampEn);
(iii) A complexity index (CI) of the time series is calculated by integrating the SampEn values obtained for the different time scales.

Considering a group of 3 users, computation for the dominant person identification starts from the three vectors containing the kinematics measurement of the users (e.g., position) plus one vector containing the kinematics measures relative to the polygon approximating the group (e.g., perimeter, center of gravity of the polygon).

The identification of the dominant user (*leader*) is given by:

(i) compute the Complexity Index for each vector;
(ii) select the users displaying the lowest Complexity Index (potential leader);
(iii) compute the Pearson's product moment correlation of the Complexity Index of each potential leader with the Complexity Index of the polygon's center of gravity;
(iv) select the user displaying the highest correlation (*leader*).

See (Glowinski et al 2010) for more details.

3.3.5 References

N. Bianchi-Berthouze, P. Cairns, A. Cox, C. Jennett, and W.W. Kim. On posture as a modality for expressing and recognizing emotions. In Workshop on the role of emotion in HCI, 2006.

D. Bernhardt and P. Robinson. Detecting Affect from Non-stylised Body Motions. Lecture Notes in Computer Science, 4738:59, 2007

J. K. Burgoon and N. E. Dunbar. Nonverbal expressions of dominance and power in human relationships. In V. Manusov and M. Patterso, editors, The Sage Handbook of Nonverbal Communication, CA: Sage, 2006. Eds. Thousand Oaks.

A. Camurri, Varni G., Volpe G. (2010). "Computational Model of Entrainment within Small Groups of People: Toward Novel Approaches to KANSEI information Processing", in Proc. of Intl. Conference on Kansei Engineering and Emotion Research 2010 (KEER2010), Paris (ISBN 978-4-9905104-0-4)

A. Camurri, B. Mazzarino, and G. Volpe. Expressive interfaces. Cognition, Technology & Work, 6(1):15–22, 2004

M. Costa, A.L. Goldberger, and C.K. Peng. Multiscale entropy analysis of biological signals. Physical Review E, 71(2):21906, 2005

S. Dahl, F. Bevilacqua, R. Bresin, M. Clayton, L. Leante, I. Poggi, and N. Rasamimanana. Gestures in Performance. Musical Gestures: Sound, Movement, and Meaning, page 36, 2009.

J.W. Davidson. Bodily communication in musical performance. Oxford University Press, USA, 2005.

D. Gatica-Perez. Automatic nonverbal analysis of social interaction in small groups: A review. Image and Vision Computing, 27(12):1775 – 1787, 2009. Visual and multimodal analysis of human spontaneous behaviour
B. De Gelder. Towards the neurobiology of emotional body language. Nature reviews. Neuroscience(Print), 7(3):242–249, 2006

Glowinski, D., Coletta, P., Volpe, G., Camurri, A., Chiorri, C. & Schenone, A. (2010). Multi-scale entropy analysis of dominance in social creative activities. ACM Multimedia Intl Conference, 2010: 1035-1038.

A. Kapoor, W. Burleson, and R.W. Picard. Automatic prediction of frustration. International Journal of Human-Computer Studies, 65(8):724–736, 2007.

A. Kapur, A. Kapur, N. Virji-Babul, G. Tzanetakis, and P.F. Driessen. Gesture-Based Affective Computing on Motion Capture Data. Lecture Notes in Computer Science, 3784:1, 2005

S.M. Khan and M. Shah. Detecting group activities using rigidity of formation. Proceedings of the 13th annual ACM international conference on Multimedia, pages 403–406, 2005.

R. Laban and L. Ullmann. The Mastery of Movement. Plays, Inc., 8 Arlington Street, Boston, Mass. 02116,1971.

Mancas, M., Glowinski, D., Volpe, G., Coletta, P., Camurri, A., “Gesture Saliency: a Context-aware Analysis”, in S. Kopp and I. Wachsmuth (Eds.), Gesture in Embodied Communication and Human-Computer Interaction, Lecture Notes in Artificial Intelligence (LNAI), Vol. 5934, p. 146-157, ISBN 978-3-642-12552-2, Springer Verlag Berlin / Heidelberg, 2009

M.Meijer.The contribution of general features of body movement to the attribution of emotions. Journal of Nonverbal Behavior,13(4):247–268, 1989.

S. Mota and R.W. Picard. Automated posture analysis for detecting learners interest level. 2003.

F.E. Pollick, H.M. Paterson, A. Bruderlin, and A.J. Sanford. Perceiving affect from arm movement. Cognition, 82(2):51–61, 2001

A. Vinciarelli, M. Pantic, and H. Bourlard. Social signal processing: Survey of an emerging domain. Image and Vision Computing, 27(12):1743–1759, 2009.

H.G. Wallbott. Bodily expression of emotion. Eur. J. Soc. Psychol, 28:879–896, 1998.

3.4 Kinematic measures: The case of conductor-orchestra dialogue

Music orchestras represent a particularly interesting instance of sensorimotor coordination between several players and a conductor. However, a rigorous testing of inter-individual coordination in an ecological scenario poses a series of technical problems. In fact, rigorous laboratory research typically has three methodological limitations: i. Actions are simple, short, and stereotyped; ii. No real communication is implemented among participants; iii. Both sides of the interaction have to be parameterized to measure the behavioral response of the other agent. Although these limitations have been regarded as necessary given the large space of parameters, the use of artificial actions performed in unrealistic communicative scenarios actually burdens the generalization power of results.

Here we aim to use a rather different approach by studying music orchestras (a violins section and a conductor) in an ecological rehearsal scenario thus excerpting no particular interference on participant’s behavior. Here, we will record violinists' bows and conductor's baton kinematics via an unobtrusive passive infrared optical system. The rationale is that movement kinematics of one individual must have some statistical relation with the kinematics generated by another individual, to let us infer coordination between them. We will search for directed influences, and modulation thereof, among actions of the participants without imposing any artificial constraint.

Directed influences between participants were computed by using the Granger Causality (GC) method (Granger, 1969; Geweke, 1982). Since the typical concept of causality is that causes always precede effects, all events taking place at a certain point in time must have had their cause at an earlier stage. Granger’s proposal is that if a time-series y causes (or has an influence on) x, then knowledge of y should help predict future values of x. Thus, causality (or directed influence) is framed in terms of predictability of one's behavior given the past behavior of another participant.

3.4.1 Data analyses and computational tools

A first data preprocessing will be used to handle the missing data in the 3D trajectories (spline method; Deboor, 1978). The spline method interpolates the data with continuous third order derivatives. We will compute the magnitude of the acceleration from each 3D trajectory. There is a two-fold motivation behind this choice. First, Granger-causality requires the time-series to be covariance-stationary. Trajectories are often non-stationary so we will differentiate the signal (as it is common practice) to obtain a stationary signal. Second, we believe that, in terms of transfer of information (concerning musical expressiveness) between conductor and violinists and among violinists, accelerations of bows and baton are more informative than their trajectories and velocities.

Finally the (magnitude of) acceleration time-series will be demeaned, detrended, normalized (to z-scores) and windowed (overlapping windows=. Granger causality inference will be carried out at each window. In order to assess whether the window length affect the causal relations inference (see Seth, 2010).

Granger causality, in its standard and linear formulation, is based on (linear) Autoregressive Models (AM). AR models belong to the family of the Linear Dynamical Systems, which has been extensively used in modeling human motion (Del Vecchio et al., 2003; Lu and Ferrier, 2004; Bissacco, 2005; Bissacco and Soatto, 2006). An AR(k) model of a time series y is defined as:

$$\textbf{(1)} \quad x(t) = \sum_{j=1}^{l} a_j x(t-j) + \varepsilon_R(t)$$

where x(t) is the value of the time series y at time t, l is the order of the model (i.e., the length of the history observed in the model), ai (i = 1..l) are the weights for the history (the model parameters), and is the residual (prediction error).

There are two widely used criteria for selecting the optimal order of a linear predictor (i.e., the order that guarantees the best goodness of fit of the model): the Akaike's Final Prediction Error Criterion (AIC) (Akaike, 1974) and the Schwarz's Bayesian Criterion (BIC) (Schwartz, 1978). The ai parameters can be computed by using Ordinary Least Squares.

Since Granger causality is based on AR models the validity of the inferred causal relations depends on the validity of the AR models (more specifically of the unrestricted AR models, see next section). To assess the validity of an AR model different test can be carried out, ranging from tests of the non-correlation of the residuals to tests of the goodness-of-fit of the model (for example, the goodness-of-fit can be measured as sum of squares of the residuals).

A time series X is said to "Granger cause" a time series Y, if the past values of X provide statistically significant information to predict the next value of Y (Granger, 1969). The prediction is computed using AR models. Two AR models are required: an unrestricted AR model where the history of all time series is assumed to contribute to the prediction of the current value of a time series; and a restricted AR model where the time series of which the causality values (on the other time series) is computed is excluded from the history. Given two time series X and Y, the unrestricted model is defined as:

(2)

$$x(t) = \sum_{j=1}^{l} a_{U,j} x(t-j) + \sum_{j=1}^{l} b_{U,j} y(t-j) + \varepsilon_U(t)$$

$$y(t) = \sum_{j=1}^{l} c_{U,j} x(t-j) + \sum_{j=1}^{l} d_{U,j} y(t-j) + \eta_U(t)$$

While the restricted model is defined as:

(3)
$$x(t) = \sum_{j=1}^{l} a_{R,j} x(t-j) + \varepsilon_R(t)$$

$$y(t) = \sum_{j=1}^{l} d_{R,j} y(t-j) + \eta_R(t)$$

Then the magnitude of the causality from X to Y and from Y to X can be measured respectively as:

(4)
$$\mathcal{F}_{x \to y} = \ln \frac{\mathrm{H}_R}{\mathrm{H}_U}$$

$$\mathcal{F}_{y \to x} = \ln \frac{\mathrm{E}_R}{\mathrm{E}_U}$$

where E and H are the model error variances:

(5)
$$\mathrm{E}_R = var(\varepsilon_R(t)),\ \mathrm{E}_U = var(\varepsilon_U(t)),\ \mathrm{H}_R = var(\eta_R(t)),\ \mathrm{H}_U = var(\eta_U(t))$$

When the interaction of more than two time series is addressed, repeated pair-wise Granger causality computations can lead to misleading results. To avoid that, a simple extension of Granger causality, sometimes referred to as Conditional Granger causality, has been proposed by Ding and colleagues (Ding et al., 2006). Suppose we have three time series X, Y and Z, then the Conditional Granger causality from Y to X given Z is defined as the log ratio of the error variance of the restricted model where only Y is excluded from the history (when modeling X) and the variance of the unrestricted model, where the history of all time series X, Y and Z is included.

Once the Granger causality values have been computed, we need to test their statistical significance, i.e., we need to infer the significant causal relations. A significance test can be done by carrying out a F-test of the null hypothesis that the model parameters referring to the time series of which we compute the “causal strength” (on the other time series) are all zero (e.g., parameters $b_{U,j}$ in model (2) to test the significance of $\mathcal{F}_{y \to x}$). When more than two time series are analyzed some corrections (e.g., the Bonferroni correction) will be applied to the F-test.

The Granger causality analysis, including AR models validation and statistical tests of the causal interactions, will be carried by using the “Granger Causality Connectivity Analysis” MatLab toolbox (Seth 2010).

3.4.1.1 Non-linear Granger causality

Granger causal relations might be erroneously inferred (or ignored) when the linearity assumptions of the linear AR models are wrong, i.e., when there are significant non-linear interdependencies between the observed times series. Several solutions to extend Granger causality to the non-linear case have been proposed (see Freiwald et al., 1999 and Chen et al., 2004 for example). We will implement the non-linear Granger causality method proposed by Ancona et al., 2004 and based on kernel-based AR models where kernels are radial basis functions.

3.4.2 References

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Autom. Control 19, 716-723.

Ancona, N., Marinazzo, D. and Stramaglia, S., 2004 Radial basis function approaches to nonlinear Granger causality of time series. Physical Review E 70, 056221.

Bissacco, A., 2005. Modeling and learning contact dynamics in human motion. In CVPR '05, pp 421–428, Washington, DC, USA, 2005. IEEE Computer Society.

Bissacco, A. Soatto, S., 2006. Classifying human dynamics without contact forces. In CVPR '06, pages 1678–1685,Washington, DC, USA, 2006. IEEE Computer Society.

Chen, Y.H., Rangarajan, G., Feng, J.F., Ding, M.Z., 2004. Analyzing multiple nonlinear time series with extended Granger causality. Phys. Lett. A 324, 26–35.

Deboor, C., 1978. A Practical Guide to Splines. Springer-Verlag Berlin and Heidelberg GmbH & Co. K, December 1978.

Del Vecchio, D. Murray, R. M. and Perona, P., 2003. Decomposition of human motion into dynamics based primitives with application to drawing tasks. Automatica, 39:2085–2098, 2003.

Ding, M., Chen, Y., and Bressler, S.L., 2006. Granger causality: Basic theory and application to neuroscience. In Schelter. S., Winterhalder, N., & Timmer, J. Handbook of Time Series Analysis. Wiley, Wienheim.

Freiwald, W.A., Valdes, P., Bosch, J., Biscay, R., Jimenez, J.C., Rodriguez, L.M., Rodriguez, V., Kreiter, A.K., Singer, W., 1999. Testing non-linearity and directedness of interactions between neural groups in the macaque inferotemporal cortex. J. Neurosci. Methods 94, 105–119

Geweke J, Measurement of linear dependence and feedback between multiple time series. J Am Stat Ass. 1982;77, 304-313.

Granger CWJ, Investigating causal relations by econometric models and cross-spectral methods. Econometrica. 1969;37, 424-438.

Lu, C. and Ferrier, N.J., 2004. Repetitive motion analysis: Segmentation and event classification. IEEE Trans. Pattern Anal. Machine Intell., 26(2):258–263, 2004.

Schwartz, G., 1978. Estimating the dimension of a model. The Annals of Statistics 5, 461-464.

Seth, A.K., 2010. A MATLAB toolbox for Granger causal connectivity analysis. Journal of Neuroscience Methods. 186:262-273

3.5 Audio measures

3.5.1 Methodological issues and techniques for the acquisition of auditory measures

The current section presents an overview of the most crucial of those methodological issues and techniques potentially involved in sound acquisition and processing to be carried out around both SIEMPRE experiments and related recordings. Along with a contextual description of the main challenges involved, we provide a short discussion on the foreseen technical difficulties and possible workarounds, along with initial directives on needed techniques.

The two main topics introduced here are audio acquisition (sound capturing) and audio processing. These, while being strongly related, will be treated in two different parts. While in the acquisition part the focus will be put on the interesting signals (sound channels), on their relevance to the objectives of the project, and on the best available means for accessing them; the second part will deal with the necessary sound processing techniques involved both in the data acquisition part (also in relation to indirect acquisition of instrumental gestures) or in the analysis stage (sound descriptors and automatic segmentation).

Regarding the specific implications that audio acquisition and processing have on SIEMPRE project experiments and data analysis, two main aspects are considered here: (1) the use of lower-level audio features (combined with acquired instrumental gesture parameters) for the segmentation (score-performance alignment) of the recorded performances, and (2) the extraction of higher-level audio features (based on those low-level features) suitable to be used as a source for the analysis of synchronization processes (e.g. dynamics, articulation descriptors).

3.5.1.1 Acquisition of Auditory Signals

The acquisition of audio signals from a music performance scenario should be considered an important and relevant problem in the context of this project, and the reason behind that is the strong importance of sound as among the main (if not the most important) channels for communication taking place. Both in the case of inter-performer interaction and in the case of performer-audience interaction, sound carries an important role in driving music making and music perception tasks taking place in music performance real contexts, having the sound produced by the instruments and the sound perceived by both the performers and the audience as three principal sonic entities.

In the context of SIEMPRE, sound acquisition represents a difficult issue, because of the constraints imposed by real performance conditions. Those constraints, derived from the intrusiveness caused by the experimental setup (audio acquisition), are indeed limiting both the amount and quality of relevant sound signals that are available to be acquired.

As mentioned before, there exist three main groups of sound relevant signals. Next we proceed to provide an outline of important considerations and affordable sources for acquiring sound in each one of the three cases, also highlighting the corresponding restrictions.

Sound produced by musical instruments

The sound produced by the musical instruments (in this case bowed strings) represents the most important group of sound channels that provide sources for communication during performance. Three main methods, basically defined by attending to the placement of the acquisition device and the nature of the signal to be captured, will be considered for acquiring instrumental sound.

String vibration

METHOD

The string vibration signal, commonly acquired by means of piezoelectric transducers embedded in the bridge of the instruments, constitutes the best available source for the study of individual instrumental sound. The reasons behind this fact reside on the absence of mixed information coming from the rest of the instruments, and also on the absence of the reverberation and resonant characteristics of the body of the instrument. These result into a cleaner sound signal that is more suitable to the application of spectral-domain sound processing techniques for the extraction of reliable timbre related descriptors. As an alternative to the use of piezoelectric bridge pickups, contact microphones (or electret) close to the bridge may also provide the means for acquiring a high quality signal for later use in audio analysis.

ISSUES

The most important issue is the instrument-intrusiveness resulting from replacing the bridge of a musical instrument, or even from attaching a microphone under the bridge. Form our previous experiences, a significant number of musicians tend to refuse any modification of their musical instruments, even if that does not result in any intrusiveness related to their performance. In the cases of carrying out experiments with high-level string quartets and orchestra sections who refuse the use of any of these possibilities, an alternative being currently considered is to program pre-recording sessions in which a short set of exercises are performed and recorded using both contact microphones and close-field microphones, in order to have data for constructing timbre models that could enable the extraction of relevant sound descriptors from close-field radiation signals acquired using less intrusive devices during performance.

UTILITY

Apart from the clear use of individual audio descriptors for the analysis and extraction of correlation measures between performers, it will result highly useful to apply automatic score-performance alignment techniques (combining audio descriptors and instrumental gesture parameter signals) in order to obtain a note-level transcription of the performances which is to provide the necessary ground for individual, higher-level analysis like individual beat tracking or tempo estimation, or any detailed synchronization study.

Individual close-field radiation

METHOD

In order to acquire the sound produced by each instrument, one of the alternatives that allow capturing further timbre properties provided by the body of the instruments is the use of directional microphones, one per instrument. When conveniently located and pointing to the right direction, it is expected that each will provide a signal from which a number of audio features can be extracted (without facing the problem of source separation) as mostly describing the corresponding instrument. One of the possibilities is to attach each to a music, in case the performers are using one.

ISSUES

As it happened with the string vibration signal, intrusiveness may result to be an important issue depending on the context. In this case, the most relevant factor would merely be aesthetic: during recordings taking place in a real concert of high level quartets and orchestra sections, the presence of microphones may be refused by the performers. Another issue is the quality of the acquired signals, which may impede the extraction of individualized sound descriptors: on one hand, the presence of reverberation both from the bodies of the instruments and from the concert hall will make more difficult to reliably estimate onset times or extract accurate timbre descriptions; in the cases in which acquired signals contain an equal amount of information from the other sources the application of source separation techniques would be needed, but we consider that to be out of the scope of the project.

UTILITY

The use of these signals (once the maximally individualized sound descriptors have been obtained from them) would also be (as in the case of the string vibration signals) the use of individual audio descriptors for the analysis and extraction of correlation measures between performers, and therefore be able to apply automatic score-performance alignment techniques (combining audio descriptors and instrumental gesture parameter signals) for obtaining a note-level transcription of the performances which is to provide the necessary ground for individual, higher-level analysis like individual beat tracking or tempo estimation, or any detailed synchronization study.

Joint ensemble sound

METHOD

The acquisition of ensemble sound is to be achieved by means of ambient (non directional) microphones. On one hand, they are to be located at the physical center the ensemble under study at a height from the floor equivalent to the performer’s heads. On the other hand, non directional microphones are to be located between the ensemble and the audience (when recordings are carried out in real concerts), at a distance from the floor to that will depend on the stage/venue particular conditions.

ISSUES

This audio signal can be considered as a better representation of the complete musical message, especially from a perceptual point of view and also from the perspective of the audience. However, the necessity of analyze individual sources in order to extract information of the performance of each musician makes this approach as less promising for analyzing instrumental sound. Attempting to extract individual instrumental sources from one common sound signal would require to study and apply source separation techniques, and such area of research is considered as to be out of the scope of this project.

UTILITY

For the cases in which intrusiveness constraints impede to use one of the two previously mentioned methods, the acquired ensemble audio signal will at least be used to obtain overall intensity descriptors (based on spectral-domain analysis techniques) that could result complementary to other signals captured when studying the sound perceived by the audience or by the musicians.

Sound perceived by the musicians

During performance, for the cases to be considered in SIEMPRE, when considering the sound perceived by the musicians as one of the groups of sonic entities, we neglect any sound produced by the audience as to be of relevant importance to the performance. Given this restriction, it is going to be assumed that the sound arriving to the ears of the musicians is only generated by the musical instruments.

Depending on the performance conditions (e.g. acoustic characteristics of the venue), the sound perceived by the musicians will present certain differences with respect to the sound produced by the musical instruments. Such differences, which are to be assumed quasistationary and only varying with changes in the orientation of the instruments and in the musician's head rotation, will strongly depend on the acoustic characteristics of the venue, and the positions of the members of the ensemble with respect to each other. It appears as an interesting pursuit to extract or infer measures that result relevant in the study of those constraints.

In an attempt to capture the sound perceived by the ensemble musicians during performance by considering binaural-oriented perception models to be of relevance to the synchronization and entrainment processes taking place in ensemble performance, an ideal situation would be the one in which each musicians is wearing binaural microphones, so that two sound streams are acquired for each performer. From the audio signal captured by each pair of microphones, specific audio analysis techniques would be applied by attending to measures of instrument and head orientations, and also to the reverberation characteristics of the space.

The ideal scenario described above, however, results impractical in certain situations (especially in real concerts with high –level quartets or orchestras) given certain intrusiveness and also due to aesthetic reasons. In those cases, apart from ignoring time variations of the auditory differences mentioned above, the first and most promising approach is to program a series of pre-recordings ahead of the performance time and study those differences in off-line conditions. During those pre-recording sessions, the aim would be to acquire a number of relevant signals from which those variations can be parameterized. Obtained parametric models would then be used, as a post-processing step, in order to obtain a reliable estimation of the perceived sound by only using the acquired sound streams captured from the musical instruments (using bridge pickups/microphones or directional microphones, if possible as previously described).

Sound perceived by the audience

As it happens for the case of sound perceived by musicians, when considering the sound perceived by the audience as one of the groups of sonic entities we neglect any sound produced by the audience as to be of relevant importance to the performance. Given this restriction, it is going to be assumed that the sound arriving to the ears of the audience is also generated solely by the musical instruments.

The acquisition of sound signals from within the audience should be carried out using, again, non-directional, ambient (cardioids) microphones conveniently located in different positions of the audience. The aim of acquiring sound at different positions resides on the (potentially) significant differences in arrival times (as compared to the visual channels) and on the reverberant characteristics of the concert hall. These auditory measures, together with measures of other nature to be acquired from the audience in SIEMPRE recordings, will hopefully help to the relation between the characteristics of the space, the spatial location of the audience members, and their emotional response to the performance.

Overview of devices and uses for acquiring auditory measures

We have compiled in Table 1 the four main types of sound capturing devices together with their expected placement during the SIEMPRE experiments, the conditions in which they will be used (recording sessions versus off-line pre-recording experiments), and the aims of their use.

As a concluding remark related to (i) development cost and (ii) chances of success when applying the needed audio processing techniques, we foresee that most of the relevant auditory measures within the SIEMPRE prospects are expected to be the result from processing individual audio stream for each instrument.

Table 1. Overview of devices and uses for acquiring auditory measures

Device	Where	When	Aim
Ambient (cardioid) microphones	All around (stage / audience)	All recording conditions	Capture overall and local sound streams from which to extract basic audio features
Directional microphones	Stage (different positions)	Upon agreement with musicians (both during pre-recording sessions and during performance)	Acquire individual instrumental sound for extracting refined audio features
Bridge pick-ups (microphone / transducers) or contact microphones	Instrument bodies (bridge)	Upon agreement with musicians (both during pre-recording sessions and during performance)	Acquire individual instrumental sound (string vibration signal)
Binaural microphones	Performers	Only during pre-recording sessions	Capture sound perceived by musicians

3.5.1.2 Audio Analysis Techniques

The topic of audio analysis is of clear relevance to the objectives of SIEMPRE project. For the case of synchronization between performers, it appears as strongly limited by the quality of audio signals acquired from performance recordings. We have seen in the previous section that obtaining clearly separated audio signals may result unfeasible in certain conditions. While the ideal situation of having nearly-independent sound signals for all musicians is assumed when defining the methods for audio processing and their use in the extraction of high level performance –related features to be used in synchronization analysis, non-ideal scenarios have been considered when devising an audio analysis roadmap. However, given that the highly challenging, still unsolved problem of source separation still falls out of the scope of the project, the SIEMPRE Consortium should devote effort to facilitate the acquisition of independent audio signals for each performer.

Audio feature extraction

The first step will consist on the extraction of a series of audio descriptors from the acquired audio signals. The extraction of this first set of descriptors will in most of cases be applied to all acquired sound streams, although for the case of some higher level descriptors, they will be specific to a subset of them.

The extraction of descriptors will be performed in off-line conditions, as a post-processing step, once the sound streams have been conveniently uploaded to the SIEMPRE repository (currently under development). Once the most commonly needed and used set of audio features has been decided, the corresponding algorithms will be implemented and offered as web-services to which the users of the repository will have access therein. Those web-services for audio feature extraction will return a new signal or set of signals with the computed descriptors, and those signals will be automatically added to the repository so that any user can have access to them.

Lower-level descriptors

All lower-level descriptors will be computed in a frame-by-frame fashion. At the same time, the majority of such descriptors will be computed in the spectral domain. It is expected that the set of low-level descriptors will constitute a basis for subsequent analysis, both only based on audio, and combining different sources of information. Next we provide a preliminary list of the first set of low-level audio descriptors to be extracted [Peeters, 2004].

- **Bark-scale spectral envelope descriptors**
  - spectral band amplitudes
  - spectral skewness
  - spectral kurtosis
  - spectral spread
- **Linear-scale spectral envelope descriptors**
  - spectral band amplitudes
  - spectral skewness
  - spectral kurtosis
  - spectral spread
  - spectral centroid
  - spectral complexity
  - spectral decrease
  - spectral energy
  - spectral flatness
  - spectral crest
- **Mel Frequency Cepstrum Coefficients** (MFCCs)
- **Dissonance**
- **High-frequency content**
- **Zero-crossing rate** (time domain)
- **Silence-rate** (time domain)

Higher-level descriptors

For the extraction of higher-level descriptors from audio features, it is assumed that (i) the score of the piece to be performed is known, and (ii) the musicians respected the score to a significant extent. The extraction of higher-level descriptors will be carried out by combining time-domain and spectral-domain approaches. Higher-level descriptors will be obtained either from individual sound sources (assuming they have been acquired separately) or from joint ensemble sound, depending on the application.

- **Fundamental frequency**

  This set of descriptors will be a strong basis for the analysis of intonation synchronization, vibrato extraction and parameterization, vibrato synchronization, and individual score-performance alignment [Maher, 1994] [de Cheveigne, 2002].

  - pitch estimation
  - pitch estimation confidence
  - pitch salience

- **Beat/tempo tracking**

  The higher level descriptors referred to tracking beats will be of relevant importance for the estimation of local and global tempo, timing synronization, and individual score-performance alignment [Dixon, 2001] [Grosche, 2010].

  - predominant tempo candidates
  - predominant tempo candidate confidences
  - local tempo candidates
  - local tempo candidate confidences
  - beat times

- **Tonal analysis** (alternative method)

  The extraction of tonal descriptors from audio signals might be used to pursue the problem of joint ensemble score-performance alignment when only ambient microphones have been used to acquire the sound produced by the musical instruments as a group [Gomez, 2006].

- **Predominant pitch estimation | Basic techniques for multi-pitch estimation** (alternative method)

  For the cases in which only ambient microphones have been used to acquire the sound produced by the musical instruments, or in the cases for which the audio signals acquired using directional microphones when signals have been acquired, the use of these techniques may help in separating audio sources, although they fall out of the scope of the project [Klapuri, 2003].

Reverberation

As contrasted in the past with professional ensemble members, it is hypothesized here that the acoustic properties of the space in which an ensemble is performing will affect the synchronization between them. A simplified measure of reverberation is needed in order to extract relationships between the space and synchronization capabilities. Here the commonly known t60 value is chosen as a standard measure used for the characterization of the reverberation properties of the venue, which in turn will affect the synchronization phenomena in ensemble playing.

A thorough study with different concert halls and performance spaces appears as a difficult task. Because of that, a preliminary study is planned first with a violin duet using silent electric violins, and applying different real-time artificial reverberation effects to their bridge pickup signals in order to study how the reverberation time (*t60*) or other characteristics of the space are indeed affecting a number of synchronization measures.

Audio analysis in relation to instrumental gesture / performance controls

Specific audio analysis techniques, mostly based on spectral-domain processing, will be used for constructing non-linear mapping models between sound descriptors(spectral envelope parameterizations) and instrumental gesture features [Perez, 2009].

A clear application of such mapping models will be trying to overcome the very restrictive limitation dictated by the impossibility of installing motion capture sensors in the instruments during performance: the mapping models will be built from data gathered during off-line pre-recording sessions, and then be applied for indirectly inferring instrumental gesture parameters from the audio later captured from the real performance. The inferred instrumental gesture parameters or features will contribute to crucial steps like score-performance alignment, or to the extraction of high level performance features (dynamics, articulation, etc.).

Score-performance alignment

The problem of score-performance analysis is of crucial relevance to the SIEMPRE project. Failing at providing automatic or semi-automatic tools for facilitating the segmentation of the performances into notes as related to the nominal scores would imply devoting much effort to carry out those tasks manually. We have considered two main scenarios: using only audio information (features), or combining audio information and instrumental gesture parameters.

Audio-based

Assuming that an individual audio stream has been acquired for each of the instruments, techniques inspired on dynamic programming are to be applied [Cano, 1999].

When only joint ensemble sound is acquired, a common score-performance alignment might be pursued by also applying dynamic programming, but in this case using tonal information extracted from the ensemble audio.

Combining audio and instrumental gestures

The most promising approach to obtaining a reliable tool for score-performance alignment is the combined use of audio features and instrumental gestures. The only question is whether instrumental gestures will be acquired in a 'direct' manner using motion capture systems (e.g.

Qualysis) or will be acquired in an ‘indirect; manner from audio features. In any of the two cases, dynamic programming appears as the most affordable technique to use [Maestre, 2009] [Perez, 2009].

3.5.2 References

Peeters, G. (2004). A large set of audio features for sound description (similarity and classification) in the CUIDADO project. *CUIDADO IST Project Report*, *54*, 1–25. Retrieved from http://www.citeulike.org/user/ajylha/article/1562527.

Dixon, S. (2001). Automatic Extraction of Tempo and Beat From Expressive Performances. *Journal of New Music Research*, *30*(1), 39-58. Routledge. doi: 10.1076/jnmr.30.1.39.7119.

Maher, R. C., & Beauchamp, J. W. (1994). Fundamental frequency estimation of musical signals using a two-way mismatch procedure. *Journal of the Acoustical Society of America*, *95*(4), 2254

De Cheveigné, A., & Kawahara, H. (2002). YIN, a fundamental frequency estimator for speech and music. *Journal of the Acoustical Society of America*, *111*(4), 1917. Ensemble hand-clapping experiments under the influence of delay and various acoustic environments

Farner, S., Solvang, A., Sæbø, A., & Svensson, P. (2006) Ensemble hand-clapping experiments under the influence of delay and various acoustic environments. Audio Engineering Society 121st Convention, San Franciso, USA. http://www.pvv.ntnu.no/~farner/pub/pdf/farner06c-aes.pdf

Cano, P., Loscos, A., & Bonada, J. (1999). Score-Performance Matching using HMMs. *In Proceedings of the International Computer Music Conference* (Vol. 1, pp. 441-444).

Klapuri, A. P. (2003). Multiple fundamental frequency estimation based on harmonicity and spectral smoothness. *IEEE Transactions On Speech And Audio Processing*, *11*(6), 804-816.

Müller, M., Konz, V., Scharfstein, A., Ewert, S., & Clausen, M. (2009). Towards Automated Extraction of Tempo Parameters from Expressive Music Recordings. *International Conference on Music Information Retrieval* (pp. 69-74).

Gómez, E. (2006). Tonal description of polyphonic audio for music content processing. *INFORMS Journal on Computing*, *18*(3), 294-304.

Grosche, P., & Mueller, M. (2010). Extracting Predominant Local Pulse Information from Music Recordings. *Audio Speech and Language Processing IEEE Transactions on*, (99), 1–1.

Cano, P., Loscos, A., & Bonada, J. (1999). Score-Performance Matching using HMMs. *In Proceedings of the International Computer Music Conference* (Vol. 1, pp. 441-444).

Maestre, E. (2009). Modeling instrumental gestures: an analysis/synthesis framework for violin bowing. PhD thesis, Universitat Pompeu Fabra.

Perez, A. (2009). *Enhancing Spectral Synthesis Techniques with Performance Gestures using the Violin as a Case Study*. PhD thesis, Universitat Pompeu Fabra.

3.6 Measurement of physiological signals

3.6.1 Physiological signals and their use in augmenting emotion recognition for human-machine interaction

This section introduces some of the concepts and motivations in choosing specific physiological signals for the SIEMPRE project. The primary use of these signals will be as an indicator of the emotional state of performers and the synchronization of that emotional state with other performers and, most revolutionarily, the audience. Much of the background material below was recently published and has been directly excerpted from [Knapp11a]. That reflects the fact that the measurements used in SIEMPRE are not chosen at random. Rather they draw on sustained research in the area, built up through a sequence of earlier studies, and reported in relevant publications.

3.6.1.1 Introduction

The three indicators of emotional state currently used in emotion research have been "evaluative reports, overt actions, and physiological responses" [Bradley07]. Informed by many diverse fields of science including neuro-physiology, psychophysiology, and human-computer interaction (HCI), SIEMPRE will explore incorporating the use of physiological signals in emotion recognition and the use of this information as part of studying the relationship between and among performers and between performers and the audience.

Emotions are partly neuro-physiological processes (e.g., [Cacioppo 1999]), and attempting to decipher emotional state in the absence of direct measures of physiological changes is indeed ignoring a wealth of relevant and sometimes vital information. In spite of this, the focus of much of the recent research on emotion has been on facial, gestural, and speech recognition. One argument for this research bias is based on the incorrect assumption that humans cannot detect physiological changes unless they are revealed by overt actions. However, as has been stated many times, "it is a mistake to think of physiology as something that people do not naturally recognize. A stranger shaking your hand can feel its clamminess (related to skin conductivity); a friend leaning next to you may sense your heart pounding; students can hear changes in a professor's respiration that give clues to stress; ultimately, it is muscle tension in the face that gives rise to facial expressions. [Picard01]" It should also be noted that revealing changes that are not detectable by human contact indeed might be the most important contribution of physiological measures to emotion recognition.

3.6.1.2 A Very Brief History of Physiological Measurement and Emotion

History is worth knowing in the field of emotion and physiology, because it is an area where popular conceptions of science are deeply attached to old ideas. William James captured the public imagination when he claimed (in 1884) that the essence of emotion was awareness of visceral changes that occur in response to extreme situations (such as meeting a bear in the woods) [Cannon27]. If that were so, then measuring the visceral changes directly would allow artificial systems to detect a person's emotions as well as the person him- or herself – or possibly better. In reality, even contemporary experts who are sympathetic to James accept that his idea captures only part of the truth. Injuries that prevent visceral feedback do not nullify emotion; visceral states are not as sharply distinguished as emotions are; experienced emotion can be changed by manipulating cognitive state, but not visceral states; and so on.

Not long after James released his idea, prominent figures (including Jung) took up the idea that skin conductivity could reveal otherwise hidden psychological events. Lie detection quickly became a high-profile application for it and related techniques. The idea captured the public imagination to the extent that employers were held liable for failing to use lie detection before hiring employees who went on to commit crimes. However, by 1959, it was becoming clear that standard forms of lie detection had questionable scientific validity [Lykken98]. It was not the physiological measures that were the focus of the uncertainty, but the methodology of the protocol and specifically the questions that were being asked. In studying emotion and, most importantly, physiology and emotion, it was becoming clear that context and induction techniques were critical.

Over the ensuing decades, psychophysiologists have built up an enormous body of evidence on relationships between physiological measures and emotion. Two good summaries are in [Caciopo00] and [Bradley07]. This research has shown that if other variables are meticulously controlled, there are associations between emotional states and physiological variables. The phrase 'meticulous control' is key. Physiological signals are subject to multiple influences. They are affected not only by emotion, but also by almost any kind of effort, mental or physical. Traditional experiments dealt with these problems by creating situations that prevented irrelevant variables from intruding. Even with meticulous control, recent literature surveys such as found in [Kreibig07] show that correlations between physiological changes and changes in emotional state are not always consistent.

3.6.1.3 Definitions of Emotion-Related Physiological Signals

For SIEMPRE, we must define the physiological signals we will be using and briefly explain how they are typically measured (more detail on the synchronous data acquisition of all eMAP signals will be be reported in deliverable 3.1). There are a great many textbooks and articles that define these signals and explore their usage in fields ranging from medicine and psychology to bioengineering and instrumentation. Two classic books are "The Handbook of Psychophysiology" [Cacciopo07] which defines each physiological signal and summarizes its history and usage within the broad field of psychophysiology and "Medical Instrumentation, Application and Design" [Webster98] a classic engineering handbook that reviews physiological signals in the context of the medical instrumentation needed to capture each signal.

Among the many methods for measuring changes in human physiology, this section will focus on briefly defining those physiological parameters or signals which are currently being used in the fields of psychophysiology and emotion *and* are measurable in a relatively unobtrusive way (and thus could be envisaged to be part of SIEMPRE). It should be emphasized that many physiological signals only recently have that meet these criteria due to advances in measurement technologies. Thus, one can expect that the introduction of new technologies will serve to expand this list over the coming years (or months!). A good review (although not entirely complete) of so-called ambulatory monitoring systems, systems that can be used in mobile environments and are relatively unobtrusive, can be found in [Ebner-Priemer07]. In defining these key physiological signals, we will categorize them into physiological signals that originate from the autonomic and somatic components of the peripheral nervous system and physiological signals that originate from the central nervous system (see Figure 1 for a taxonomy of the nervous system).

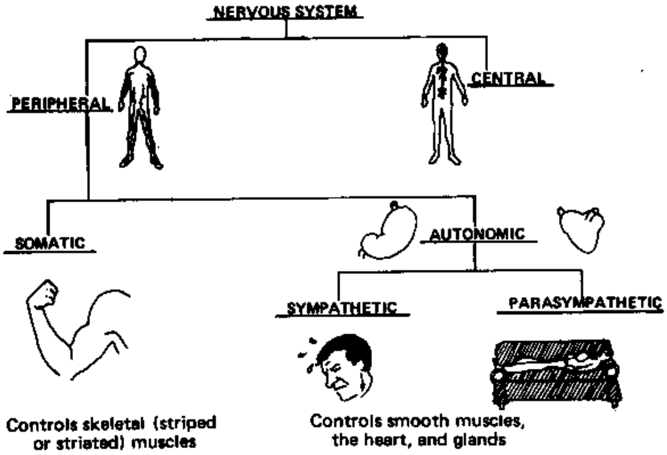

NERVOUS SYSTEM
PERIPHERAL
CENTRAL
SOMATIC
AUTONOMIC
SYMPATHETIC
PARASYMPATHETIC
Controls skeletal (striped or striated) muscles
Controls smooth muscles, the heart, and glands

Figure 1: Taxonomy of the Nervous System (from online.sfsu.edu/~psych200/unit5/u5m5.gif)

3.6.1.4 Physiological Data Acquisition

In collecting data from physiological sensors, the signals must ultimately be captured by a data acquisition system connected to a digital system ranging from an embedded processor to a super-computer. The choice of available data acquisition systems is vast. Some of the important factors in narrowing this range of choices include:

- Signal isolation: In order to reduce the chance of electrical shock, **the data acquisition system must electrically isolate the individual from "wall" or "mains" power as well as from the computer itself**. This can be achieved by using a battery-powered wireless data acquisition system or by using a system specifically designed for physiological data acquisition that already incorporates optical or magnetic isolation.
- Obtrusiveness: There is a considerable trade-off in the number and type of physiological signals chosen and the obtrusiveness of the measurement. At the limits either not enough data is being measured or the individual cannot participate in the interaction in any way that is not interfered with by the measurement equipment. While this is one of the most difficult choices in using physiological measures for emotion measurement, there is little to no literature on the impact of the ecology of the measurement on the emotional estimation "accuracy".

  It should be pointed out that while so-called "wireless" data acquisition systems are usually superior to wired systems in terms of their obtrusiveness, they might indeed include many wires connecting the physiological sensors to the wireless transmitter. There are very few commercially available systems that combine the sensor with the transmitter to eliminate all wires and are most commonly found in the consumer sports arena from companies such as Polar, Suunto, or Nike.
- Sampling rate and resolution: These will be based on the bandwidth and dynamic range of the physiological signal acquired. The range of sampling rates is considerably reduced for wireless data acquisition systems.
- Synchronization of multiple data streams: In many cases more than one channel of physiological data is analyzed simultaneously (or combined with video or audio streams). The data acquisition system – both hardware and software – must have the capability of synchronizing multiple streams of data with multiple sampling rates.
- Shielding and differential amplification: The amplitude of the voltage of physiological signals can be as small as a fraction of a μV. In order to avoid contamination from other signals, so-called "noise" signals, electrical shielding of the signal wires from sensor to

data acquisition system is extremely important. Also, data acquisition systems using a technique known as differential amplification, amplifying only the difference between two sensor signals, should be used if possible. By amplifying only the difference between two sensor signals, any noise that is common to both sensors will be considerably reduced.

Autonomic and Somatic Nervous System

The somatic component of the peripheral nervous system is concerned with sensing information that happens outside the body and is responsible for the voluntary control of our skeletal muscles to interact with this external environment. Signals that measure this voluntary control of the muscles are measuring aspects of the somatic nervous system. The autonomic (ANS) component of the peripheral nervous system is responsible for sensing what happens within the body and regulating involuntary responses including those of the heart and smooth muscles (muscles that control such things as the constriction of the blood vessels, the respiratory tract and the gastrointestinal tract). There are two components of the ANS, the parasympathetic and the sympathetic (see Figure 2).

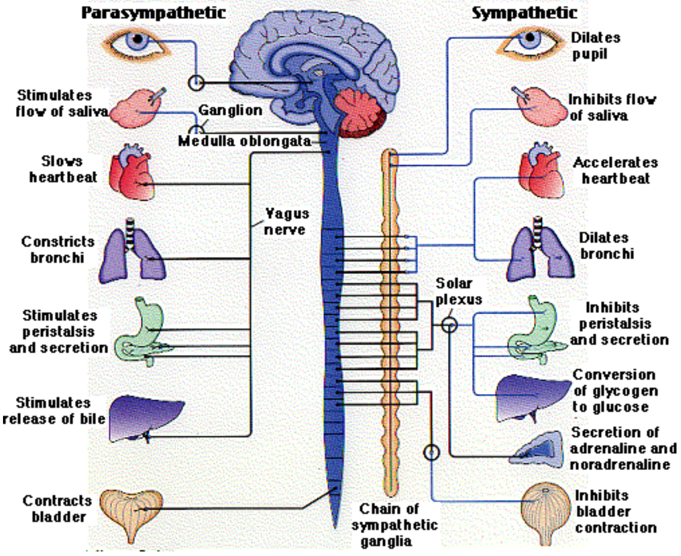

Parasympathetic
Sympathetic
Stimulates flow of saliva
Ganglion
Medulla oblongata
Slows heartbeat
Vagus nerve
Constricts bronchi
Stimulates peristalsis and secretion
Stimulates release of bile
Contracts bladder
Chain of sympathetic ganglia
Solar plexus
Dilates pupil
Inhibits flow of saliva
Accelerates heartbeat
Dilates bronchi
Inhibits peristalsis and secretion
Conversion of glycogen to glucose
Secretion of adrenaline and noradrenaline
Inhibits bladder contraction

Figure 2: Autonomic Nervous System (from http://www.wickersham.us/anne/images/autonomic.gif)

The parasympathetic component is responsible for slowing the heart rate and relaxing the smooth muscles. The sympathetic component of the ANS is responsible for the opposite, i.e., raising the heart rate and constricting the blood vessels which causes, amongst other effects, an increase in blood pressure. It is also responsible for changes in skin conductivity. The sympathetic response is slower and longer lasting than the parasympathetic response and is associated with the so-called, "flight or fight reaction". Physiological signals that measure the involuntary responses of the peripheral nervous system are measuring aspects of the autonomic nervous system.

Autonomic Nervous System

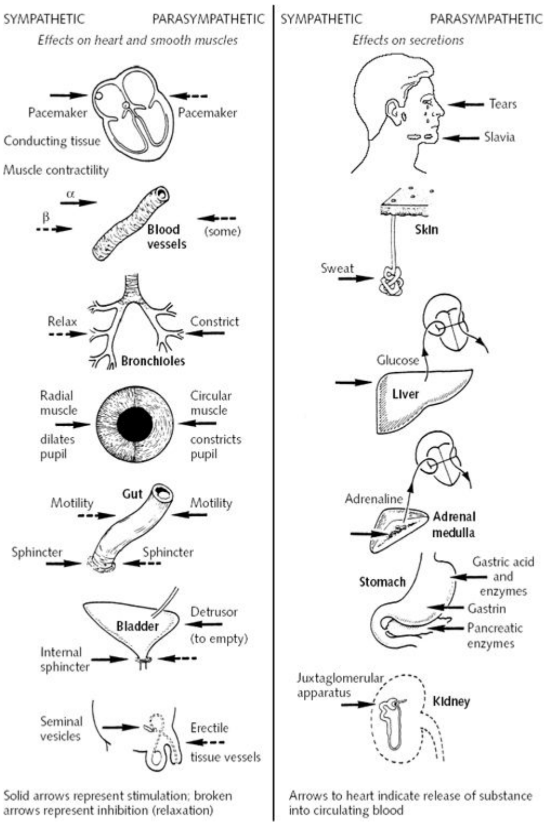


Figure 3: Autonomic Nervous System (After Jennett, S. (1989) Human Physiology)

Electrodermal Activity (EDA)

Measurement of the electrodermal activity or EDA is one of the most frequently used techniques to capture the affective state of users, especially for exploring attention and varying arousal in emotion. It is controlled by the sympathetic nervous system. EDA sensors (electrodes) measure the ability of the skin to conduct electricity. A small fixed voltage is applied to the skin through the electrodes and the skin's current conduction or resistance is measured (this is preferred over applying a current and measuring the voltage produced). The value of this conductivity is usually in the range of 2-20uS (500kOhm – 50kOhm). The skin conductivity consists of two separate components. There is a slow moving tonic component or skin conductance level (SCL) that indicates a general activity of the perspiratory glands from temperature or other influences and a faster phasic component or skin conductance response (SCR – also known as Galvanic Skin Resonse or GSR) that is influenced by emotions and the level of arousal. For example, when a subject is startled or experiences anxiety, there will be an increase in the skin conductance due to increased quantity of sweat in the sweat ducts of the glands.

Ecological Measurement of EDA

EDA is most significant on the palm of the hands and the bottom of the feet. The two most common measurement techniques are thus:

1. to place an electrode on each of two fingers (usually the thumb and index finger)
2. to place two electrodes across the palm

For the most accurate recordings, the electrodes consist of silver-silver chloride cup electrodes. However, in using the measurement of EDA as part of an HCI system, the EDA is commonly measured using standard conductive metal plates.

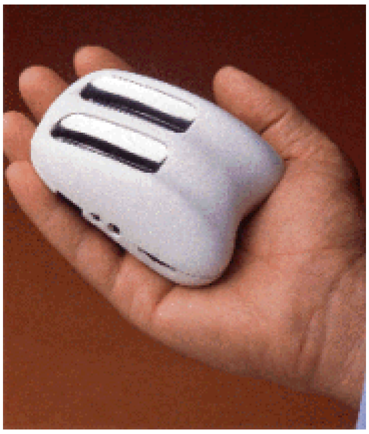


Figure 4: Mouse used to measure EDA (from Thought Technology)

Changes in physical activity, environmental conditions such as temperature and humidity, movement of the electrodes on the skin, and changes in pressure on the electrodes all serve to confound the measurement of EDA and must be mitigated. Physical activity and ambient conditions primarily affect the SCL and should be measured using other modalities (motion sensors, temperature and humidity sensors, etc.) in an attempt to limit the artefact.

Important Features of the EDA Signal

In addition to the amplitude of the SCR and SCL, temporal parameters of the SCR such as latency, rise and decay time are all important features of the EDA used in determining attention and emotional state.

Cardiovascular System

Heart Rate and Heart Rate Variability

Another important physiological correlate of emotion is the frequency or period of contraction of the heart muscle. As shown in Figure 3, the heart rate (HR) is controlled by both the parasympathetic response (decreasing HR) and sympathetic response (increasing HR). The heart rate or period can be derived from many measurement techniques as will be discussed below. The higher frequency changes (0.15 – 0.4 Hz) are heavily influenced by breathing (respiratory sinus arrhythmia (RSA)) especially with younger and more physically fit individuals. The lower frequency changes (0.05-0.15Hz) are not influenced by the RSA and can reveal other aspects of the ANS.

Ecological Measurement of Heart Rate and Heart Rate Variability

Electrocardiography (ECG):

Electrocardiography measures electrical changes associated with the muscular contraction of the heart. More specifically, the ECG results from the Sino-Atrial (SA) node and Atrio-Ventricular (AV) node of the heart electrically activating the first of two small heart chambers, the atria, and then the two larger heart chambers, the ventricles. Particularly, the contraction of the ventricles produces the specific waveform known as the QRS complex as shown in **Errore. L'origine riferimento non è stata trovata.**. The heart rate is most commonly derived from the ECG by measuring the time between R components of the QRS complex – the so-called R-R interval.

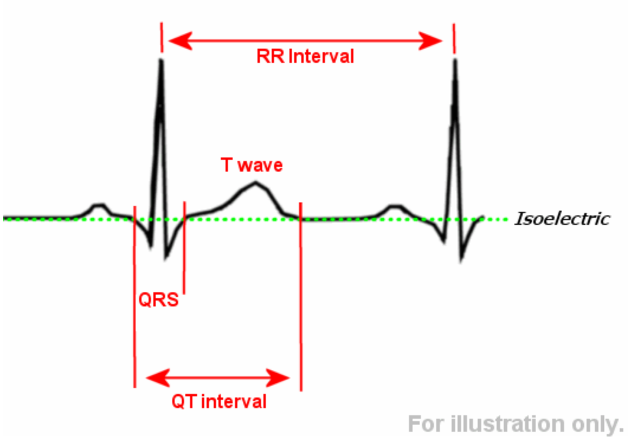


Figure 5: ECG Wave

The ECG is measured at the body surface across the axis of the heart. Electrodes are placed on the skin which transduce the electric field caused by the previously mentioned electrical activity of the heart to electron flow in the measurement lead. This signal is then high-pass filtered to remove long term offset and amplified by a factor of around 1000. In medicine, the ECG is most commonly measured using a standard 12 lead (electrode) configuration, however, when only the heart rate is being measured, this can be reduced as low as a 2 lead configuration as shown in Figure 5. Indeed, as is done with exercise equipment, heart rate measurements can even be recorded simply by measuring the voltage potential from hand to hand (although not nearly as accurate as chest leads).

In addition to measuring heart rate, one other aspect of the ECG wave that is important to emotion research is the amplitude of the T wave shown in **Errore. L'origine riferimento non è stata trovata.**. This is because the T wave amplitude has been found to be an indicator of activity of the sympathetic nervous system.

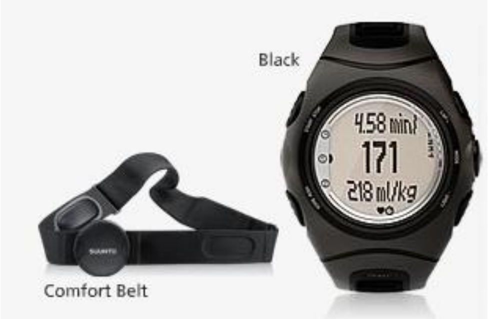


Figure 5: Watch with Wireless ECG-derived Heart Rate Monitor (www.suunto.com)

Photoplethysmography and Blood Volume Pulse (BVP)

An alternative to measuring heart rate directly from the electrical activity of the heart is to measure the pulsation of blood flow through the vascular system. The most common technique for achieving this is to shine light from an infra-red LED into the skin and measure either the amount of light transmission through the skin or the amount of light reflection from the skin (or both). As the heart beats, the perfusion of the blood vessels underneath the light source ebbs and flows and thus the absorption characteristics of the light pulses with the heart beat. As shown in Figure 6, the finger tip is the most common location for measurement of the BVP, although it can also be measured other places on the extremities of the arm feet or even earlobe. Photoplethysmography becomes quite inaccurate due to even minor motion of the body and is often coupled with the use of accelerometers to detect motion and to attempt to compensate (e.g., [Morris06]).

Figure 6: Photoplethysmography (from http://www.medis-de.com)

It is important to note that two other physiological parameters associated with changes in emotional state can be measured with photoplethysmography: peripheral blood perfusion (e.g [Kunzman05]) and blood oxygen saturation (SpO2) levels (e.g. [Karekla04]). The effect of the change in magnitude of the BVP because of changes in peripheral blood perfusion can also cause the measurement of HR and HRV to be somewhat less accurate due to missed beats.

Impedance plethysmography works similarly to photoplethysmography except that instead of measuring the change in blood perfusion by measuring the change in light reflectance/transmission, impedance plethysmography measures the change in the skin's capability to conduct electricity. This is most often measured by applying a small AC current across the chest and measuring the change in voltage. This technique has enabled the creation of clothing that can measure HR and HRV.

A new system developed in conjunction with Biocontol systems integrates both of these sensors on a single finger.

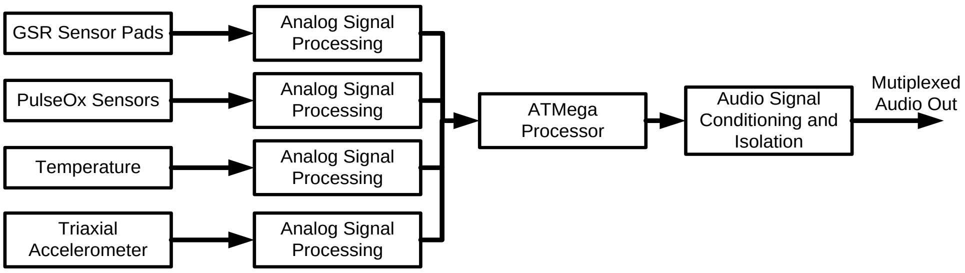


Figure 7: Block Diagram of the MobileMuse

As shown in Figure 7, four sensors are integrated into the MobileMuse. This was because, upon further consideration, and because of the ease of design, a temperature sensor was also added to the interface. Skin temperature change (in relationship to the environment) has been shown to be indicative of long term mood [Baumgartner06] and it was thought that this might prove beneficial in assessment of emotional state. A tri-axial accelerometer was also added to the circuit for artifact removal.

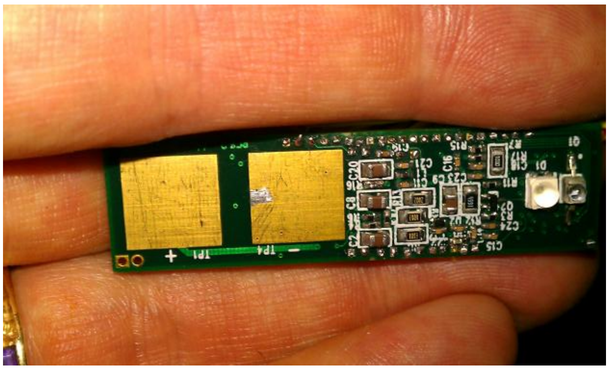


Figure 8: First implementation of the MobileMuse: The two large pads are for GSR measurement and the two LED's on the far right are for pulse oximetry.

As shown in the block diagram in Figure 7, all of the sensor signals are amplified, processed, and conditioned and then connected to an ATMega processor. Choosing this processor means that the MobileMuse can be used as a custom Arduino board with all of the advantages that creates - most importantly, ubiquitous software availability. The ATMega processor is used to frequency-division multiplex the sensor signals in order to create one single audio data stream. The signal is then re-converted to an analog stream using the pulse-width-modulation output of the processor and subsequent signal conditioning. Finally, magnetic isolation is used to remove any shock risks and to eliminate line noise.

Blood Pressure and Peripheral Perfusion

In addition to modulation of the heart rate, emotional changes can influence the contractility of the blood vessels. This change can affect blood pressure and perfusion as well as skin temperature.

Ecological Measurement of Blood Pressure and Peripheral Perfusion

As mentioned previously, peripheral perfusion can be quite ergonomically measured by photoplethysmography. Just as simply, skin temperature can be measured by placement of a thermistor on the skin surface (although this is confounded considerably by ambient temperature). Ecological measurement of blood pressure, however, has still not been achieved. This is because the most common (and most accurate) non-invasive techniques for measuring blood pressure require that a cuff is placed on the arm and is inflated to cut off arterial blood flow. The acoustical changes that appear as this flow is returning to normal (first the appearance of a heart rhythm (k) sound and then the disappearance of the k sound) as the cuff is deflated indicate the systolic and diastolic blood pressure, respectively. This inflation and deflation of the cuff make it difficult to measure continuous blood pleasure changes. Although expensive and large, one of the few ambulatory devices available is the Portapress monitor shown in Figure 9.

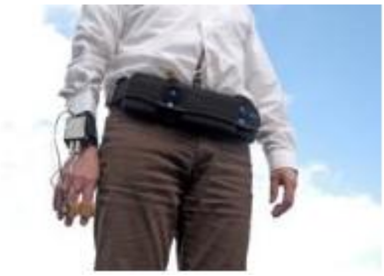


Figure 9: Portapress Ambulatory Blood Pressure Monitor (www.finepress.com)

It is argued that pulse wave velocity, the measurement of the time it takes for a pulse to move down the arm, is proportional to changes in blood pressure. See [Harata06] for a review. The pulse arrival time can be measured at two locations using photoplethysmography and then the pulse transit time (and thus pulse wave velocity) can be calculated. An example of a current attempt at using the ecological measurement of pulse wave velocity for emotion tracking can be seen in Figure 10. There is relatively general agreement that this measurement cannot be used for measuring the absolute value of blood pressure due to the affects of the contractility of the arm’s vascular system.

Figure 10: Pulse Wave Velocity Measurement of Blood Pressure (www.exmocare.com)

Respiratory System

Respiration Rate and Depth

The respiratory system is one of several systems of the body that are under both autonomic and voluntary (somatic nervous system) control. Thus respiration rate and depth as an indicator of emotional state must be used with caution and always viewed in context. In controlled environments, it has been found that variation of respiration rate generally decreases with relaxation. Startle events and tense situations may result in momentary respiration cessation and negative emotions generally cause irregularities in the respiration pattern.

Ecological Measurement of Respiration Rate and Depth

The respiration signal (breathing rate and intensity) is commonly acquired by using a strain gauge or piezo sensor embedded in an elastic band worn around the chest. The sensor measures the expansion and contraction of the band, which is proportional to respiration rate. Two bands are often used, one placed on the upper chest and another around the lower abdomen, to measure the depth of inspiration and exhalation.

Another technique for measuring respiration is impedance plethysmography – using the fact that the impedance of the chest cavity varies with respiration. This is highly advantageous because both heart rate and respiration can be determined from one chest strap or article of clothing.

One other ecological technique of measuring respiration rate is to analyze the respiratory sinus arrhythmia (RSA) from the ECG waveform (as discussed above). As with impedance plethysmography, both heart rate and respiration rate can be determined using a single chest band. Although not nearly as accurate as piezo or strain-gauge bands (due to base-line variation and noise in the ECG signal), with improving signal processing techniques, the so-called ECG Derived Respiration (EDR) rate, has become an increasingly common method of determining respiration rate [e.g. Yeon07].

Blood Oxygenation

Blood oxygen saturation levels (SPO2) as well as the pCO2 levels can be used as another parameter in quantifying emotional response. The cause of changes in the oxygenation of the blood is multi-factorial. That is, it is not simply a function of respiration or cardiovascular activity or any one other system, but a combination of many systems. Photoplethysmography as discussed previous, is the most ecological technique for measuring blood oxygenation.

3.6.2 Effects of Autonomic Physiological Changes: Face thermography

Physiological indicators of the emotional state of performers, the synchronization of that emotional state with other performers and the audience may be also recorded via novel and less intrusive technologies such as thermography. In fact, infrared termography is the technique that uses an infrared imaging and measurement camera to "see" and "measure" invisible infrared energy being emitted from an object. Facial temperature is indeed related to the autonomous nervous system activity and may be partially correlated with other more established measures such as galvanic skin responses (GSR), heart rate (HR), and electromyography (see Jarlier, Grandjean, et al., 2011, in press). Thermographic images have the critical advantage of enabling completely ecological experimental set-ups. In fact, the camera can be hidden from participants' sight and do not require cabling nor electrodes to be placed on their bodies. In this context participants may be actively performing a given task (i.e. musicians playing together) or listen/watch to an experimental stimulus (i.e. an audience attending a concert), while thermal images are acquired non invasively. However, the use of such technology in behavioral research is extremely innovative and there is no computational tool nor guidelines for acquisition and analysis of this kind of data. Thus, thermography appears clearly an extremely interesting new method for the study of social interactions and emotional entrainment between multiple individual, although it require a strong initial efforts for its validation and the development of new ad hoc computational tools.

3.6.2.1 Basic principles of thermography

Thermography is based on the fact that, unlike visible light, in the infrared spectrum, everything with a temperature above absolute zero emits infrared electromagnetic energy. The higher the temperature of the object, the greater the infrared radiation emitted. All objects, cold or hot, radiate heat in the form of infrared energy. As an object increases in temperature, it radiates more energy, and the wavelength gets shorter. More specifically, thermal imaging cameras detect radiation in the infrared range of the electromagnetic spectrum (roughly 9000–14,000 nanometers or 9–14 µm) and produce images of that radiation. Since infrared radiation is emitted by all objects near room temperature, according to the black body radiation law, thermography makes it possible to see one's environment with or without visible illumination. The amount of radiation emitted by an object increases with temperature; therefore, thermography allows one to see variations in temperature. When viewed through a thermal imaging camera, warm objects stand out well against cooler backgrounds; humans and other warm-blooded animals become easily visible against the environment, day or night. Thermography has a long history, although its use has increased dramatically with the commercial and industrial applications of the past fifty years. Government and airport personnel used thermography to detect suspected swine flu cases during the 2009 pandemic. Firefighters use thermography to see through smoke, to find persons, and to localize the base of a fire. Maintenance technicians use thermography to locate overheating joints and sections of power lines, which are a tell-tale sign of impending failure. Building construction technicians can see thermal signatures that indicate heat leaks in faulty thermal insulation and can use the results to improve the efficiency of heating and air-conditioning units. Some physiological changes in human beings and other warm-blooded animals can also be monitored with thermal imaging during clinical diagnostics.

3.6.2.2 Data recorded with thermography

Thermal images, or thermograms, are actually visual displays of the amount of infrared energy emitted, transmitted, and reflected by an object. Because there are multiple sources of the infrared energy, it is difficult to get an accurate temperature of an object using this method. A thermal imaging camera is capable of performing algorithms to interpret that data and build an image. Although the image shows the viewer an approximation of the temperature at which the object is operating, the camera is actually using multiple sources of data based on the areas surrounding the object to determine that value rather than detecting the actual temperature. This phenomenon may become clearer upon consideration of the formula Incident Energy = Emitted Energy + Transmitted Energy + Reflected Energy where Incident Energy is the energy profile when viewed through a thermal imaging camera. Emitted Energy is generally what is intended to be measured. Transmitted Energy is the energy that passes through the subject from a remote thermal source. Reflected Energy is the amount of energy that reflects off the surface of the object from a remote thermal source. If the object is radiating at a higher temperature than its surroundings, then power transfer will be taking place and power will be radiating from warm to cold following the principle stated in the Second Law of Thermodynamics. So if there is a cool area in the thermogram, that object will be absorbing the radiation emitted by the warm object. The ability of both objects to emit or absorb this radiation is called emissivity. Emissivity is a term representing a material's ability to emit thermal radiation. Each material has a different emissivity, and it can be quite a task to determine the appropriate emissivity for a subject. A material's emissivity can range from a theoretical 0.00 (completely not-emitting) to an equally-theoretical 1.00 (completely emitting); the emissivity often varies with temperature.

3.6.2.3 Thermography in medicine

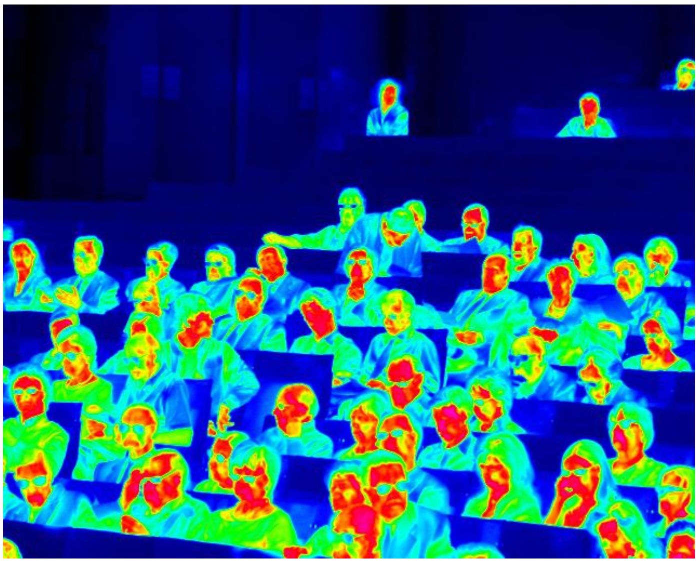
One of the most common clinical use of thermography is based on the principle that metabolic activity and vascular circulation in both pre-cancerous tissue and the area surrounding a developing breast cancer is almost always higher than in normal breast tissue. Thus digital infrared imaging uses extremely sensitive medical infrared cameras and sophisticated computers to detect, analyze, and produce high-resolution diagnostic images of these temperature variations. These temperature variations may be among the earliest signs of breast cancer and/or a pre-cancerous state of the breast. Other recent studies have demonstrated important experimental applications in other clinical branches such as sexual medicine (for the study of erectile disfunctions) or metabolic disorders, repetitive strain injuries, pain syndromes, arthritis, vascular disorders among others (See the IACT (International Association of Clinical Thermology) website: http://www.iact-org.org).

3.6.2.4 Thermography in neurophysiology and in the SIEMPRE project

Although thermography proved very useful in clinical medicine, very little has been done in the study of emotional responses of patients or healthy subjects (Murthy, Pavlidis, 2006; Sun, Pavlidis, 2006; Fei, Pavlidis, 2006; Garbey et al., 2007; Shastri et al., 2009; Fei, Pavlidis, 2010; Murthy et al., 2010; Jarlier et al, 2011). One possible reason for this is that qualitative or simple analyses such as hand-drawn regions of interest mean temperature of a still thermogram, are quite easy. These methods may be sufficient in clinical environment but are inadequate for the accurate measures required in basic sciences. In fact the emotional state triggered by a stimulus certainly evolves in time, and may migrate on the subject's body thus forming complex patterns of temperature changes. Furthermore, it's necessary to extract relevant features in a semi-automatic manner for large amounts of data. Thus far few applications have been shown such as those presented by the group of Pavlidis (Pavlidis et al., 2002a; Pavlidis et al., 2002b; Pollina et al., 2006). The works of Pavlidis mostly revolves around deception research, showing a great potential for the use of thermography in measuring automatic emotional responses. However there are several issues with movement artifacts and which features are most relevant for the detection of emotional states changes.

Our aim is to measure multiple subjects at the same time without constraining their movements too much for ecological purposes (i.e. the audience scenario). This requirement adds further complexity to the data analysis and requires novel computational tools we will develop ad hoc. At present we are developing multiple computational tools to automatically remove data artifacts, extract relevant features and correlate these features with the experimental variables manipulated in the stimuli. The present approach draws from computer vision methods for the data-driven analysis and image segmentation.

The first step is to define temperature changes caused by head (body, hands, etc.) movements alone and those elicited by stimuli (Zhou at al., 2009). Large movement artifacts will be removed by segmenting single images in increasingly smaller areas. Then, temporally consecutive areas that change their temperature abruptly (not following a physiological rate of change) will be considered as artifact. Small movement artifacts will be removed by subsampling frame rate (from 100 to 1 Hz) choosing images whose temperature contours are maximally overlapping (difference images). The second step regards the actual extraction of temperature features from the data. This step may diverge into two different approaches. The first one, partially hypothesis-driven will use state of the art facial tracking algorithms to automatically draw region of interest over subjects' faces (i.e. Buddharaju et al., 2007). The second approach will extract overall temperature changes across all image in the physiological temperature range shown in literature (29°C-34°C). One approach searches for specific changes in facial temperature, upper to lower part temperature ratio or left to right ratio (Zhu et al., 2007; Shastri et al., 2008). The other will instead search for global changes in all participants as if they were a single body exposed to the same emotional stimulus. The final step will be that of correlating the temporal evolution of temperature features and other manipulated features in the stimuli relevant for the SIEMPRE project (i.e. musical or auditory dimensions). We are also using Principal Component Analysis (PCA) and Region Of Interest (ROI) in order to investigate the patterns of temperature changes related to emotions elicited by music and other kinds of stimuli (see Jarlier et al, 2011 for a description of these methods for thermography analysis).



Snapshot from the SIEMPRE feasibility study performed by IIT in collaboration with UNIGE at a concert at Teatro dell'Opera Carlo Felice (November 2010). Measurement of an audience by thermocamera.

3.6.3 References

Murthy JN, van Jaarsveld J, Fei J, Pavlidis I, Harrykissoon RI, Lucke JF, Faiz S, Castriotta RJ. Thermal infrared imaging: a novel method to monitor airflow during polysomnography. Sleep. 2009 Nov 1;32(11):1521-7.

Fei J, Pavlidis I. Thermistor at a distance: unobtrusive measurement of breathing. IEEE Trans Biomed Eng. 2010 Apr;57(4):988-98.

Jarlier, S., Grandjean, D., Delplanque, S., N'Diaye, K., Cayeux, I., Velazco, M.-I.,Sander, D., Vuilleumier, P., & Scherer, K.R. (in press). Thermal Analysis of Facial Muscles Contractions. IEEE Transactions on Affecitve Computing , 2 (1).Shastri D, Merla A, Tsiamyrtzis P, Pavlidis I. Imaging facial signs of neurophysiological responses. IEEE Trans Biomed Eng. 2009 Feb;56(2):477-84.

Shastri D, Tsiamyrtzis P, Pavlidis I. Periorbital thermal signal extraction and applications. Conf Proc IEEE Eng Med Biol Soc. 2008;2008:102-5.

Zhou Y, Tsiamyrtzis P, Pavlidis IT. Tissue tracking in thermo-physiological imagery through spatio-temporal smoothing. Med Image Comput Comput Assist Interv. 2009;12(Pt 2):1092-9.

Zhu Z, Tsiamyrtzis P, Pavlidis I. Forehead thermal signature extraction in lie detection. Conf Proc IEEE Eng Med Biol Soc. 2007;2007:243-6.

Sun N, Pavlidis I. Counting heartbeats at a distance. Conf Proc IEEE Eng Med Biol Soc. 2006;1:228-31.

Fei J, Pavlidis I. Analysis of breathing air flow patterns in thermal imaging. Conf Proc IEEE Eng Med Biol Soc. 2006;1:946-52.

Garbey M, Sun N, Merla A, Pavlidis I. Contact-free measurement of cardiac pulse based on the analysis of thermal imagery. IEEE Trans Biomed Eng. 2007 Aug;54(8):1418-26.

Buddharaju P, Pavlidis IT, Tsiamyrtzis P, Bazakos M. Physiology-based face recognition in the thermal infrared spectrum. IEEE Trans Pattern Anal Mach Intell. 2007 Apr;29(4):613-26.

Pollina DA, Dollins AB, Senter SM, Brown TE, Pavlidis I, Levine JA, Ryan AH. Facial skin surface temperature changes during a "concealed information" test. Ann Biomed Eng. 2006 Jul;34(7):1182-9.

Murthy R, Pavlidis I. Noncontact measurement of breathing function. IEEE Eng Med Biol Mag. 2006 May-Jun;25(3):57-67.

Pavlidis I, Levine J. Thermal image analysis for polygraph testing. IEEE Eng Med Biol Mag. 2002a Nov-Dec;21(6):56-64.

Pavlidis I, Eberhardt NL, Levine JA. Seeing through the face of deception. Nature. 2002b Jan 3;415(6867):35.

3.6.4 Visible (Overt) Effects of Autonomic Physiological Changes: Tears, Eye Blinks, Pupil Dilation, and "Goose Bumps"

There are several overt changes that can directly indicate activity of the autonomic nervous system. While this chapter focuses on physiological changes, since these overt changes can be a direct (uncognitively mediated) function of the underlying physiology, they are worth mentioning. Overt properties of the eye that fall into this category include tears (tear volume and ocular hydration), eye blinks, and pupil dilation which can be measured with various visual recognition systems that can co-exist with HCI. All have been associated with various changes in emotional state. Visual recognition systems have also been used in attempting to quantify "goose bumps" or "goose flesh" or the pilomotor reflex. The anecdotal reporting of the pilomotor reflex is commonly mentioned in descriptions of emotional response.

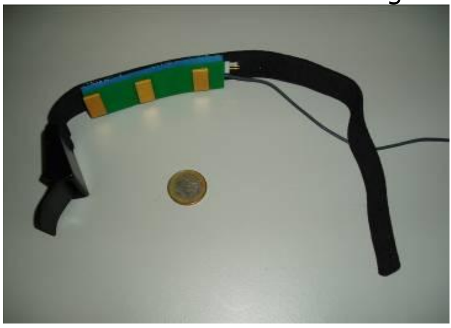
Somatic Nervous System

Muscle Activity

As has been discussed in other chapters of this book, it is well know that overt facial gestures are a well-studied indicator of emotional state. However, before visible movement occurs on the face, activation of the underlying musculature must occur. Indeed, there are many circumstances where there are measureable changes in the activation of the facial muscles and no visible facial gesture. This can be due to "rapid, suppressed, or aborted" expressions [Cacioppo92] or due to the actual attachment of the muscular structure to the skin. Thus, the ability to measure muscular activation in the face is another point in which physiological measurement can supplement measurement of overt changes. Changes in muscular tension in other areas of the body such as the arms may also indicate an overall level of stress or, as with the face, indicate the presence of overt gestures that cannot be viewed with visual observation.

Ecological Measurement of Muscle Activity

Measurement of muscle tension is commonly achieved using surface electromyography (sEMG or just EMG). Surface electromyography measures muscle activity by detecting the electrical potential that occurs on the skin when a muscle is flexed. This electrical potential is created by motor neurons depolarizing or "firing" causing the muscle fibres to contract. The rate (frequency) of the depolarization is proportional to the amount of contraction (until the individual motor neuron begins to saturate). At the same time more and larger motor neurons are recruited and begin to fire simultaneously. Thus, as shown in , an increase in muscle contraction is seen as an increase in amplitude of the EMG as well as a modulation of the frequency spectrum. The structure of motor neurons and the muscle fibres they innervate is called a motor unit and the potential measured by the surface EMG is commonly referred to as a Motor Unit Action Potential or MUAP. As with measurement of the ECG, EMG measurement involves the use of electrodes which are placed on the skin surface.



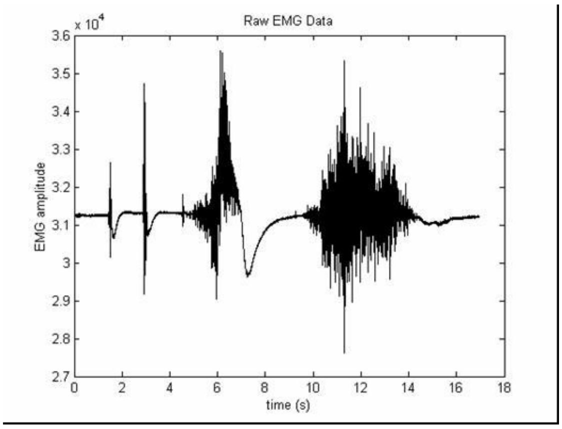
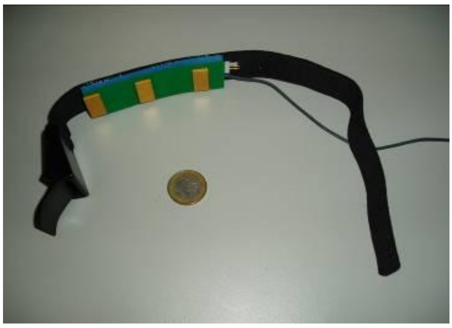


Figure 11: EMG Waveform. Note the four increasingly large contractions events indicated by an increase in amplitude and spectral complexity of the EMG



.

Figure 12: An ecological EMG band for wearing on arm or leg (www.infusionsystems.com)

Central Nervous System: Brain imaging

The central nervous system (CNS) is composed of the brain and the spinal cord and is responsible for processing information and controlling the activity of the peripheral nervous system. Over the past decade, functional imaging of the brain (imaging of the dynamic function rather than the static condition of the brain) during emotion inducing activities has been yielding an increasing body of knowledge of how and where emotions are processed in the brain [Bradley07]. The capability to compare and correlate CNS activity as measured with functional imaging with the activities of the ANS as measured using the techniques discussed previously has the potential to reveal mapping functions between emotional stimulation and physiological response. Unfortunately, because of the enormous size, cost and operational/environmental constraints, the most important functional imaging techniques such as positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) are unlikely to be used as part of a human-machine interaction scenario for several years to come (see Figure 14). However, due to its temporal and spatial resolution, brain imaging devices and particularly fMRI because of its non-invasiveness are a fundamental tool to study the brain at work. In particular, as far as SIEMPRE is concerned, fMRI will allow us to investigate the inter-individual correlates of regional brain activity during listening of music.

In the third part of the project, fMRI will be employed to elucidate, from a neurophysiological point of view, the common background shared by action, language and music by investigating, both in expert players and in naïve subjects, the cortical pattern of activation induced by watching an action, listening to the sound of that action, watching/listening to music (known or unknown) and watching/listening to a speaker (native or foreign language). The analysis of the cortical superimpositions between these different conditions would allow a better understanding of the cortical basis of what has been called ‘supramodal syntax’ (Fazio et al., 2009).   Moreover, we are planning to adapt to our data the technique originally proposed by Hasson et al. (2004) that has been used with two different approaches: voxel-by-voxel intersubject correlation and reverse correlation. Using fMRI we will also investigate how listeners are able to attribute emotions to musical excerpts adopting a “model based” approach: the signals derived from a computational model for a specific cognitive process (for example the mechanisms related to emotional attribution to music) are correlated against fMRI data from participants performing a relevant task to determine brain regions showing a response profile consistent with that model, see for example O'Doherty JP, Hampton A, Kim H, 2007).

Voxel-by-voxel intersubject correlation. In this analysis one searches for inter-subject correlation by using the spatiotemporal activation profile in a source brain to predict the activation in other fellow brains. To that end we will normalize all brains into a Talairach coordinate system, spatially smooth the data, and then use the time-course of each voxel in a given source brain as a predictor of the activation in the corresponding voxel of the target brain. The strength of this across-subject correlation measure stems from the fact that it allows the detection of all sensory-driven cortical areas without the need of any prior assumptions as to their exact functional role.

Reverse correlation. In order to identify the source of such powerful common 'consensus' between different individuals listening the same music or watching the same movie, we will adopt an analysis approach loosely analogous to the 'reverse correlation' method used for single-unit mapping (Ringach et al 2002). In this analysis, we will use the peaks of activation in a given region's time-course to recover the stimulus events that evoked them. Thus, constructing a regionally specific 'movie' which was based on the appended sequence of all frames which evoked strong activation in a particular region of interest (ROI), while skipping all weakly activating time points.

Figure 13: fMRI Machine

Central Nervous System: EEG

Electroencephalography (EEG) measures the electrical activity of the brain as it appears at the surface of the skull. As the millions of neurons within the brain "fire", the electric field generated by the electrochemical process can be measured using surface recording electrodes that function in a similar manner as the electrodes used for EMG and ECG. However, unlike ECG signals which are typically measured using up to twelve electrodes, EEG recordings can use anywhere from 3 to 256 (and even more!) electrodes (see Figure 14). The choice of electrode quantity will depend on the desired number of locations on the head to be measured and the desired spatial resolution within any given location. The measurement of the EEG is also considerably more difficult than either the EMG or ECG because many of the important features of the EEG signal are more than 1000 times smaller than the EMG or ECG signal. Measurement of signals below 1μV are common and can require that more attention is paid to cleaning the electrode site and applying electrolyte. Additionally, considerable signal averaging (comb filtering), spectral and spatial filtering is required to remove non-EEG signals, "noise", from the EEG signal. These "noise" signals include EMG and movements of the eyes causing baseline potential shifts (EOG).

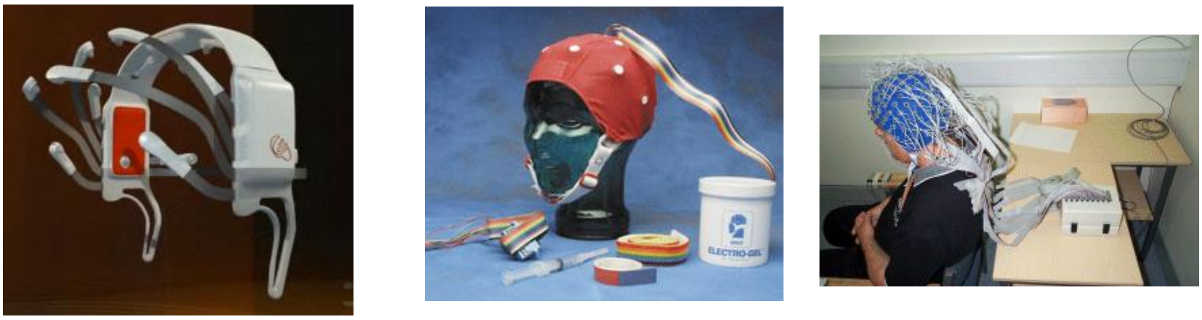


Figure 14: Three examples of the range of electrode quantity found in EEG interfaces. From left to right: the emotivEpoc (www.emotiv.com), a standard 10-20 EEG array (www.biopac.com), and a 256 lead array (www.biosemi.com).

There are three general areas of research into the use of EEG measurement as an indicator of emotion:

1. Spatial location and distribution of EEG signals. While emotions are not located in any one particular location of the brain [Phan04], amplitude and temporal asymmetries in EEG response patterns can yield insight into emotional state [e.g. Costa06].

2. The temporal response patterns of the EEG to stimuli – Evoked Response Potentials (ERP). The pattern of the time, location, and amplitude of the EEG response to auditory, visual, or emotional imagery can yield clues to emotional state and processing. While the ERP is studied at locations across the skull, the time response is limited to a narrow range from 100mS (fast) to 1000mS (slow). This is one of the largest areas of research in the physiological correlates of emotional state – especially in the area of the correlation of ERP's and emotionally evocative visual stimuli (e.g., [Holmes03] and [Schupp03]).

3. The frequency structure of the EEG signal – "Brain Waves". Synchronization of neural activity underneath the recording electrodes gives rise to oscillatory behavior in a collection of frequency bands. These bands range include:

   - Delta Band (1-4Hz)
   - Theta Band (4-8Hz)
   - Alpha Band (8-13Hz)
   - Beta Band (13-30Hz) and
   - Gamma Band (36-44Hz)

Correlations between the presence, timing, and location of these frequencies on the skull have been found to be related to several aspects of attention, vigilance, and emotional state (e.g. [Sebastiani03]).

EEG and entrainment

There are any number of instances where we can observe entrainment (see above for a definition) take place but especially so where music is concerned. Hence, it can be found at many levels, from conscious behaviors such as head-bobbing or foot-tapping to the beat of a song, to the alteration of heart and respiration rates in a listener (Etzel, Johnsen, Dickerson, Tranel, & Adolphs, 2006; Khalfa, Roy, Rainville, Dalla Bella, & Peretz, 2008), to the modulation of the amplitude of brainwaves for example at corresponding sound frequencies (Will & Berg, 2007). This of course, depends on how strict one's definition of entrainment is. But if one accepts the wider definition of entrainment as "the coordination of temporally structured events through interaction" (Clayton, Sager & Will, 2005), we can see that in the context of music listening, entrainment could not only apply to processes such as physiological tempo entrainment where the interacting oscillators are internal rhythms like heart rate and external ones like the period of a piece, but also other to other kinds of internal oscillators, such as the regular firing rate of neural populations in a given brain area for example, since this type of signal too can be described by a sinusoid. Naturally, because of its high temporal resolution, EEG is easily the best-suited technique for the study of entrainment between rhythmic stimuli and brainwave oscillatory activity.

The concept of brainwave entrainment was explored by Will and Berg in their 2007 study on brain wave synchronization. In this EEG study, ten participants listened passively to periodic drum or click sounds with repetition rates that varied between 60 and 480bpm. The authors found increased phase synchronization in all frequency bands in addition to three distinct components, one of which represents an entrainment response in repetition rates between 60 and 300bpm in the corresponding EEG frequency bands (i.e. 1 and 5Hz) according to the authors. These promising results were brought about through the use of stimulus related phase coherence analyses, which we therefore suggest be used in future SIEMPRE experiments dealing with rhythmic stimuli.

But other techniques are being explored as well. In a different study where participants listened passively to either text or music, Bhattacharya, Petsche and Pereda (2001) looked at the level of synchrony between evoked gamma band activities between two brain regions by using a similarity index allowing for the detection of asymmetric interdependency.

We would like to propose that entrainment happens at various points in the perceptual process, beginning with a first kind of entrainment where neural population in primary and secondary auditory areas synchronize with the beat of a musical stimulus, and then a second type entrainment where the already entrained neural populations in the auditory cortex entrain neural populations in other areas, such as premotor and motor areas. The interaction between these regions and others, especially at the subcortical level will be also an interesting topic to better understand how our brain is able to represent rhythms and how such external stimuli are able to entrain musicians and audiences sharing a common feeling. Within this context, it seems reasonable to use a combination of the techniques described above in order to further our understanding of both kinds of entrainment and the processes involved in the perception of rhythm at individual level but also in social interactions (dyadic EEG experiments are planned).

3.6.4.1 References

P. Fazio, A. Cantagallo, L. Craighero, A. D'Ausilio, A.C. Roy, T. Pozzo, F. Calzolari, E. Granieri, L. Fadiga. Encoding of human action in Broca's area. Brain. 2009 Jul;132(Pt 7):1980-8.

U. Hasson, Y. Nir, I. Levy, G. Fuhrmann, R. Malach. Intersubject synchronization of cortical activity during natural vision. Science. 2004 Mar 12;303(5664):1634-40.

O'Doherty JP, Hampton A, Kim H. (2007). Model-based fMRI and its application to reward learning and decision making. Ann N Y Acad Sci., 1104:35-53.

D.L. Ringach, M.J. Hawken, R.M. Shapley. Receptive field structure of neurons in monkey primary visual cortex revealed by stimulation with natural image sequences. Journal of Vision, 2002; 2, 12-24.

Bhattacharya, J., Petsche, H., & Pereda, E. (2001). Long-range synchrony in the gamma band: role in music perception. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 21(16), 6329-6337.

Clayton, M., Sager, R., & Will, U. (2005). In time with the music: The concept of entrainment and its significance for ethnomusicology. European Seminar in Ethnomusicology, 1, 3-75.

Etzel, J. A., Johnsen, E. L., Dickerson, J., Tranel, D., & Adolphs, R. (2006). Cardiovascular and respiratory responses during musical mood induction. International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology, 61(1), 57-69. doi:10.1016/j.ijpsycho.2005.10.025

Khalfa, S., Roy, M., Rainville, P., Dalla Bella, S., & Peretz, I. (2008). Role of tempo entrainment in psychophysiological differentiation of happy and sad music? International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology, 68(1), 17-26. doi:10.1016/j.ijpsycho.2007.12.001

Large, E. W., & Kolen, J. F. (1994). Resonance and the perception of musical meter. Connection Science: Journal of Neural Computing, Artificial Intelligence and Cognitive Research, 6(1), 177-208.

Will, U., & Berg, E. (2007). Brain wave synchronization and entrainment to periodic acoustic stimuli. Neuroscience Letters, 424(1), 55-60. doi:10.1016/j.neulet.2007.07.036

3.7 Physiology and Human-Machine Interaction

Another question characterizing SIEMPRE is: "if, to find correlates of emotional states, physiological signals must be measured in meticulously controlled environments, how then can they be used as part of an ecological human-computer interaction paradigm?". To see how researchers have answered this, we must examine what has occurred in three separate fields:

3.7.1 Physiological control of music

The first endeavours to use physiological signals to control machines occurred well before mice and GUI's and even PC's existed. One of the most interesting examples, as shown in **Errore. L'origine riferimento non è stata trovata.**, was Alvin Lucier's piece in 1965 entitled "Music for Solo Performer". In this performance he sonified the alpha activity of his EEG. While the idea was to listen to changes of cognitive state, i.e., there was no direct intention of quantifying emotional state, the synchronization of alpha activity is proportional to relaxation and so there was probably more than a little self-induction of a low activation emotion.

Figure 15: From Alvin Lucier's *Music for Solo Performer* - Controlling Sound with EEG

In 1978, Dick Raaijmaker used EMG, EDA, ECG, and acoustic measurement of respiration to sonify the level of stress of an individual dismounting a bicycle over the course of 30 minutes. While a large component of the changes in physiological state was caused by physical exertion, changes in emotional state throughout the course of the piece clearly influenced the sound.

The increasing use of physiological state (and by possible unintentional consequence, emotional state) as a tool for artistic expression led to the development by Knapp and Lusted in 1987 of the BioMuse [Knapp90]. This was one of the first commercially available systems which enabled musical performers to use physiological state (in this case EMG, ECG, and EEG) to control consumer electronic musical instruments and introduced the capability to control the newly ubiquitous PC.

3.7.2 Physiological control of computers (without emotional assessment)

At this time, research on using physiological signals to augment human-computer interaction became established with work on EMG [Putnam93], EOG [LaCourse90], and the new field of Brain Computer Interfaces (BCI) [Wolpaw98] using EEG. Much of this work was targeted at improving interaction for those with disabilities [Lusted96]. For example, ERP's from the EEG were combined with EOG to augment mouse control [Patmore96].

Research on physiologically augmented interfaces has continued to expand into many arenas including assistive living and computer gaming. From these investigations new, more ecological interfaces are being created [Knapp05] from which forms the foundation for applying physiological measurement of emotion to HCI.

3.7.3 Physiological control of computers with emotional state assessment

Combining the research on physiological interfaces with the ongoing research on psychophysiology and emotion, Picard and her colleagues at the MIT Media Lab began developing what she termed "affective interfaces" and "affective clothing" [Picard97]. Interfaces ranging from jewellery to gloves (see Figure 16) were being used to investigate whether correlates of emotional state could be found with ecological interfaces.

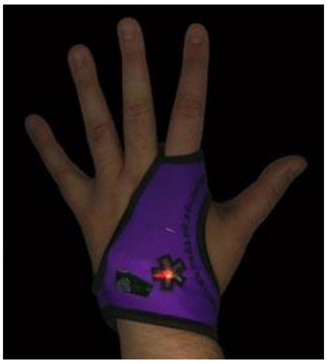


Figure 16: The Galvactivator - Mapping EDA to Light

Many new physiologically based systems are currently being created in research centres and from commercial enterprises. Some of these can be seen in Figure 4, Figure 12, and Figure 14. In the field of music, some researchers are beginning to look at using emotion as part of conducting [Nakra00] and performing [Knapp05]. It still remains to be seen, however, whether measurements using any of these new interfaces, even in conjunction with other measures such as facial recognition or speech recognition and operating in constrained environments, can accurately assess emotional state. This is the ultimate question of the current research, "Can an automatic recognition system be developed that can process physiological signals and other indicators of emotional state and come to any reasonably consistent result?".

3.8 SIEMPRE Physiological Indicators

Based on the above background work, SIEMPRE will focus on a specific subset of the ecological physiological indicators of emotion based on the SIEMPRE scenarios. These can be divided into the three scenarios:

1. For quartets:
   a. The mobilemuse: GSR, HRV, temperature, and motion.
   b. The BioWave: Single channel EEG, EOG, and facial EMG.
   c. Respiration (depending on agreement with performer).
   d. The BioFlex: Multiple EMG channels depending on performer and instrument.

2. For Conductor:
   a. The mobilemuse: GSR, HRV, temperature, and motion.
   b. The BioWave: Single channel EEG, EOG, and facial EMG.
   c. Respiration (depending on agreement with performer).
   d. The BioFlex: Multiple EMG channels on arm.

3. For Audience:
   a. The mobilemuse: GSR, HRV, temperature, and motion.

In line with the psychological evidence sketched above, it will not be assumed that sensor outputs can be translated directly into statements about emotion. Rather they will be correlated with evaluative reports, overt actions and context, to arrive at reasonable judgments about audience and performer states. Hence they form part of a package with the psychological measures related to quality of experience.

3.8.1 References

Baumgartner, T., Esslen, M., & J□ancke, L. (2006). From emotion perception to emotion experience: Emotions evoked by pictures and classical music. International Journal of Psychophysiology, 60 (1), 34-43.

Margaret M. Bradley and Peter J. Lang, "Emotion and Motivation" in Handbook of Psychophysiology, Cambridge University Press, pp 582- 607, 2007.

John T. Cacioppo, L.K. Bush, & L.G Tassinary. (1992). "Microexpressive Facial Actions as a Function of Affective Stimuli: Replication and Extension," Personality and Social Psychology Bulletin, 1992. vol. 18, pp. 515–26.

John T. Cacioppo and Wendi L. Gardner, "Emotion," Annu. Rev. Psychol. 1999. 50:191.214.

John T. Cacioppo, G. G. Berntson, J.T. Larsen, K.M. Poehlmann, and T.A. Ito. "The Psychophysiology of Emotion," in Handbook of Emotions, Edited by Michael Lewis, Jeannette M. Haviland-Jones, Guilford Press, 2000, pp. 173-191.

John T. Cacioppo, Louis G. Tassinary, and Gary G, Bernston, eds., Handbook of Psychophysiology, 3rd edition, Cambridge University Press, 2007.

W. B. Cannon, "The James-Lange Theory of Emotions: A critical examination and an alternative theory." American Journal of Psychology, vol. 39, pp. 106–127, 1927.

T. Costaa, E. Rognonib and D. Galati, "EEG phase synchronization during emotional response to positive and negative film stimuli", Neuroscience Letters, Volume 406, Issue 3, 9 October 2006, Pages 159-164.

R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J. G. Taylor, "Emotion recognition in human-computer interaction," IEEE Signal Processing Mag., vol. 18, pp. 32–80, 2001.

Guillaume Chanel, Julien Kronegg, Didier Grandjean, Thierry Pun: Emotion Assessment: Arousal Evaluation Using EEG's and Peripheral Physiological Signals, MRCS06, Istanbul, 530-537, 2006.

C. Conati, R. Chabbal, and H. Maclaren. A study on using biometric sensors for detecting user emotions in educational games. In 3rd Workshop on Affective and Attitude User Modeling, Pittsburgh, USA, June 2003.

Ulrich W. Ebner-Priemer and Thomas Kubiak, "Psychological and Psychophysiological Ambulatory Monitoring, A Review of Hardware and Software Solutions," European Journal of Psychological Assessment, Vol. 23, No. 4, pp. 214-227, 2007.

[Glowinski 2010] Glowinski, D., Camurri, A., Noera, C., Volpe, G., Cowie, R., McMahon, E., Knapp B., Jaimovich, J. "Using induction and multimodal assessment to understand the role of emotion in musical performance" in Christian, P. Crane, E., Fabri. M, Agius, H., Axelrod, L. (eds), Proceedings of the 4th Workshop on Emotion in Human-Computer Interaction, Liverpool, UK, ISBN: 978-3-8396-0089-4, Fraunhofer verlag, Stuttgart, 2010

A. Haag, S. Goronzy, P. Schaich, and J. Williams, "Emotion recognition using bio-sensors: First steps towards an automatic system," in ADA 2004, 2004, pp. 36–48.

A. Holmes, P. Vuilleumierb and M. Eimera, "The processing of emotional facial expression is gated by spatial attention: evidence from event-related brain potentials." Cognitive Brain Research, Volume 16, Issue 2, April 2003, Pages 174-184

K. Harata, M. Kawakami, and M. O'Rourke, “Pulse Wave Analysis and Pulse Wave Velocity: A Review of Blood Pressure Interpretation 100 Years After Korotkov,” Circulation Journal: Official Journal of the Japanese Circulation Society, vol. 70, no. 10, 2006, pp. 1231-1239.

Hudlicka, E. “To Feel of not to Feel: The Role of Affect in Human-Computer Interaction,” International Journal of Human-Computer Studies, 59, 1-32, 2003.

Center for the Study of Emotion and Attention [CSEA-NIMH], The International Affective Picture System: Digitized Photographs, Gainesville, FL: Center for Research in Psychophysiology, University of Florida, 1995.

W. James, *The Principles of Psychology*. New York: Holt, 1890.

K. H. Kim, S. W. Bang, and S. R. Kim, “Emotion recognition system using short-term monitoring of physiological signals,” Medical & Biological Engineering & Computing, vol. 42, pp. 419–427, 2004.

Jonghwa Kim, Elisabeth Andre, Matthias Rehm, Thurid Vogt and Johannes Wagner. Integrating Information from Speech and Physiological Signals to Achieve Emotional Sensitivity. In Proc. of the 9th European Conference on Speech Communication and Technology, 2005.

Maria Karekla, John P. Forsyth and Megan M. Kelly, “Emotional avoidance and panicogenic responding to a biological challenge procedure,” Behavior Therapy, Volume 35, Issue 4, Autumn 2004, Pages 725-746.

R. B. Knapp and H. S. Lusted, "A Bioelectric Controller for Computer Music Applications,” Computer Music Journal, MIT Press, Vol. 14, No. 1, pp. 42-47, Spring 1990

R. B. Knapp and Hugh S. Lusted, “Designing a Biocontrol Interface for Commercial and Consumer Mobile Applications: Effective Control within Ergonomic and Usability Constraints,” *Proceedings of the 11th International Conference on Human Computer Interaction*, Las Vegas, NV, July 22-27, 2005.

R. B. Knapp and P. R. Cook, “The Integral Music Controller: Introducing a Direct Emotional Interface to Gestural Control of Sound Synthesis,” *Proceedings of the International Computer Music Conference (ICMC)*, Barcelona, Spain, September 4-9, 2005.

R. B. Knapp, Jonghwa Kim, and Elisabeth André, "Physiological signals and their use in augmenting emotion recognition for human-machine interaction," in Emotion Oriented Systems: The HUMAINE Handbook on Emotion, Springer, 2011.

R. B. Knapp and Brennon Bortz, " MobileMuse: Integral Music Control Goes Mobile," Accepted for publication in *Proceedings of the New Interfaces for Musical Expression Conference,* Oslo, Norway, June, 2011.

S. D. Kreibig, F. H. Wilhelm, W. T. Roth and J. J. Gross, “Cardiovascular, electrodermal, and respiratory response patterns to fear- and sadness-inducing films,” Psychophysiology, vol. 44, 2007, 787–806.

U. Kunzmann, and D. Gruhn, (2005). Age differences in emotional reactivity: The sample case of sadness. Psychology and Aging, vol. 20, pp. 47–59.

J. R. LaCourse and F. C. Hludik, “An eye movement communication - control system for the disabled,” IEEE Trans. Biomed. Eng., vol. 37, pp. 1215 - 1220, 1990.

H. S. Lusted and R. B. Knapp, “Controlling Computers with Neural Signals,” *Scientific American*, October 1996.

DT Lykken. *A Tremor in the Blood: Uses and Abuses of the Lie Detector*, New York: McGraw-Hill Book Company, 1998.

Margaret Morris, Terry Dishongh, Farzin Guilak, “Apparatus for monitoring physiological, activity, and environmental data,” USPTO Applicaton #: 20080154098.

Nakra, T.M., “Inside the Conductor’s Jacket: Analysis, Interpretation, and Musical Synthesis of Expressive Gesture,” M.I.T. Media Laboratory Perceptual Computing Section Technical Report, no. 518, 2000.

F. Nasoz, K. Alvarez, C. Lisetti, and N. Finkelstein, "Emotion recognition from physiological signals for presence technologies," International Journal of Cognition, Technology, and Work - Special Issue on Presence, vol. 6(1), 2003.

National Research Council. The Polygraph and Lie Detection. Committee to Review the Scientific Evidence on the Polygraph. Division of Behavioural and Social Science and Education. Washington, D.C: The National Acadmenic Press, 2003.

D. W. Patmore and R. B. Knapp, "A Cursor Controller Using Evoked Potentials and EOG," Proc. of the RESNA '95 Annual Conf., Vancouver, Canada, pp. 702-704, June 9-14, 1995.

K.L. Phan, T.D. Wager, S.F Taylor, and I. Liberzon, "Functional Neuroimaging Studies of Human Emotions," CNSSpectrums, vol. 9, pp. 258–266.

R.W. Picard and J. Healey, "Affective Wearables," Personal Technologies, vol. 1 no. 4, 1997, pp. 231-240.

R. Picard, E. Vyzas, and J. Healy, "Toward machine emotional intelligence: Analysis of affective physiological state," IEEE Trans. Pattern Anal. and Machine Intell., vol. 23, no. 10, pp. 1175–1191, 2001.

W. L. Putnam and R. B. Knapp, "Real-Time Computer Control Using Pattern Recognition of the Electromyogram," *Proc. of the IEEE International Conf. on Biomedical Eng*., San Diego, CA, pp. 1236-1237, October 27-29, 1993.

J. A. Russell, "Core Affect and the Psychological Construction of Emotion," Psychological Review, Vol. 110, No. 1, 2003, pp. 145–172

L. Sebastiani, A. Simoni, A. Gemignani, B. Ghelarducci and E. L. Santarcangelo , "Human hypnosis: autonomic and electroencephalographic correlates of a guided multimodal cognitive-emotional imagery," Neuroscience Letters, Volume 338, Issue 1, 20 February 2003, Pages 41-44.

H.T Schupp, M. Junghofer, A.I. Weike, and A.O. Hamm, "The selective processing of briefly presented affective pictures: An ERP analysis," Psychophysiology, vol. 41, 2003, pp. 441–449.

Yeon Sik Noh, Sung Jun Park, Sung Bin Park, and Hyung Ro Yoon, "A Novel Approach to Classify Significant ECG Data Based on Heart Instantaneous Frequency and ECG-derived Respiration using Conductive Textiles," 29th Annual International Conference of the IEEE EMBS, 22-26 Aug. 2007, pp: 1503-1506.

J. Wagner, J. Kim, and E. André, "From physiological signals to emotions: Implementing and comparing selected methods for feature extraction and classification," in ICME'05, Amsterdam, July 2005.

John G. Webster, editor, "Medical Instrumentation, Application and Design, 3rd edition" John Wiley and Sons, 1998.

J.R. Wolpaw, H. Ramoser, D.J. McFarland, and G. Pfurtscheller, "EEG-based Communication: Improved Accuracy by Response Verification," IEEE Transactions on Rehabilitation Engineering, vol. 6, num. 3, 1998, pp. 326-333.

4. APPENDIX

In response to discussions that took place at the theoretical workshop of December 3rd 2010 in Geneva, it was decided to define a template both for the description and circulation of experimental designs and results amongst the partners: the detailed description of the specific experiments will be included in Deliverables D2.1 and D2.2.

The specifications of the template were discussed at the workshop and were later formatted by the UNIGE-CH group which has been since used by various partners for the description of pilot studies and planned studies alike.

Template for the specification and report of SIEMPRE experimental designs

Title	
Question of interest	
Leaders	
Other SIEMPRE groups involved	
Referent scenario	
Research objectives	
Theoretical hypotheses	
Operational hypotheses	
Relationship with the objectives of the project	
Time schedule	
Methods	
Participants	
Materials	
Data format	
Experimental protocol/procedure	
Measures	
Results	
Descriptive results	
Inference statistics	
Additional results	
Discussion	

As can be seen, the template has the advantage of acting as a guideline for the design of experiments related to SIEMPRE from the onset by virtue of specifying how they relate to the objectives and scenarios of the project. When partners fill it, the description of the performed or planned studies should be brief yet precise enough for partners to be able to replicate them should they wish to.

1. *Title*: the title of the experiment should be brief and self explanatory, quickly informing the reader about the important variables, dependent and independent, and expected links or the population under study if relevant (one sentence max).

2. *Question of interest*: the question of interest briefly describes the question the experiment is trying to investigate or to answer (in one sentence).

3. *Leaders*: the group that is primarily responsible for the study.

4. *Other SIEMPRE groups involved*: if the study is a collaboration between two or more partners the group that is not primarily responsible for the study should be mentioned here.

5. *Referent scenario*: partners are to indicate which SIEMPRE scenario the study is using. That is either the (i) musician-musician, (ii) conductor-musicians, (iii) music-listener, or (iv) musician-listener scenario as described in Annex 1.b.

6. *Research objectives*: in this section partners may describe in a little more detail than point 2 what the study in question is trying to achieve if they wish (~30 words or less).

7. *Theoretical hypotheses*: this section describes the expected results in terms of one or several theoretical hypotheses, i.e. expected links between variables, expected results according to conditions or general predictions.

8. *Operational hypotheses*: this section operationalizes the hypotheses described in the previous section by referring to the measures used for the study of each variable instead of the general concept under study (e.g. "arousal" operationalized in terms of a score on a self-reported scale or GSR).

9. *Relationship with the objectives of the project*: this section describes how the study relates to one or more of the three areas of focus of the project, i.e. entrainment, emotional contagion, and co-creation.

10. *Time schedule*: describes the time frames for different parts of the study, i.e. expected dates for beginning recordings for example or date by which preliminary results are to be expected.

11. *Methods*: this section describes how the study is being planned or was performed and contains points 12 through to 16 (it is not filled itself).

12. *Participants*: all relevant characteristics of the population are mentioned in this section such as total number, sex, age, whether they were musicians or not and other information as deemed relevant to the question of interest.

13. *Materials*: this section describes the equipment used, including video, audio, or other type of recording equipment, including physiological or neurological (EEG, etc). This section also includes a brief description of the stimuli used (video, audio) as well as any tests or questionnaires used with up-to-date references.

14. *Data format*: this section describes the format of the measures used.

15. *Experimental protocol/procedure*: this section describes the procedure of the experiment in chronological order from beginning to end including the recruitment procedure, instructions, the type of consent given and debriefing. This section also includes the duration of each part of the experiment (conditions, stimuli etc.) and total duration of the experiment.

16. *Measures*: this section describes the measures that were either planned and/or taken during the study. This is not the same as point 13, for e.g. a respiration belt is not a respiration measure, it is a material.

17. *Results*: this section describes the results of the study and includes points 18 through to 20. This section should only report the results and not comment on them.

18. *Descriptive results*: this section summarizes descriptive statistics as deemed relevant to the question of interest.

19. *Inference statistics*: this section includes the results to statistical tests of significance or other statistical tests as deemed relevant in relation to the operational hypotheses.

20. Ad*ditional results*: this section includes results to statistical tests of significance or other statistical tests that do not directly relate to the original question of interest or operational hypotheses but were deemed sufficiently important to be mentioned or proved promising.

21. *Discussion*: here the partners may briefly comment on the results (points 18 through to 20) obtained and how they relate to the question of interest or operational hypotheses. Here partners my make mention of what was and wasn’t achieved as well as limitations and problems encountered during the study.