# SIEMPRE

## D4.1 - Results from the first series of experiments and first evaluation report

| *Version* | *Edited by* | *Changes* |
|---|---|---|
| 0.9 | Antonio Camurri, Donald Glowinski | First draft |
| 1.0 | Roddy Cowie | Input from QUB, IIT |
| 1.1 | Roddy Cowie & all partners | Input to all sections |
| 2 | Roddy Cowie & all partners | Finalising all sections |

# TABLE OF CONTENTS

# 0. INTRODUCTION

This deliverable describes the preliminary results from the first series of SIEMPRE experiments which were previewed in D.2.1, and performed in the project's first year. The results of this first series of experiments will be used to refine the proposals to be set out in D2.2 "Second series of experiments" (month 24).

The aim of the report is not just to show that work has been done, but rather to provide an overview of progress that lends itself to a) evaluating the work and b) clarifying what the next steps should be. Hence this section previews the material within an overall framework that was developed by the partners in order to show how the parts of the research connect.

Fuller descriptions are given in the later sections. For ease of reference, they follow the numbering used in D2.1. For each experiment, there is an abstract summarising what has been done, and an evaluation of what it has achieved. Technical detail is kept brief, partly by referring to other sources when that is possible. Procedure is described by reference to the information set out in D2.1, avoiding duplication. Where changes have been made, they are described explicitly. In many cases, the results have already been published. There, the report refers to the published source rather than duplicate it. The relevant sources are listed at the end of the document.

## 0.1 An overall framework

Figure 1 summarises the overall conception behind the SIEMPRE experiments. The function of the experiments is to shed light on an intimately interconnected system, where multiple kinds of communication take place, resulting in a unique and highly valued kind of experience.

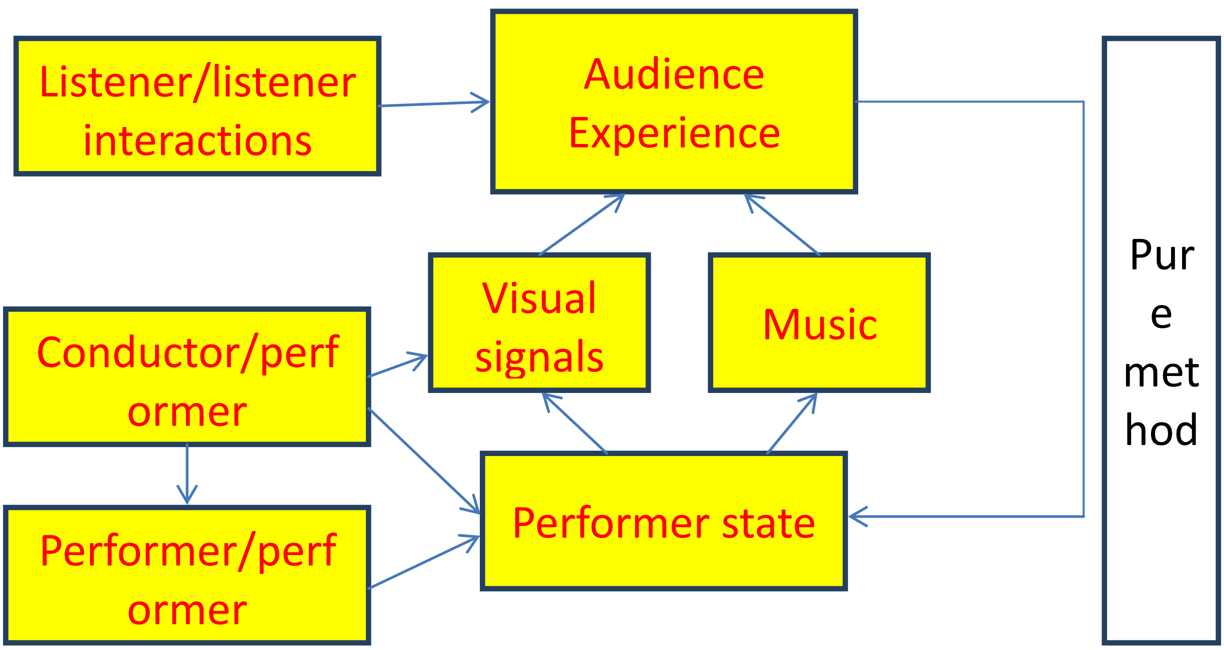



Figure 1: an outline of the SIEMPRE domain

The names of the main sub-areas in Figure 1 (in red) should be reasonably self-explanatory in context. Fuller explanations are given below along with the summaries of the work that has been done in each. Each one is a fascinating subject in itself, but the topics are not chosen simply because they are fascinating. They are chosen to build up an overall understanding of the way live performance works, and engages audiences.

A second level of structure needs to be considered alongside the topic areas. Part of it is shown as part of Figure 1. Some research is needed not to clarify a particular area, but to resolve methodological issues that affect several areas. Figure 1 classifies that as ‘pure method’. In addition, certain systematic divisions run through the topic areas. Three in particular are important.

- *Kinetic / dynamic* There is a systematic distinction between two ways of describing musicians’ behaviour in particular: kinetic (describing positions and movements of selected points) and dynamic (describing the forces that generate movements and sounds). In general, we expect these to provide complementary types of information.
- *Experimental / ecological* Behavioural research typically recognises a need to combine studies that allow control and precise measurement at the expense of artificiality (experimental) and studies that engage with natural situations or realistic models of them at the expense of control (ecological).
- *Mental / physical* At various points in figure 1, the variables that are logically important are at the mental level (e.g. what is experienced or communicated), but the variables that can be most satisfyingly measured are physical (e.g visceral changes or patterns of movement).

The research strategy in SIEMPRE tries to ensure that it does not rely inappropriately on information from one of those levels when in fact, both are needed.

## 0.2 Area by area summaries

This section provides a brief overview of what has been achieved in the various key areas. It gives brief summaries of study outcomes rather than descriptions of method or quantitative results. That level of detail is given in later sections. For ease of reference, experiments are identified by the numerical codes which were used in D2.1, and which are also used in the later sections.

**Conductor/performer interactions** Experiments 3.5 and 3.6 have shown that conductor behaviour influences performer state, and hence music. They recorded violinists' and conductors’ movement kinematics during execution of Mozart pieces, searching for causal relationships among musicians by using the Granger Causality method (GC). They show that the increase of conductor-to-musicians causal influence, together with the reduction of musician-to-musician coordination (an index of successful leadership) is reflected in quality of execution, as assessed by musical experts judgments. A technical description of results has been given in a full paper submitted for publication in a high impact journal.

**Performer/performer interactions** Experiments 2.1 and 2.2 tested a number of dynamic interdependence measures before concluding on the measure that provides the most satisfying results. The final chosen measure appears to confirm the hypothesis that violinists are influenced by each other’s intonation, and also by each other’s tempo when performing together, at least for the simple cases of non-professional musicians.

Experiments 3.3 and 3.4 explored measures of interdependence for kinetic measures. A set of experiments has been designed to investigate behavioural features that account for non-verbal communication within the music ensemble. A more detailed aspect of social interactions, namely dominance, was studied and preliminary findings revealed that dominance over others may be achieved through the regulation of individual and group's behavior complexity.

**Performer state** Dynamic measures are particularly suited to this issue. Experiment 3.1b explored how best to place electrodes in order to measure muscle tensions related to posture and execution. For a single musician, they provide rich data which relates to distinctions in style (emphatic or academic). Analysis of the data is challenging, and both standard and novel methods are being explored. Technically, it is difficult to envisage extending the technique to larger numbers (because of issues such as interference).

**Music** Music is central, and it is considered in multiple ways. Several of the performer behaviour studies mentioned consider the music in terms of global descriptions (such as quality of performance or academic/emphatic). Experiments 2.1 and 2.2 have studied how the intonation and tempo of music are controlled. In terms of listener response, experiments 1.1 and 1.2 examine the emotion that music is perceived to express, and the experimental part of 1.4 has identified physiological reactions that it evokes. Both are at a level that allows the judgments to be correlated with features of the music. Subtler questions emerge from other studies. Experiment 3.3 indicates that there are particular points in the music where ensemble playing differs markedly from solo playing, and sets the challenge of understanding what they are.

Overall, the studies highlight the challenges involved in finding a description of music that can bind its structure to the various factors relevant to its production and its effects. It is by no means an isolated problem. Similar challenges arise with other media, particularly speech.

### Audience Experience

Audience experience is both complex and difficult to pin down, and it has been approached at multiple levels.

Experimental studies have covered both mental and physical indicators. They have dealt with individuals rather than audiences. If convincing correlates can be found in individuals, then they provide hypotheses that are worth testing in the more challenging situation of a true audience.

Experiment 5.1 studied possible correlates of emotional engagement in the central nervous system. It built on influential theories which suggest that synchronies between brain subsystems are characteristic of emotion, and asked whether music induced measurable synchronies. Analysis is under way.

Experiment 1.4a studied correlates in peripheral systems associated with emotion (particularly heart rate and skin conductance). Preliminary results identify several physiological features that show significant differences between emotion types.

Experiments 1.1 and 1.2 moved into the domain of psychological rather than physical measures. They tested the ability of 'trace' techniques to capture time-varying emotional responses to music across the multiple dimensions identified in the GEMS scales. They established that there is robust agreement between raters on the way emotional content varies from moment to moment.

Ecological studies have also looked across physical and psychological measures.

Experiments 4.2 and 4.4 explored the potential of thermographic techniques to capture emotional responses. Statistical analysis of the data revealed that the emotional content of the music affected the facial temperature for specific parts of the face (left and for the upper). The use of thermography to measure the audience response to music was not included in the technical annex. In fact, this idea is one of the most positive outcomes coming from the meeting of the consortium. Thermography is potentially very useful for the scope of the project although it poses a series of technically challenging problems for data acquisition and analysis. During the project both IIT and UNIGE-CH have spent considerable efforts on these issues, reaching extremely interesting initial results and building a strong expertise on this technique.

Experiment 1.4b explored the use of measures that deal directly with mental states in an ecological setting. The literature points to a wide variety of factors that are likely to contribute to the entire "quality of experience": examples are emotion, social, attentional/flow, sense of refreshment, etc. These have been integrated into a questionnaire. We also developed a continuous self-report mechanism for use by audience members in a live concert scenario. A relatively ecological paradigm was developed to test the measures. A volunteer audience attended a concert consisting of two contrasting performances, one designed to engage them strongly, the other to be relatively alien. The questionnaire showed significant differences between the two performances on all the scales, which confirms that it does capture differences in the quality of experience. However, the continuous self-report mechanism was rarely used, and adaptions are clearly needed if it is to be useful.

**Pure method**

A substantial part of the work in this phase was concerned with testing and optimising sensor configurations. Experiment 2.3 studied (and confirmed) the feasibility of using the Polhemus motion sensors. Experiment 3.1 examined placement of EMG electrodes, and experiment 3.2 identified the MoCap Qualisys markers needed to measure non-verbal social communication in a quartet. Experiments 4.1 and 4.3 explored the usefulness of thermographic measures, and compared the measures yielded by 2 cameras.

A second issue of method was ensuring the reliability of the collected multimodal data. A series of recordings were carried out for that purpose

A different kind of methodological issue involves segmentation of musical performance. In experiment 2.2, key differences between Solo Vs Group performance appear to involve the interpretation of the attack and conclusion of musical phrases (and subphrases), in cadences. and in differences in the communicative intention in gesture at such moments.

That analysis pinpoints a crucial issue, which is the identification of key episodes. For example, let us consider the experiments with string quartets “Solo Vs Group Performance”. The following question emerges: are there particularly significant segments in the stream of multimodal data where the difference between solo and group performance are most evident (by a human observer)? Identification of these key episodes is critical for the application of automated analysis techniques.

## 0.3 Assessment of progress

We have established the capacity to make measurements in almost all of the areas shown in figure 1, in ways that allow us to explore most of the connections shown in the figure. These involve progress not only in technology, but also in psychological instrumentation and appropriate design. In some areas, we have identified powerful ways of exploring connections (notably the sequence of connections between conductor, musician/musician interaction, and music). The data collected is in itself a major resource, and analysing it will be a key activity in the next phase of the project.

The important gaps that remain are reasonably clear, and we have begun discussions on ways to address at least some of them. For example, given the techniques that we have, we can see in principle how to study the impact of listener/listener interactions and feedback from audience to performer.

Overall, we believe that the experiments keep SIEMPRE on course to achieve its goal of a uniquely full analysis of large-scale interactions.

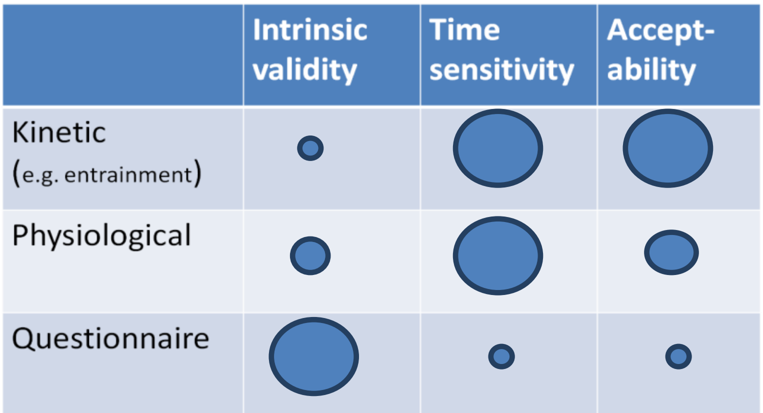
## 0.4 Study by study summaries

The summary above covers some 20 studies. It tries to convey the broad picture that they paint, but that is not compatible with going into very much detail about any one study. The following sections summarise each individual study in more detail.

# 1. SUBJECTIVE RATINGS

A major objective of the SIEMPRE project is the understanding of the mechanisms underlying the communication between performers and listeners. There are different ways of investigating self-reports from an audience depending on what we want to explore. For instance, Likert scales, adjectives checklists and free reports are among the most used (Zentner & Eerola, 2010). Although these are the methods that are used the most, the answers given by listeners are often delayed and more importantly, all these methods are static and therefore unable to account for the dynamic aspects of music and emotion. However, it seems necessary to capture this time flow to better understand the emotional responses to music. Thus, it is important for the SIEMPRE project to find suitable continuous self-report methods.

One set of options is summarized in the table below. Measures of audience movement and physiology have the necessary time sensitivity, and could be applied during a performance, but it is not obvious a priori what they tell us about the audience's feelings. On the other hand, questionnaires lack time sensitivity, and completing them thoroughly disrupts the listening experience. However, showing that the movement and physiological measures capture differences identified by questionnaires provides a way of establishing which aspects of experience they measure.




Another option is to use 'trace' or 'dynamic judgment' techniques. The pioneering studies by Cowie et al (2000) and Schubert (2001, 2004) in this field demonstrated the effectiveness and the reliability of continuous measurements. The main advantage offered by these continuous measurements is that we can follow the subjective ratings of individuals at each time point and thus make a link between these dynamic judgments and the study of musical structure or acoustic parameters presents in the signal in order to better understand the mechanisms underlying the subjective feeling or the perception of the emotions expressed by music.
This methodology has been used in different musical contexts (laboratory context vs. concert/ live performance) both for emotions expressed by music and felt emotions.
We present below the first wave of pilot studies with preliminary results using these methods in collaboration with different groups of professional musicians during both live performance and laboratory contexts.

## 1.1 Dynamic judgment of the audience during live performance

and

## 1.2 Dynamic judgment of a small audience during Workshop

| |
|---|
| Investigation of the agreement of the audience on emotional dimensions expressed by music during a live performance (exploratory study); testing the reliability of dynamic judgments performed by an audience on emotional dimensions expressed through music. |
| UNIGE-CH |

**Abstract**

In this experiment we tested the reliability of dynamic judgments performed by an audience on emotional dimensions expressed through music, during an ecological setting. The results show that people are able to recognize and assign emotions expressed by music, with this new method of continuous measurement. As reported in the D2.1, we worked with the Quartet Terpsycordes, the Cappella Mediterranea Ensemble (Ambronay Festival) and the Quartetto di Cremona. We report here the results obtained with the Quartetto di Cremona. The aim of this exploratory study was to investigate to what extent, on a given GEMS dimension (Zentner, Grandjean & Scherer, 2008), participants agreed on the emotion expressed by music. To this end, we collaborated with the famous Italian Quartet, "Quartetto di Cremona", during different settings. The first one was during a concert at the Eglise Saint-Germain in Geneva (live performance). The second one was during a workshop with the quartet at the University of Geneva. This experiment concerns the study of the emotions expressed by music but UNIGE-CH also plan to do this kind of experiment for felt emotions.

**Method**

The musical pieces and their GEMS dimensions are presented in the D2.1. In the two settings, we asked participants (Eglise Saint-Germain : N= 12; Workshop : N= 11) to rate continuously how the music expresses a given GEMS dimension while listening to the piece using a Flash interface during both the concert and the workshop. Participants moved the cursor up to indicate the intensity of the emotion expressed by music, while the musicians played.

**Results**

We can see with the Cronbach alphas that the results are really satisfactory, both for the concert and the workshop (Figures 1 & 2). Indeed, there is a large agreement between participants on the emotions expressed by music.

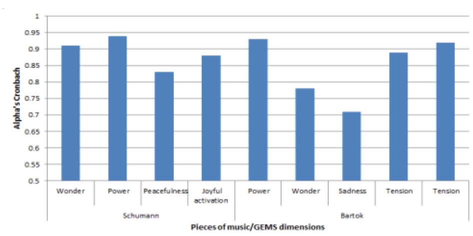



Figure 1.Cronbach Alphas of the dynamic judgments made during live performance.

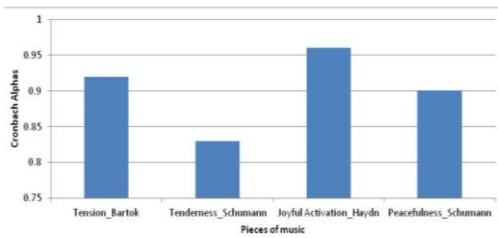



Figure 2.Cronbach Alphas of the dynamic judgments made during workshop.

The Cronbach alphas demonstrate that the method of dynamic judgment is relevant to better understand the relationship between the audience and the music produced by the musicians. Indeed, it is considered that the homogeneity of a measure is satisfactory when the value of the coefficient is at least .80 and the majority of our Cronbach Alphas is about .80 and .90.

Figures 3 and 4 show an example of the z-scores of the dynamic judgments made during the workshop and the concert, respectively.

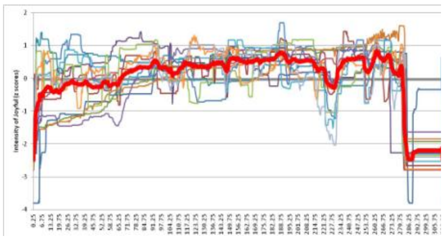


Figure 3. Workshop: Evolution profile of judgments of the emotion expressed by music in terms of Peacefulness. Haydn, *String Quartet n°2,* op.54.

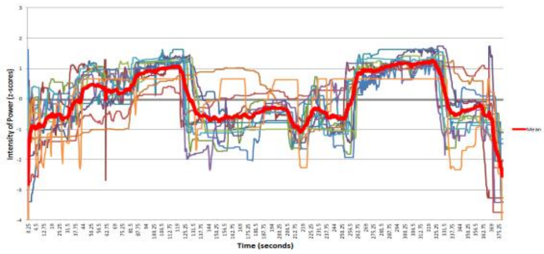


Figure4. Live performance: Evolution profile of judgments of the emotion expressed by music in terms of Power. Schumann, *String Quartet n°3 in A major op.41.*

There are the time on the x-axis and the intensity of the emotion expressed by music on the y-axis. The average is represented by the red line. There is variability of the judgments around the average but we observe a large agreement between participants throughout the musical pieces. In a second step, we want to investigate the cues (acoustic parameters and musical structure) that allow participants to recognize and assign an emotion expressed by music. To this end, we are collaborating with the Geneva University of Music in order to construct a typology of the musical structure relevant in the process of attribution of emotion to the music.

We also have two musical excerpts which were played in both live performance and workshop. Figure 5 (a and b) shows an example for a Schumann's movement and demonstrate again amazing results. These results will also be compared in our future analysis.

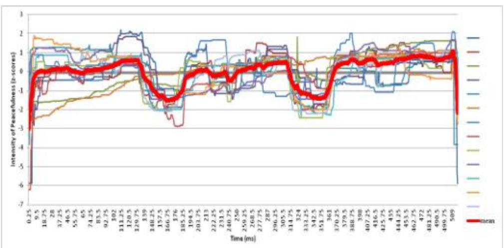


Figure 5 a) Dynamic judgments of a Schumann's movement made during the workshop.

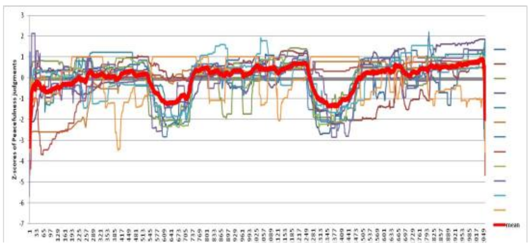


Figure 5 b) Dynamic judgments of a Schumann's movement made during the live performance.

## 1.3 Recordings and judgments of different types of musical expressiveness

| |
|---|
| Construct of musical stimuli in order to investigate the agreement of people on an emotional dimension expressed by music during a dynamic task.<br>Understand the acoustic characteristics and the musical cues related to these dynamic judgments. |
| UNIGE-CH |


**Abstract**

In this experiment, we collaborated with a famous French violinist, Renaud Capuçon, in order to create musical stimuli that will be evaluated during a laboratory context with the method of dynamic judgement (cf. Experiment 1.1 and D2.1), for both felt emotions (cf. Experiment 1.4) and emotions expressed by music.


**Method**

The aim of this study was twofold. Firstly, we want to investigate the impact of different musical expressiveness in a task of dynamic judgment. Secondly, we want to investigate the dynamic judgment of attribution of an emotion to the music across different kinds of musical expressiveness. To this end, 9 pieces for solo violin were recorded at UNIGE-CH in September of 2010, pieces corresponding to the GEMS dimensions. Each piece was performed in 3 different styles: academic, emphatic and natural. See the D2.1 for the description of the musical styles. The 27 resulting pieces were then used as stimuli in a judgment paradigm where we asked participants (N=89) a) to rate continuously the musical expressiveness and later answered how strongly they thought the musical excerpt was expressing the nine emotional dimensions of the GEMS, using the sliders (N= 44) and b) to rate continuously the emotion expressed by music, based on the GEMS dimensions and then answered how intense they found the musical piece with a single slider (N=45)**.**

**Results**

Cronbach alphas analysis revealed a heterogeneity in the agreement between participants for the musical expressiveness (min: .06; max: .98) and very good results for the agreement regarding the emotions expressed by music (min: .75; max: .99). Figure 1 presents an example of the z-scores of the dynamic judgments of the intensity of musical expressiveness with the *Allemande* of the Partita n°2 in D minor for violin solo, BWV 1004 by J-S Bach, in a) academic mode; b) emphatic mode; c) natural mode.

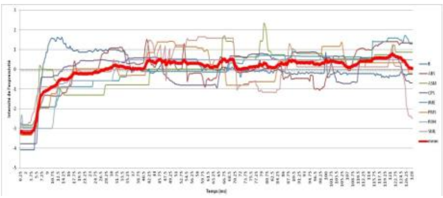


Figure 1.a) Z-scores of the dynamic judgment for the Bach's movement played in an academic mode.

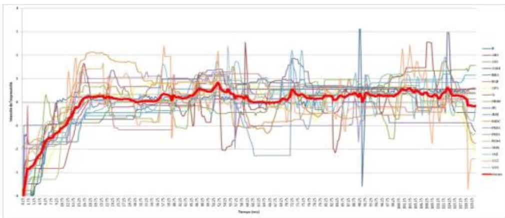


Figure 1. b) Z-scores of the dynamic judgment for the Bach's movement played in an emphatic mode.

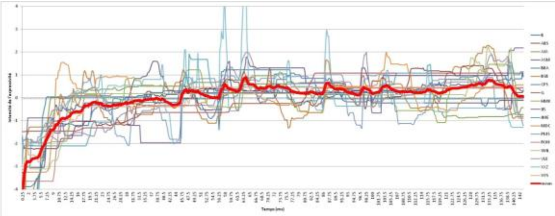


Figure 1.c) Z-scores of the dynamic judgment for the Bach's movement played in a natural mode.

We did permutation analysis in order to see significant differences between the academic and the emphatic excerpt. As shown in Figure 2, there are indeed significant differences between the dynamic judgments in academic and emphatic modes.

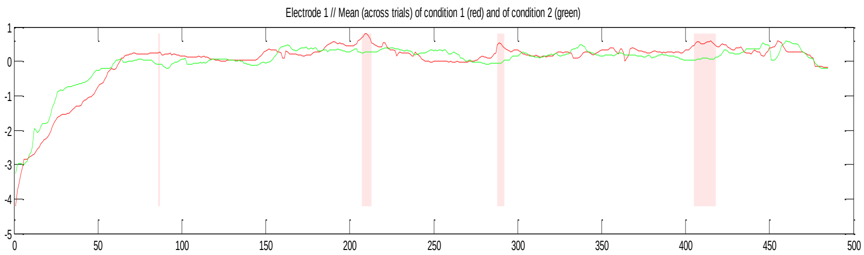



Figure 2. Comparison of the two conditions "emphatic" vs. "academic" for the Bach's movement. (Red line = emphatic mode ; green line : academic mode).

The time is represented on the x-axis while the intensity of the musical expressiveness is represented on the y-axis.

Regarding the evaluation of the emotion expressed by music, we found a quite beautiful result with the Mozart's movement judged on the Joyful Activation. Figure 3 shows the overlap of the three averages obtained during the task of dynamic judgment.

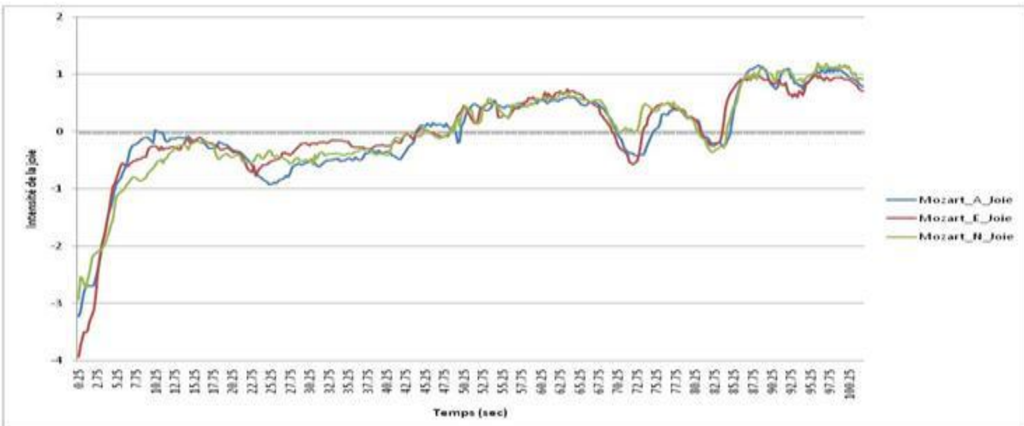



Figure 3. Averages of the three dynamic judgments obtained during the task of evaluation of the emotion expressed by music, here with the Joyful Activation dimension.

Further analysis is ongoing and we also plan to use the typology of the musical structure and the impact of acoustic parameters to predict these judgments using a Granger Causality method.

## 1.4 Evaluation of the qualitative judgment of felt emotions, listening individually and as a group and investigating rhythmic entrainment.

| 1.4a Testing subjective measurements that assess Quality of Experience(QoE) for a non-live setting: Physiological measures |
|---|
| **QUB** |

**Abstract**
In order to inform which are the specific physiological measurements that are shared between individuals (if any) while listening to music, we have designed an on-going experiment that collects GSR and HR data whilst participants listen to randomly chosen songs. Songs have been grouped in four classifications, which were chosen to elicit the four combinations of high and low valence and arousal. Participants were asked to complete a brief self-report questionnaire after each song, rating their emotional response. This experiment has been implemented in Dublin, New York, Genoa and will be installed in Bergen from the start of December 2011. Preliminary results show agreement between the song emotional classification and the self reports, and several features extracted from the physiology are showing significant differences between classifications.

**Method**
This expanded ideas that were at an early stage in D2.1.
The aim of this study is to determine what are the relationships between the properties of an excerpt of music (dynamics, rhythm, emotional intent, etc), the self-reported emotional response, and the ANS response, as measured through features extracted from GSR and HR. In order to achieve this, an experiment was designed and implemented as a terminal installation to be presented in public venues, in order to build a large database of physiological and self-report data. The experiment at the Science Gallery – Dublin lasted for three months (June-August 2010), having nearly 4000 participants and over 12000 listening samples. The music selection included in its 53 excerpts a wide variety of genres, styles and structures, which were divided in four groups that were chosen to elicit high valence, low valence, high arousal and low arousal.
The participants' age ranged between 10 and 80, and the majority (67.3%) were under 30 years old. Gender was divided in 53% female, 47% male. When asked about their nationality, 62% stated to be Irish versus the remaining 38% who declared themselves as nationals from a different country. 61% stated not having a musical background. Regarding the musical genres the participants declared to listen regularly, the results were the following: Rock 23%, Pop 20%, Classical 12%, Dance 12%, World 9%, Hip-Hop 9%, Jazz 8%, Traditional Irish 6%, None 1%.

**Results**
The emotional content classification of the songs was compared with the participants' self-report answers to the positivity and tension questions. Analysis of variance results showed agreement with the proposed classification for both dimensions, valence ($F(3,7697) = 91$, $p<.001$, $w=0.19$) and arousal ($F(3,7697) = 148$ $p <.001$, $w=0.23$).
The physiological data was pre-processed in order to detect and remove artefacts and electrical noise. From this stage, 85.6% of the GSR signals were evaluated as fit for the analysis (with none or minimum artefact presence), versus an 84.4% of the POX signals. If we look at cases that had both GSR and POX signals above the tolerance levels, only a 73.6% of the total shared this characteristic.
POX signals were converted into Heart Rate and Heart Rate Variability vectors, and GSR signals were divided into their Tonic (LF) and Phasic (HF) components. At the time of this report, only a few exploratory descriptive measures have been calculated from the physiological vectors. Such as: mean, standard deviation, gradient, integration, linearity and frequency analysis, among others.

Nevertheless, statistical analysis using both parametric and non-parametric tests, as well as principal component analysis, already indicate the presence of some salient features both in GSR (Phasic mean, Tonic gradient) and HR (HRV mean, HRV freq. components) channels, that differentiate emotion classification of each group of songs (high valence, low valence, high arousal and low arousal). Moreover, these features have a significant correlation (Spearman, $p<.001$) with the self-report questions of engagement, positivity, activity, tension, likeness and familiarity, each feature having a significant correlation with at least 3 questions.

**Evaluation**

Due to the public nature of this study, work has mainly been focused in improving the acquisition of signals, and the algorithms that correctly identify and remove noise and artefacts. Any unaccounted variation at this stage can impact the validity of the statistical tests that use physiological measurements. It is important to point out that with the current sensor design, which requires no assistance and can be used by audiences briefed with short instructions, we are obtaining almost ¾ of valid signals. This has to be taken into account when calculating group sizes for experiments that require physiological sensing of audiences.

The analysis of the physiological measures shows high levels of dispersion between participants for the same feature, which seems to indicate that large sample sizes need to be maintained for future experiments. Furthermore, a significant amount of the participants presented little to no variation in the features extracted from GSR.

Next steps in the analysis will be focusing on additional physiological descriptors, multimodal analysis of the dataset, looking at temporal changes (versus the current whole song approach) and measures of correlation and entrainment with musical features.

| 1.4a Testing subjective measurements that assess Quality of Experience(QoE) for a non-live setting: Explicit dynamic judgments of felt emotions |
|---|
| **UNIGE-CH** |

**Abstract**

In this experiment we sought to replicate the continuous judgment techniques used in 1.1 and 1.2 in a non-live setting for felt emotions. The results show that though there is slightly less agreement between the listeners concerning their felt emotions, reliability measures are still remarkably good for the great majority of pieces.

**Method**

The aim of this study was to investigate how different versions of the same piece affect the listener in terms of his/her subjective feeling of emotion and entrainment.

To this end, 9 pieces for solo violin were recorded at UNIGE-CH in September of 2010 with the help of a professional violinist. Each piece was performed in 3 different styles: academic, emphatic and natural. The academic version consisted in a very sober rendition of the piece performed with no emotional content at all. The emphatic version on the other hand was very emotionally expressive and this was conveyed through exaggerated temporal and dynamic effects. The natural version was the way the violinist usually performs these particular pieces in concert or at home.

The 27 resulting pieces were then used as stimuli in a judgment paradigm where we asked participants (N=62) to rate continuously how they felt while listening to the piece using a Flash interface in our lab. Participants moved the cursor up to indicate that the intensity of the felt emotion increased and lowered it to indicate that it had gone down, if left at the bottom it indicated that they felt nothing at all for that particular piece thus drawing a trace on the screen as time, and the music, unfolded.

In condition a) of the experiment half the participants continuously rated how "moved" (i.e. emotionally affected) they felt in reaction to the piece first and later answered how strongly they had experienced the nine emotional dimensions of the Geneva Emotional Music Scale, GEMS, using the sliders. In condition b), the other half of participants first rated continuously how strongly they had experienced one of the nine GEMS dimensions and then answered how moved (affected) they had been by the piece with a single slider.

In either case, *Affect* and *GEMS* conditions were both followed by 12 questions about Explicit Entrainment, i.e. self-reported subjective feelings of being entrained by the beat of the music:

| Item | Question | English translation |
|---|---|---|
| **entrain1** | *...physiquement stimulé* | Physically stimulated |
| **entrain2** | *...envie de danser* | Felt like dancing |
| **entrain3** | *...senti plein d'entrain* | Felt entrained/driven |
| **entrain4** | *...physiquement énergisé* | Felt energized |
| **entrain5** | *...envie de bouger* | Felt like moving |
| **entrain6** | *...senti animé* | Felt animated |
| **entrain7** | *...physiquement excité* | Physically excited |
| **entrain8** | *...senti le rythme dans votre corps* | Felt the rhythm in body |
| **entrain9** | *...corporellement agité* | Bodily agitated |
| **entrain10** | *...envie de battre le temps, le tempo ou le rythme* | Felt like beating time, tempo or rhythm |
| **entrain11** | *...ressenti vos propres rythmes corporels changer* | Felt one's own bodily rhythms change |
| **entrain12** | *...senti votre corps résonner avec la musique* | Felt one's own body resonate with the music |

And at the end of the experiment participants filled in a detailed musical preferences questionnaire assessing their expertise and listening preferences as well as Baron-Cohen's Empathy Questionnaire (Baron-Cohen & Wheelwright, 2004).

**Results**

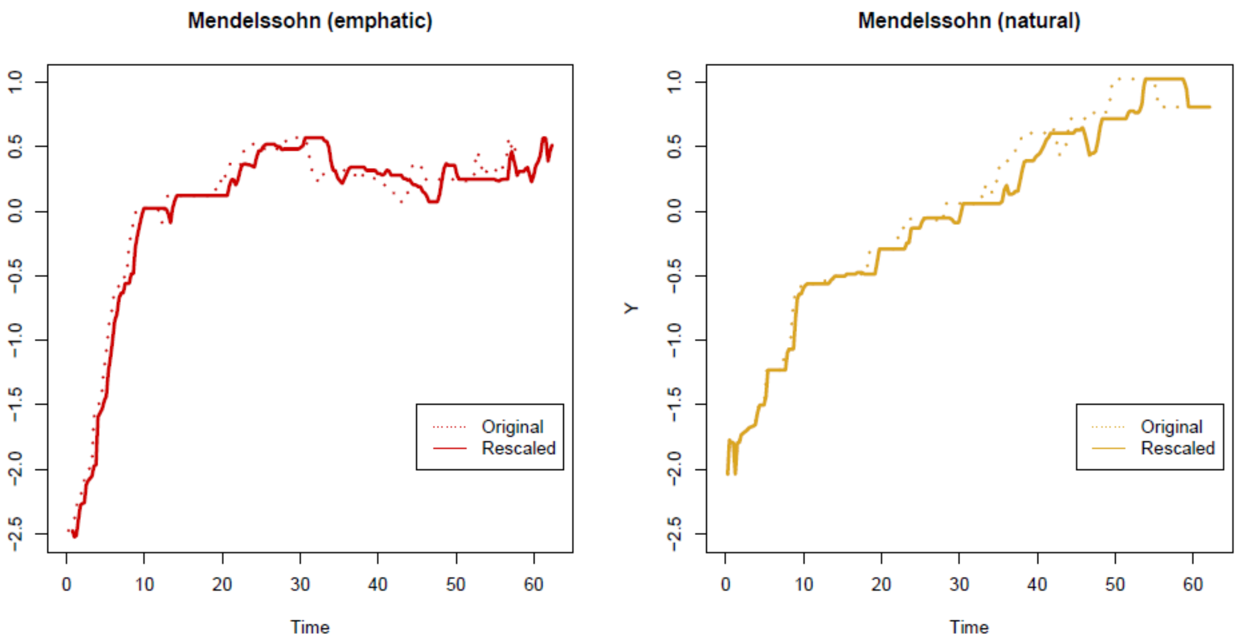
There was a surprising amount of agreement between participants despite the fact that they had to rate how they felt as can be seen from the Cronbach alphas on the continuous judgments below.

This makes us confident that the technique of continuous judgments is meaningful and reliable both for the rating of the intensity of recognized emotion in the music (experiments 1.1 and 1.2) and felt emotion in the listener (current experiment).

| Piece | Condition a) felt affect | Condition b) felt GEMS |
|---|---|---|
| Bach Academic | 0.90 | 0.83 |
| Bach Emphatic | 0.89 | 0.92 |
| Bach Natural | 0.92 | 0.93 |
| Beethoven Academic | 0.58 | 0.85 |
| Beethoven Emphatic | 0.58 | 0.76 |
| Beethoven Natural | 0.85 | 0.89 |
| Franck Academic | 0.77 | 0.92 |
| Franck Emphatic | 0.58 | 0.95 |
| Franck Natural | 0.77 | 0.95 |
| Gluck Academic | 0.92 | 0.48 |
| Gluck Emphatic | 0.93 | 0.8 |
| Gluck Natural | 0.95 | 0.76 |
| Massenet Academic | 0.89 | 0.79 |
| Massenet Emphatic | 0.88 | 0.77 |
| Massenet Natural | 0.91 | 0.93 |
| Mendelssohn Academic | 0.90 | 0.94 |
| Mendelssohn Emphatic | 0.78 | 0.96 |
| Mendelssohn Natural | 0.96 | 0.94 |
| Mozart Academic | 0.91 | 0.95 |
| Mozart Emphatic | 0.93 | 0.93 |
| Mozart Natural | 0.91 | 0.94 |
| Schumann Academic | 0.87 | 0.84 |
| Schumann Emphatic | 0.93 | 0.53 |
| Schumann Natural | 0.97 | 0.91 |
| Sibelius Academic | 0.93 | 0.71 |
| Sibelius Emphatic | 0.94 | 0.70 |
| Sibelius Natural | 0.87 | 0.62 |

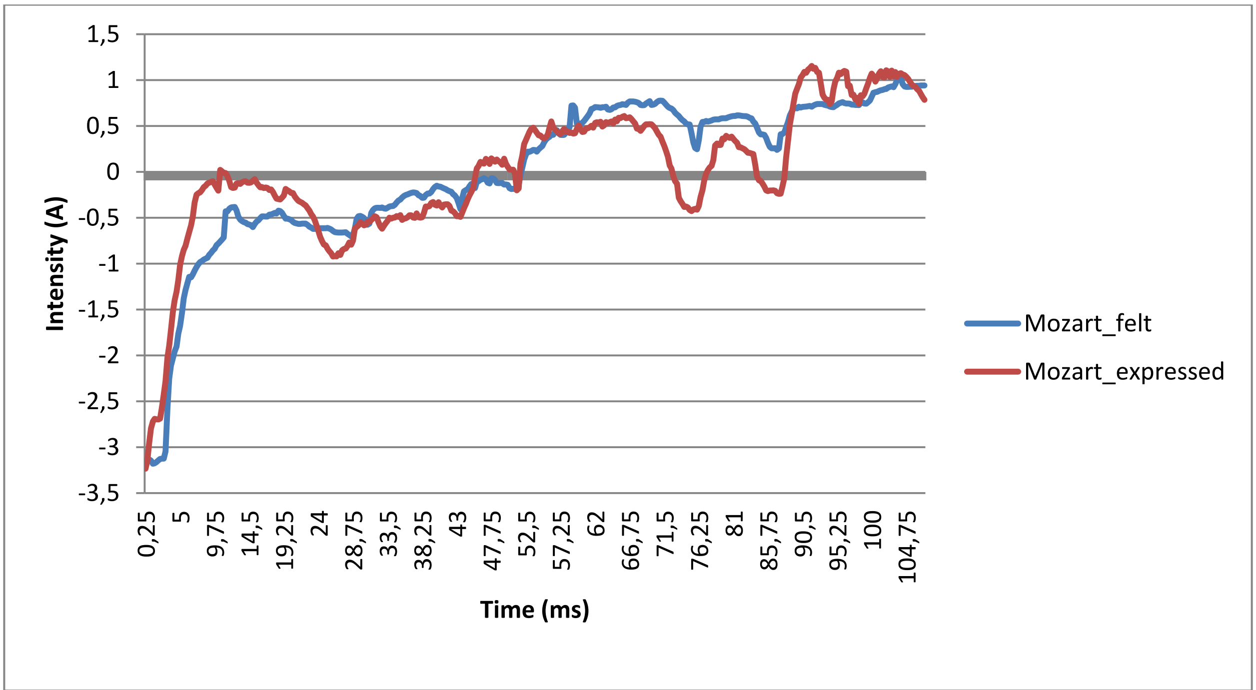
The experiment also revealed the usefulness of the entrainment questionnaire which, after performing a factor analysis, appeared to measure only one factor regardless of the piece and the style played and explained 80% of the variance.

UNIGE-CH is currently working on the rescaling of the tracings for the different versions of the same piece/stimulus to the academic version so as to be able to compare judgments according to style. This is being done through the use of markers for each measure and each beat for every piece which are then used to shift the tracings as needed as can be seen in the figure below:




## Evaluation

Analysis of the data is still ongoing and a comparative work between judgments of expressed and felt emotions for the same stimuli is planned. A foretaste of this work can be seen in the figure below where the correlation between the judgments is of $r = .82$ $(p < .001)$.



| 1.4b Testing subjective measurements that assess Quality of Experience(QoE) in a live setting. |
|---|
| **QUB, UNIGE-CH** |

**Abstract**

The Siempre DoW asks: "how can we measure the quality of audience experience"? To answer the question, we need measures that capture distinctions between the types ofsubjective experience that audience members may have. The literature points to a wide variety of factors that are likely to contribute to the entire "quality of experience": examples are emotion, social, attentional/flow, sense of refreshment, etc. These have been integrated into a questionnaire. Drawing on another literature, we developed a continuous self-report mechanism for use by audience members in a live concert scenario. A experimental paradigm was developed to test the measures. Audience members attended a concert consiting of two performances which were designed to produce contrasting engagement, each lasting 20 minutes. The participants were 19 informed volunteers of ages 18-24 (mostly undergraduate psychology students). The questionnaire showed significant differences between the two performances in all categories, confirming that it does capture differences in experience. The continuous self-report mechanism was not effective, because it was rarely used. The questionnaire is clearly a useful instrument. The problems encountered with the continuous self-report mechanism suggest modifications that may allow it to be more useful.

**Method**

This expanded ideas that were at an early stage in D2.1.

**Aim**

The aim was to pilot a battery of measures in a context where we expected audiences to have contrasting types of experience. To produce different experiences, we arranged a concert involving two contrasting types of performance, one which people who responded to the advertisement would be expected to like, the other of which is a very different and specialised taste. Three types of measure were used: a questionnaire, physiological measures, and a continuous self-report device. Video recordings were also made so that visual cues to quality of experience could be explored.

**Apparatus**

Physiological Measures: The sensor configuration developed in experiment 1.4a above was used to measure heart rate, heart rate variability and galvanic skin response.

Continuous Quantitative Response: A slider device was used to let users report when they were not enjoying the performance. The choice of a negative scale was because users are only likely to attend to a secondary task (such as pressing a cursor) when their engagement in the music is low. A spring mechanism was used so that participants could register the slider's position without having to look or concentrate.

Questionnaire

A questionnaire was developed to cover aspects of experience which, on the basis of the literature, seemed likely to contribute to audience experience. It contained 60 items covering 8 areas: emotion, attention, relationship to the performer, physical reactions, presence, refreshment, active participation, and social dimensions. The full questionnaire is in Appendix 2.

**Procedure**

19 participants were recruited for a musical performance in the Sonic Arts Lab in the SARC building in Belfast. Recruiting was via notices in the psychology and Sonic Arts buildings advertising a free performance of traditional music. Two semi-professional bands took part. In the first half of the concert, *Ni Pagani* played traditional Irish music. In the second half, *Sonic Autopsy* played experimental jazz. Each hlaf lasted apprioximately half an hour.

All of the participants completed the questionnaire twice, once at the end of each half. In addition, twelve had physiological measures taken and used the continuous measurement response; six had only physiological measures taken, and six more used the continuous measurement device. Selected portions of the audience were also recorded on video to capture visual cues to quality of experience.

**Results**

The exercise exposed problems in two areas. There were problems transferring the physiological recording techniques developed with single users in 1.4a to the multi-user scenario, and the recordings were not usable. The continuous recording technique was problematic for a different reason: users almost never pressed the slider. The likeliest explanation is users felt it was socially unacceptable to make what was in effect a negative comment on the performance. Work is in progress on two ways of overcoming the problem. One is to give slider users training before going into a test sitution. The other is to conceal the box, so there is no question of being seen to criticise.

Video data is being analysed. Impressionistically, there seem to be passages where the audience members show contrasting levels of engagement. Experiments with naïve raters will verify whether they can be reliably identified. If so, it opens the way to explored the signals that make them discriminable. Gaze appears to be a likely source of information.

The major goal was to pilot the questionnaire, and data show that it did indeed capture robust differences between the experiences generated by the two performances. Table 1.4b.1 gives a convenient summary. The results can be taken on three levels.

Table 1.4a.1 – Effects of performance type on QoE factors

| Concept | No. Items | t | df | Sig. (p-value) | Higher in | Alpha |
|---|---|---|---|---|---|---|
| Emotion | 12 | 4.419 | 16 | .000 | T | .831 |
| Social | 6 | 2.168 | 16 | .046 | T | .806 |
| Attention/Flow | 12 | -2.904 | 16 | .010 | C | .769 |
| Physical reations | 5 | .932 | 16 | .365 | T | .754 |
| Presence | 5 | 2.170 | 16 | .045 | T | .867 |
| Refreshment | 5 | 2.599 | 15 | .020 | T | .857 |
| Active participation | 4 | 4.255 | 15 | .001 | T | .677 |
| Performer | 8 | 4.429 | 14 | .001 | T | .810 |

At the most global level, the data confirm that there was something that could be called an audience response. It was not the case that each individual responded in his or her distinctive way. The significant differences indicate that for each scale, most of the audience showed the same direction of effect. Of course, we do not know whether the cohesion was a product of listening together. To test that, we would need data from a comparable set of people responding individually to comparable pieces.

With regard to the nature of the experience, the results confirm the suspicion that the audience responses evoked by different musical performances involve a considerable variety of attributes that are at least conceptually distinct. One might have thought, for instance, that scales such as renewal – which involve feeling that the world was more beautiful, life was better, and problems were less pressing – would be too esoteric to show any systematic pattern. In fact, the only scales not to show effects were among the most concrete – involving self-reported physical effects such as heart rate, posture, and breathing. To say that the scales are conceptually distinct is not to say that they capture independent factors. To test that, we would need to explore responses to music chosen to engage people, but made them feel (for instance) that the world was less beautiful, life was worse, and problems were more pressing (music by Steve Reich, Nick Drake and Jim Morrison spring to mind). We would need to ask whether changing that aspect of the music brought with it changes in others – for instance, we would obviously expect that it would affect some of the emotions. Only through that kind of study can a picture be built of the way different aspects of audience response cohere.

With regard to the questionnaire, the fact that such clear differences emerge with a small sample confirm that it is a useful instrument. It would be useful if the number of questions could be reduced, and inter-scale correlations (shown below) give preliminary indications that they may be. The presence, social and renewal scales in particular seem closely related. In contrast, the attention scale seems quite distinct from the others.

| | Presence | Social | Renewal | Emotion | Mirroring | Physi-ology | Per-former | Attention |
|---|---|---|---|---|---|---|---|---|
| Presence | | 0.709 | 0.633 | 0.591 | 0.606 | 0.398 | 0.529 | *-0.031* |
| Social | 0.709 | | 0.683 | 0.634 | 0.602 | 0.581 | 0.548 | *0.094* |
| Renewal | 0.633 | 0.683 | | 0.646 | 0.535 | 0.646 | *0.268* | *0.107* |
| Emotion | 0.591 | 0.634 | 0.646 | | 0.59 | 0.37 | 0.685 | *-0.264* |
| Mirroring | 0.606 | 0.602 | 0.535 | 0.59 | | *0.265* | *0.311* | *-0.099* |
| Physiology | 0.398 | 0.581 | 0.646 | 0.37 | 0.265 | | *0.251* | 0.355 |
| Performer | 0.529 | 0.548 | *0.268* | 0.685 | *0.311* | *0.251* | | *-0.074* |
| Attention | *-0.031* | *0.094* | *0.107* | *-0.264* | *-0.099* | 0.355 | *-0.074* | |

However, inferences from correlations need to be handled with extreme caution until the questionnaire has been applied to a wider variety of performances: some correlations in this sample clearly are an artefact of the fact that, for instance, the music that engaged people evoked positive emotions.

**Evaluation**

The experiment has achieved two main kinds of progress. It has introduced the experimental paradigm of a concert involving contrasting performances, where the contrast is chosen to produce radically different experiences in audience members. It has also provided a questionnaire which captures multiple aspects of audience experience. These provide a basis on which refinements can be built, including improved versions of the instrumentation that was used in the experiment.

# 2. EXTRACTING VALUABLE FEATURES FROM AUDIO AND INSTRUMENTAL GESTURES

Within the context of the SIEMPRE project, issues such as entrainment, inter- and intrapersonal synchronization, and musical leadership can be clarified and analyzed through the study of low-level characteristics of the captured performance, namely the audio produced and the instrumental gestures performed by the musicians in order to produce the aforementioned sound. Furthermore, an accurate note-level alignment between the performance and the score it is based on can provide a steady reference by which the performance is measured and compared.

In this first series of proposed experiments, the goal is to study synchronization in terms of intonation (realization of pitch accuracy) and tempo adaptations for the string ensemble scenario. This will be carried out using audio recordings as well as motion-captured instrumental gestures from each performance; the performance is time-aligned to its relevant score using state-of-the-art algorithms and computational tools in order to measure, at the note level, differences between performances in different experimental cases. Through the analysis of these differences we can extract quantitative features that characterize the synchronization, entrainment, and leadership aspects of the performance.

## 2.1 Analysis of intonation adjustments among violinists

| |
|---|
| 2.1 Observing and analyzing the way violinists adjust their tuning while performing in an ensemble to provide information on detecting the functional relationships (i.e. leadership) within the ensemble. |
| UPF |

**Abstract**

In a string ensemble (such as a violin duet), intonation is one of the most important factors that characterizes the cohesion of the ensemble; since the violin is a fretless instrument, the intervals between consecutive notes are not fixed to equal temperament, and therefore the effect of the relative tuning between the musicians of the ensemble is immediately audible and equally important to, say, timing. We wish to test, establish and test a set of methods which are capable of measuring the synchronization of intonation adjustments for the preliminary case of a violin duet. The experimental material used originates from a set of preliminary duet recordings carried out at UPF in September 2010, with the collaboration of IIT, UNIGE and SARC.

The method was described in D2.1 – essentially, after a set of processing steps (F0 extraction, score alignment, temporal warping et cetera) the intonation adjustments are extracted for each performer and different measures of correlation/interdependence are applied on the two time series.

A technical description of results has been given in:

- *Panagiotis Papiotis, Esteban Maestre, Marco Marchini, Alfonso Peréz* – Synchronization of intonation adjustments in violin duets: towards an objective evaluation of musical interaction (DAFX 2011 proceedings).

A more holistic view on the methodology can be also found in:

- *Panagiotis Papiotis, Marco Marchini, Esteban Maestre and Alfonso Peréz* – Measuring ensemble synchrony through violin performance parameters: a preliminary progress report (Social Behavior in Music workshop, INTETAIN 2011 proceedings).

We tested a number of interdependence measures before concluding on the measure that provides the best results, and the final chosen measure appears to validate the hypothesis that violinists are influenced by each other's intonation when performing together, at least for the simple cases of non-professional musicians. However, it has been seen that the coupling strength is dependent on a multitude of factors; namely the complexity of the piece, the skill of the violinists, and the harmonic relationship between the two performed melodic lines. In order to obtain clearer separation between solo and duet recordings and to study this synchronization phenomenon without coloration from the scores, it is necessary to push towards two main improvements: First, the use of F0 contours as the only extracted feature does not convey the real phenomenon well enough, since the pitch perception of real instruments such as the violin is strongly related to psychoacoustics factors such as the loudness and timber of the produced sound. To this end, an objective measurement of harmonic consonance/dissonance has to be included as a feature, to approach more the human perception of pitch and intonation. Second, in order to make the coupling detection independent from the performed scores, it is necessary to post-process the scores with an algorithm that analyzes the intervals between the two violins, and adjusts the expected pitch (or the lack thereof) according to the harmonicity of the interval; this way, very harmonic intervals between the two melodic lines will greatly penalize 'bad' intonation, while inharmonic intervals will contribute much less to the coupling strength.

Analyses of the recordings show that professional, skilled musicians demonstrate an impressive accuracy in reproducing the same intonation, with little difference between solo and joint performance. Standard interdependence measures (linear and rank correlation) as well as more advanced measures (mutual information, Granger causality) failed to provide significant differentiation between the solo and joint performances. However, measures borrowed from computational neuroscience (nonlinear coupling detection) did manage to show greater differentiation between the two experimental scenarios, mainly for the case of amateur musicians (see following figures 2.1.1 & 2.1.2).

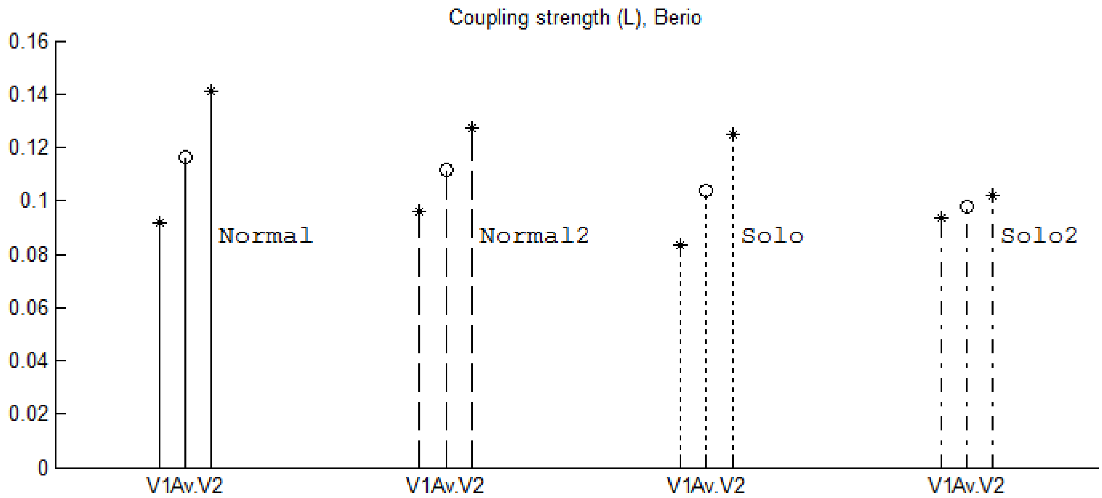



Figure 2.1.1 – Overall coupling strength for two instances of Normal (joint performance) and solo recordings of a contemporary duet, amateur musicians.

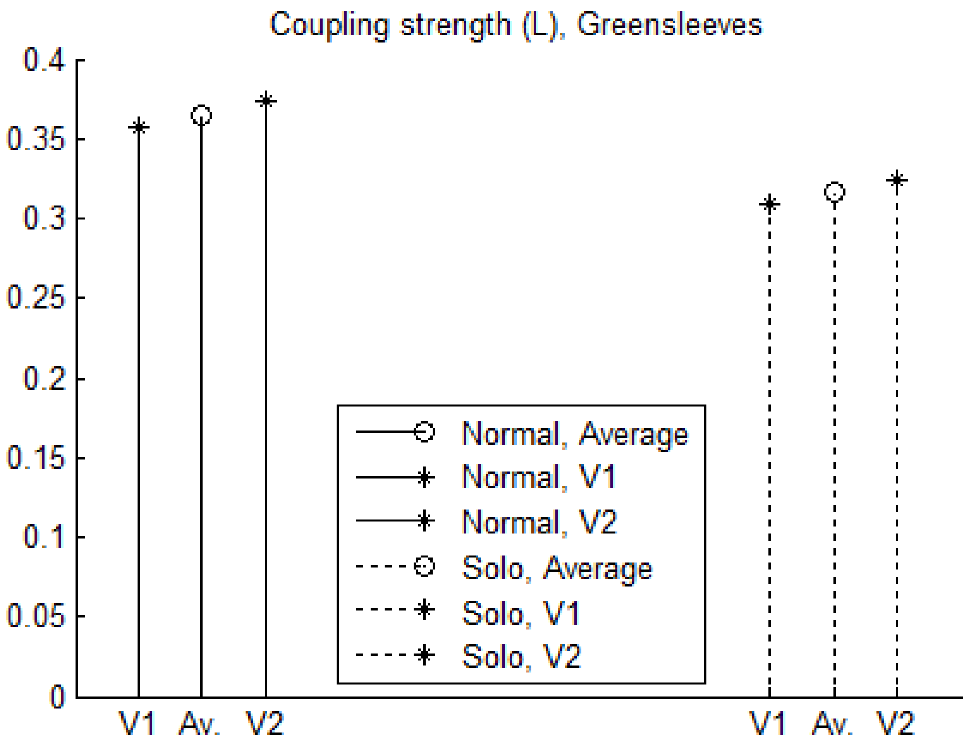



Figure 2.1.2 – Overall coupling strength for Normal (joint performance) and solo recordings of a unison melody duet, amateur musicians.

## 2.2 Synchronization in violin duets regarding adaptation to tempo changes

| |
|---|
| 2.2 Phase correction and period correction for a single musician who is listening to a metronome reference, as well as two musicians in different experimental set-ups (both listening to a metronome, musician_a listening to a metronome and the musician_b to the musician_a, with and without visual contact etc.) |
| UPF |

**Abstract**
Tempo synchronization is a key aspect of entrainment in musical performance. An understanding of its constituents mechanisms will provide new insight to the understanding and interpretation of recorded performances. Accurate measures could provide not only indications about the direction of influence in synchronization, and therefore leadership (in terms of tempo), but also about the strength of inertia/impulse that each subject is exerting on the group.

According to Repp (Bruno H. Repp, *Processes underlying adaptation to tempo changes in sensorimotor synchronization*, Human Movement Science, vol 20, issue 3, June 2001) when trying to adapt to tempo changes from a metronome signal, the two main mechanisms are that of phase correction and period correction. In a pilot experiment we aim at reproducing Repp's results of a string instrument. Furthermore, assuming his model as an hypothesis, some parameters may have to be adjusted to suit better in the case of skilled musicians and in various setups. The musicians are asked to play a sequence of steady monotonic notes in the following setups:

1. a single violinist adapting his note rate to a click metronome with tempo change
2. one violinist adapting to another violinist with eye contact and without eye contact

For each the cases several combinations of beat per minutes, strength of tempo change and note groupings have been recorded. Here is a list of what was recorded during March 2011 at QUB and in April 2011 at UPF:

Solo recordings (120 BPM)
- Violinist 1, quarter notes
- Violinist 1, eighth notes
- Violinist 1, sixteenth notes
- Violinist 2, quarter notes
- Violinist 2, eighth notes
- Violinist 2, sixteenth notes

Solo recordings (90 BPM)
- Violinist1, sixteenth notes

Duet recordings, one metronome (120 BPM)
- Violinist 1 with metronome
- Violinist 1 with metronome, no visual contact
- Violinist 2 with metronome
- Violinist 2 with metronome, no visual contact

Duet recordings, two metronomes (120 BPM)
- Two metronomes
- Two metronomes, no visual contact

The method was as described in D2.1. Onset detection and BPM estimation will be performed on the metronome clicks as well as the performed notes acquired from the recording. Bowing gesture parameters (bow velocity, bow acceleration, bow displacement, zero crossing rate, etc) will be extracted using the Polhemus MOCAP system. From the comparison of these signals we will extract the response of each musician to tempo changes. We aim at measuring these correction mechanisms for a single musician who is listening to a metronome reference, as well as two musicians in different experimental set-ups (both listening to a metronome, musician_a listening to a metronome and the musician_b to the musician_a, with and without visual contact etc.)

The analysis of the data is still undergoing. The work has mainly been focus on collecting the data and obtaining an accurate synchronization of the audio with the Polhemus gesture data and on the processing of the data to precisely define the onset point of the bow strokes.

## 2.3 Feasibility study regarding the Polhemus motion sensors

| 2.3 Feasibility study regarding the Polhemus motion sensors |
| --- |
| UPF |

**Abstract**
The interaction among musicians during the performance is a key aspect of the SIEMPRE project. Concepts such as leadership, entrainment and synchronization among musicians can be clarified through the acquisition of detailed instrumental gestures, as they are crucial to accurate audio analysis at the note level through the alignment between recorded data and the musical score.

Furthermore, the instrumental gestures captured with the Polhemus sensors appear as the most accessible method so far for extracting mid-level information regarding bowing movement (such as bow transversal velocity or bow force) accurately aligned to note onsets & offsets, which is crucial in the search to define and study in detail the interaction-specific concepts described above (entrainment, synchronization, leadership). Through the use of the Polhemus MOCAP system, bowing gesture parameters directly involved in the generation of sound are acquired. The goal is to investigate the relevance of these parameters and their acquisition methods to the objectives of the SIEMPRE project, as well as the low- and mid-level features that can be derived from these parameters.

Method was partially described in D2.1 and is here expanded. We have used the data recorded from a pilot experiment that took place at UPF during September 2010 involving additionally IIT, UniGe and SARC. We have used resulting gesture descriptors from the Polhemus setup (see figure 2.3.2) to improve existing beat tracking algorithm. The addition of bowing gestures descriptors to audio descriptors results in a dramatic improvement of the beat tracker as shown in figure 2.3.1. The audio features extraction was augmented with the instrumental gestures and it was seen that instrumental gestures improve the accuracy of feature extraction algorithms especially for audio-score alignment and tempo estimation (beat tracking).

The conclusions have been drawn partially from the joint experience that took place at UPF during September 2010 involving additionally IIT, Unige and Sarc. A number of trials were made using the current Polhemus setup, for the case of violin duos, and even though the performers did in general show certain degree of adaptation to the wires, they would have preferred a wireless sensing system (such as the Qualysis system). Also, setups involving only two musicians appeared feasible, but in cases where more musicians are to be involved, using wired sensors is perceived as to affect the performance. One of the most important measures was the time needed for setting up the sensors and calibrating each of the instruments, resulting too long as to be used in real concert situations.

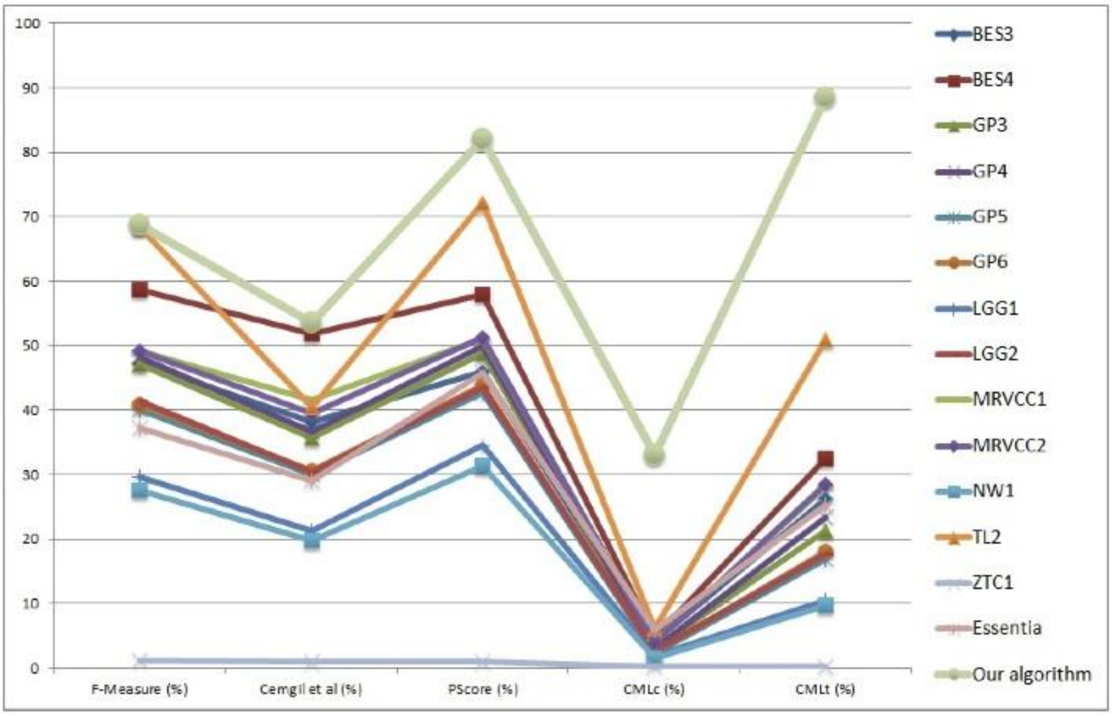



Figure 2.3.1: Graphical comparison of the tempo tracking algorithms submitted for MIREX 2010 and our algorithm.

Additionally recording a quartet would have required a longer ranger EMF emitter with the results of an increased problem of interference with ferromagnetic material eventually present in the room.

It can therefore be concluded that, although the use of Polhemus-captured instrumental gestures add a significant amount of accuracy in the extraction of audio features as well as the extraction of mid-level features derived from the gestures themselves, the improved accuracy does not in all cases merit the trade-off in intrusiveness and set-up times for the joint experiments.

Following our findings it was decided that, in the scope of joint experiments with partners, it is more convenient to acquire (using Qualysis) a sub-set of bowing gesture parameters with less accuracy, by following part of the method given in (E. Schoonderwaldt and M. Demoucron, "Extraction of bowing parameters from violin performance combining motion capture and sensors,"J. Acoust. Soc. Amer., vol. 126, no. 5, pp. 2695–2708, Nov. 2009.).

According to the Schoonderwaldt paper with a minimal setup (see figure 2.3.3) of 2 markers on the bow (one on the tip and one on the screw) and 2 markers on the instrument body (one on the scroll and one on the tailpiece) we will be able to compute an estimation of the following descriptors.

- Bowing transversal displacement
- Bowing transversal velocity
- Bowing transversal acceleration
- Bow-bridge distance (for this particular parameter, a measurement of the distance from the tailpiece marker to the bridge will have to be provided for each instrument).
- Skewness

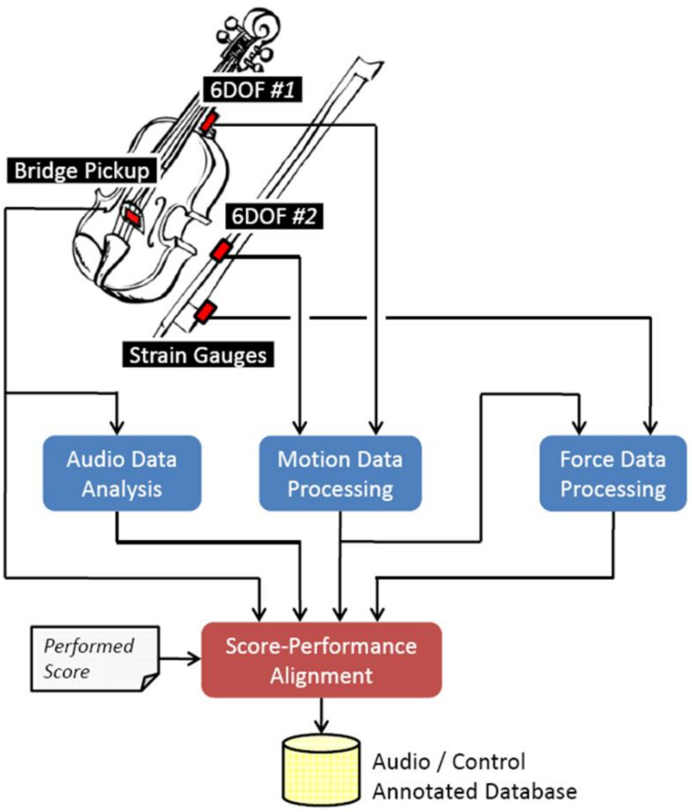



Figure 2.3.2: Polhemus sensors placement.

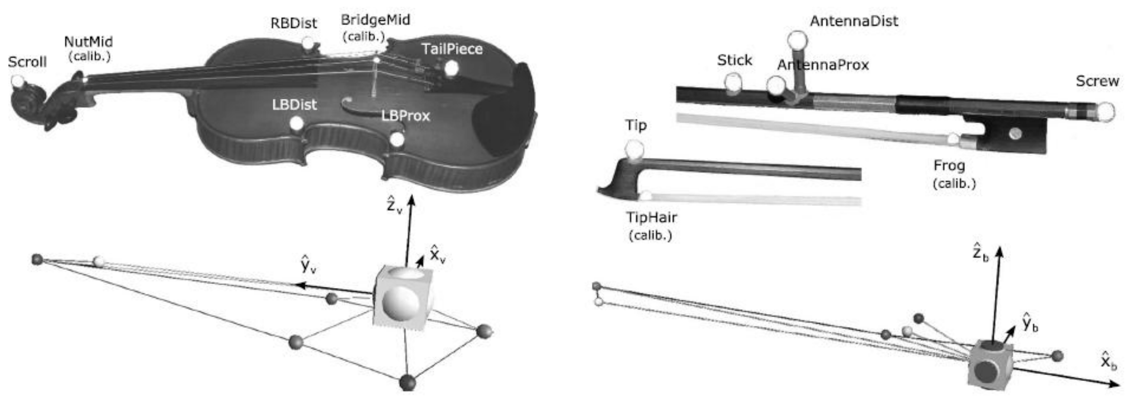



Figure 2.3.3: Qualisys Markers placement to get similar results as with the polhemus sensors.

Apart from these instrumental gestures, the high accuracy of Polhemus sensors allow us to estimate several things that Qualysis is not able to provide like string estimation (the string which is being played at each moment) or bow force (the pressure that the bow makes to the strings).

# 3. MUSICIANS' MOVEMENT ANALYSIS

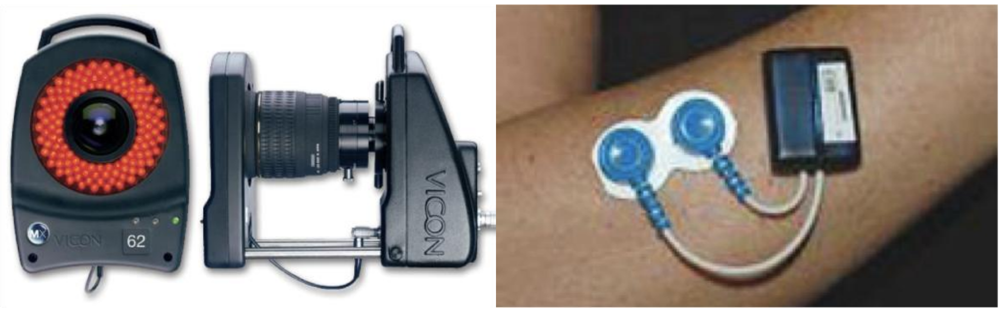
Here we aim to use a rather different approach by studying music orchestras or quartet in an ecological rehearsal scenario thus excerpting no particular interference on participant's behavior. Here, we will record violinists' bows and conductor's baton kinematics via an unobtrusive passive infrared optical 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. Furthermore we will record muscle activity to extract other parameters that are not measurable with simple kinematics, such as force and joint stability via muscle co-contraction. Measurement of muscle tension is commonly achieved using surface electromyography. Surface electromyography measures muscle activity by detecting the electrical potential that occurs on the skin when a muscle is contracted.

## 3.1 Pilot of EMG recordings in musicians

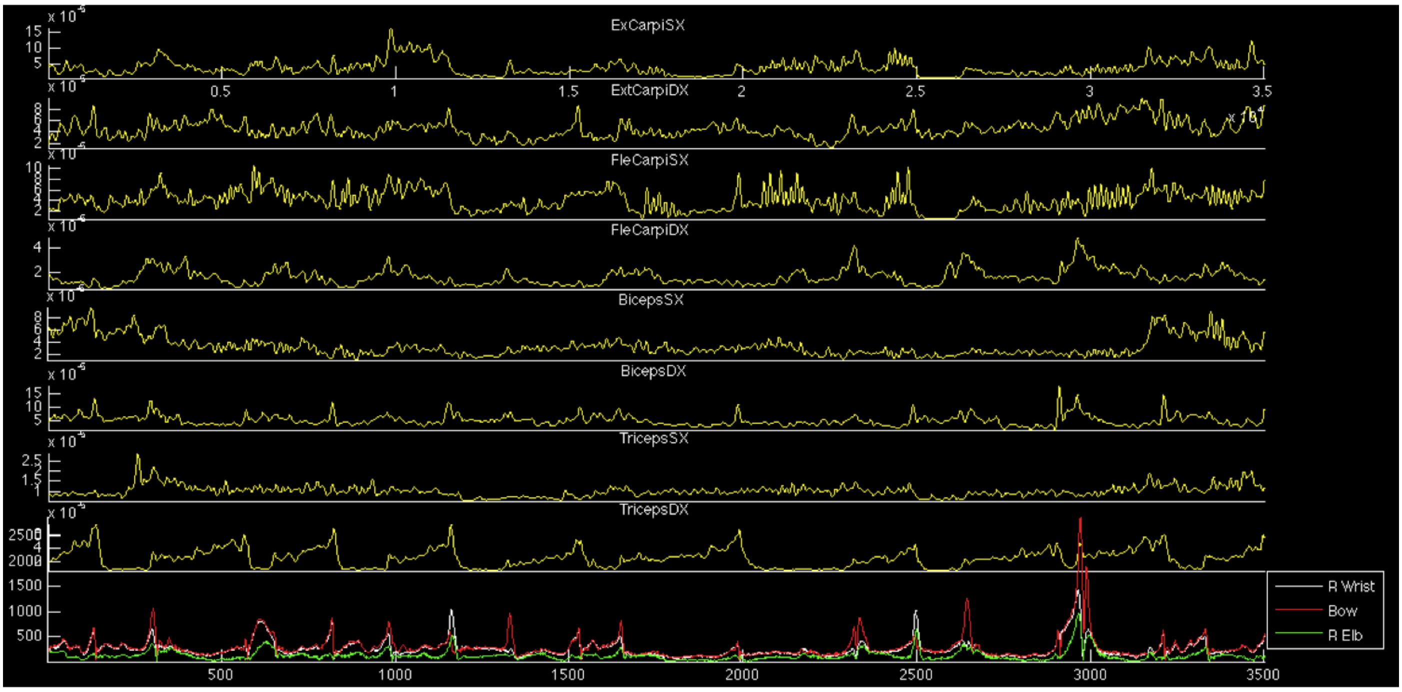
| 3.1 Defining EMG electrodes placement and comparison between emphatic and academic performances |
| --- |
| IIT |

**Abstract:** Musical expressivity can be inferred by measuring motor behavior (position data as in 3.5). However, subtle modification in musical dynamics might be difficult to detect with position data only. On the other hand, EMG data can better inform about the dynamical aspects of motor behavior and thus be a more sensible index of it. Expert musicians can indeed modulate motor control parameters in order convey very different styles and emotions. Here we asked a student musician to play the same piece at least 5 times, and by forcing two different expressive conditions. In the first they'll have to follow a metronome and reduce the expressivity (academic). In the second, they will have to use an emphatic expressivity (emphatic). Multi channel EMG can be used to extract motor synergies accounting for most of the variance observed in these two opposing behavior (academic versus emphatic). This pilot has a two-fold goal that is the definition of EMG electrodes placement and start to investigate the response pattern between emphatic and academic performances.

Method was partially described in D2.1. We recorded, from one student musician, the following muscles L/R Extensor Carpi; L/R Flexor Carpi; L/R Biceps; L/R Triceps; L/R Deltoids; L/R Tibialis; L/R Soleus; L Pectoralis. We also recorded upper torso kinematics, including L/R wrist, L/R elbow, L/R shoulder, head, C7, hip positions. The student played the same piece 5 times alternating academic and emphatic styles.




*3.1a. Vicon system for kinematic rcordings and EMG ZeroWire for muscular activity recording*




*3.1b. Example of data collected from 8 muscles in the student musician. Bottom graph shows the acceleration profile for right wrist, right elbow and bow.*

**Evaluation:** This pilot gave us important insight on the problems with these kinds of recordings. In fact, recording from several (up to 32) EMG channels on different individuals (up to 4) performing quite intense motor activity poses a series of technical challenges. Challenges are related to wireless communication and interferences due to fast body movements. However, most of these issues have been solved. Electrode positioning also depends on how many musicians are recorded at the same time and thus how many channels are available for each recording session. At present we’re still analyzing the data to extract muscle synergies that can explain the differences between academic and emphatic styles.

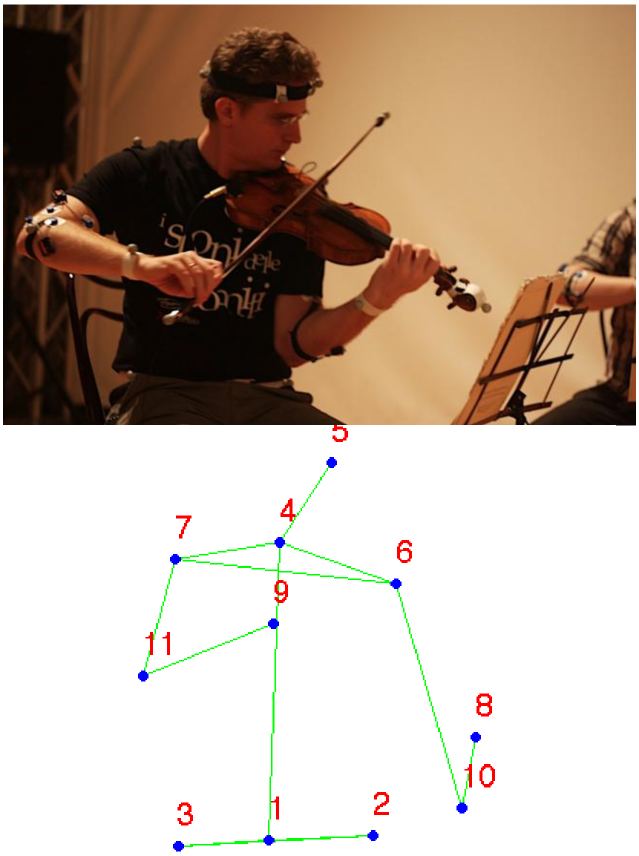
## 3.2 Quartet Preparatory experiments (MoCap)

| |
|---|
| 3.2 Identify the MoCap Qualisys markers needed to measure non-verbal social communication in a quartet; ensure the reliability of the measures; test the architecture for multimodal synchronized recordings; optimize the setup procedure.. |
| **UNIGE** |

**Abstract:** Musical expressivity can be inferred by measuring motor behavior (position data) (cf 3.1). However a number of studies considering mocap data focus on individual performance and on the analysis of expressivity. There is relatively less mocap-based studies of music ensemble. While triggering an increasing interest, analysis relies mainly on manual annotation of video. We therefore aimed at identifying a minimal set of markers for analysis of non verbal communication within the group.

**Method:** Method was described in D 2.1. The set of feasibility studies consisted in a total number of 3 experiments with students conducted in March, April and May 2011.

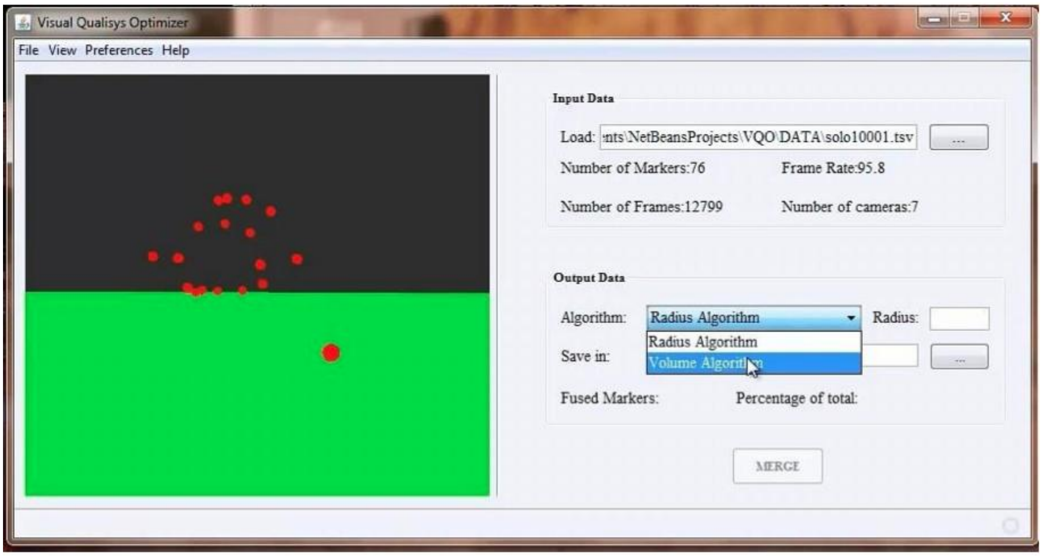
**Evaluation**: A total number of 16 Mocap Qualysis markers have been defined to study non-verbal communication within the quartet: 13 markers cover the main joints of the upper-body parts and are used to create a body-segment model of the musician, based on the Dempster's anthropometric reference (see figure below). The remaining 3 markers were used to monitor the violin and bow's moves (one marker on the scroll of the violin and two on the top and bottom of the bow).



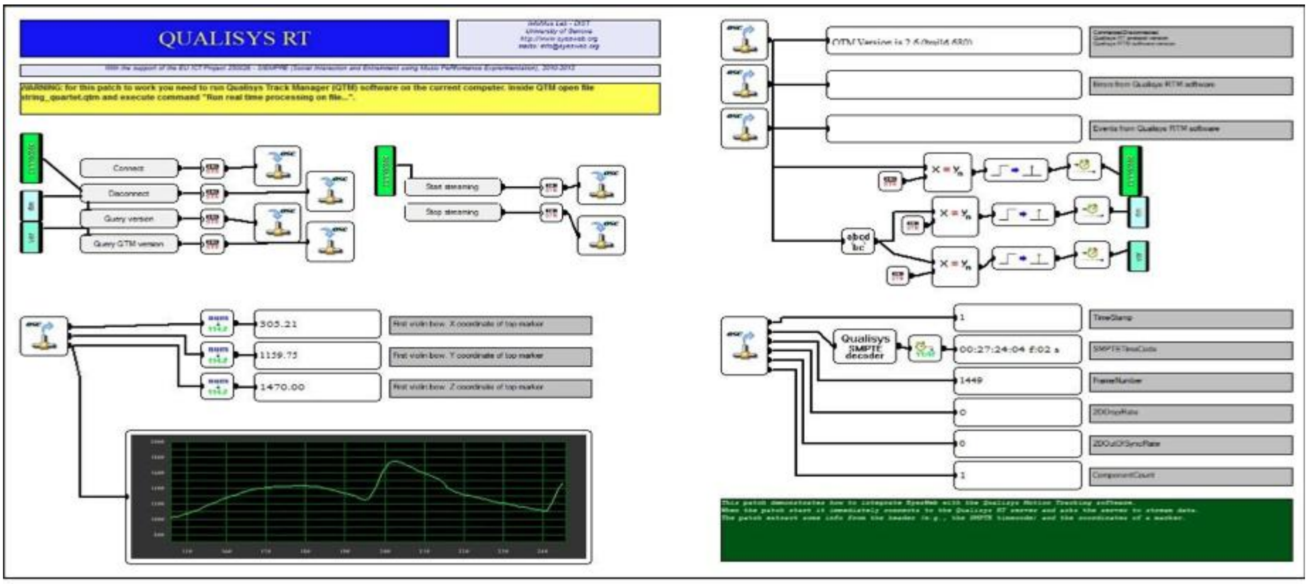

Snapshot of the Body-segment model of a musician created from the Mocap data (Cristiano Gualco, first violin of the Quartetto di Cremona).

A set of applications have been developed to tackle the issues mentioned above, in particular:
. A Java-based application was developed to ensure the reliability of the tracking process (e.g. markers can fail to be identified when they are occluded) to facilitate labelling with the Qualisys labelling software (QTM manager, see figure below)

. Real-time applications within the EyesWeb XMI software platform to synchronize and visualize Qualisys MoCap data together with video, audio and physiological data (see Figure below).




Snapshot of the developed Java-based application for the labelling of the Mocap data




Snapshot of the EyesWeb applications for real-time processing and visualization of multimedia data (including audio, physiological sensors and mocap data)

The procedure and the applications have been presented as a poster : R. Rovegno, M. Gagliardi, D. Glowinski, S. Ghisio, P. Coletta, and A. Camurri, Using the Qualisys MoCap System for Multimodal analysis of string quartet performance: a feasibility study, Intl ISSSM2011, University of Jyvaskyla (Finland), 8-18 August 2011.

## 3.3 Individual Vs Social behavior in music performance

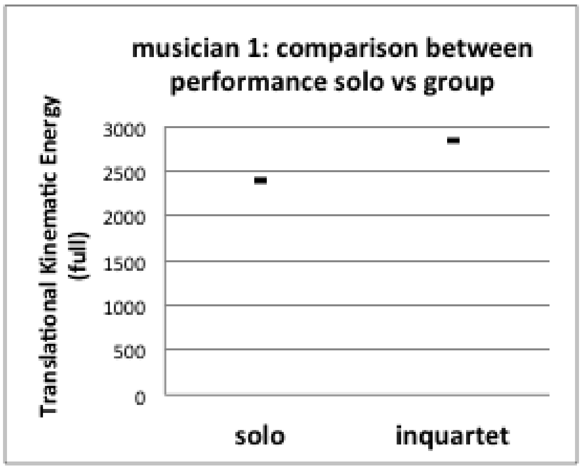
3.3 Which multimodal variables explain the difference between a soloist performance versus the

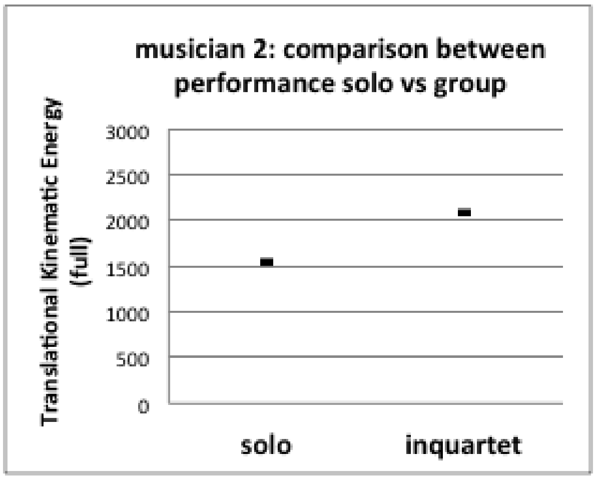
| same performance with accompaniment in an ensemble? |
|---|
| **UNIGE** |

**Abstract**: This study considers a basic issue of social aspects in music performance: the fact to play alone or with other players. In this respect, it can be considered a preliminary step for investigating detailed aspects of social interactions in music ensemble and in particular in a string quartet, such as leadership or group cohesion (see D 3.3.1)

**Method** :
Methodology was partially described in D 2.1. We've started from the mocap 3d position and audio data collected from two recording sessions of a professional string quartet (Quartetto di Cremona, see D 3.2.1 *Quartet Preparatory Experiment*), the first in July and the second in September 2011. We investigated which behavioral changes may characterize the difference between playing alone *versus* playing with others. Music piece was from the Schubert's *streichquartet*: a 2min fragment characterized by a variety of writing styles (isorhythmic parts, polyphonic phrasing with dialogic nature).
First violin and second violin were asked to play their part 7 times alone, and 5 times with the group. To disentangle possible effect of group performance on solo performance, first and second violinist had to perform 4 trials before and 3 trials after the group performance. For comparison between the two conditions, 5 trials from the solo performance were selected.
We ensure the reliability of the collected data by:
(i) having the same experimental design repeated at a distance of two months
(ii) investigating the quality of each performance as assessed by each musician through post-performance ratings (The full questionnaire is in Section 7, Appendix 1)
(iii) considering extraneous factors (personality and emotional mood) that may bias the performance: *personality* factors that may contribute to one's sensitivity of being in a group through BFI questionnaire and *emotional* mood of musicians for each session through PANAS questionnaires submitted before and after each experiment

**Evaluation**: We aimed at identifying a set of behavioral features accounting for the difference between *solo* vs *group* performance. Studies on non-verbal communication in music performance by Davidson et al. [2] and Dahl et al. [3] pointed out a number of attitudes and gesture (e.g., head nods, body sway, etc.). A first step, still under process, consisted in the manual annotation of video of the performance to retrieve the behavioral features indicated in literature
Preliminary results are reported.
A pilot quantitative analysis was conducted following approach by Camurri et al. [1] on the quantity of motion. We started from the body-segment model defined in *3.2.1 Quartet Preparatory Experiment* to compute the kinetic translational energy of musician 1 and 2 in each performance. For each musician and for each trial the sum of the translational kinematic energy of all their body segments was computed. With respect to vision-based approach adopted by [1], mocap data enable an in-depth analysis of motion both at the level of single limb and along the 3 axes.
A non-parametric *Wilconson* Exact test applied on the data from the July session revealed that energy was significantly higher when performing in a group rather than performing alone for musician 2, not for musician 1 (see figure below)




**Comparing the performance of the two musicians in the two conditions *solo* vs *group* considering the sum of the translational kinematic energy of the musician's body segments**.

Future work include the following directions
- Refine the analysis of kinetic energy by (i) including rotational movements (ii) considering the kinetic energy related to specific limbs of the upper-body (head, elbow, shoulders) (iii) detailing the component along which energy is higher (e.g., analysis of cello may require a different approach with respect to violinist as constraints are different and degree of freedom relates to upward-downward movements (*z* axis) more than on leftward/rightward and backward-forward movements) (iv) applying entropy methods to characterize the regularity of the kinetic energy overtime and (v) computing the kinetic energy on specific parts of the music piece (e.g, where *first violinist* has to explicitly coordinate the ensemble)
- Complete the kinetic analysis of body movements with audio analysis (e.g., is there any correlation between body kinetic energy and amplified contrasts between musical *nuances*?)
- Considering additional features referring to literature on analysis of communicative gesture in music performance [2,3] : (i) directionality/spread of trajectory (e.g., are violinist 1 and 2 tend to privilege specific directions when playing with other musicians such as the center of the quartet? (ii) Variations between opening/close postures [1] (iii) fluidity of movements
- Extend the analysis to another quartet such as the student's one

Perceptual experiments with partners UNIGE-CH and UPF are planned to confirm the reliability of behavioral features.

[1] A. Camurri, I. Lagerl ̈of, and G. Volpe. Recognizing emotion from dance movement: Comparison of spectator recognition and automated techniques. *International Journal of Human-Computer Studies, Elsevier Science*, 59:213–225, 2003.
[2] J. Davidson and J. Good. Social and musical co-ordination between members of a string quartet: An exploratory study. Psychology of Music, 30(2):186, 2002.
[3] S. Dahl and A. Friberg. Visual Perception of Expressiveness in Musicians' Body Movements. *Music Perception*, 24(5):433–454, 2007.

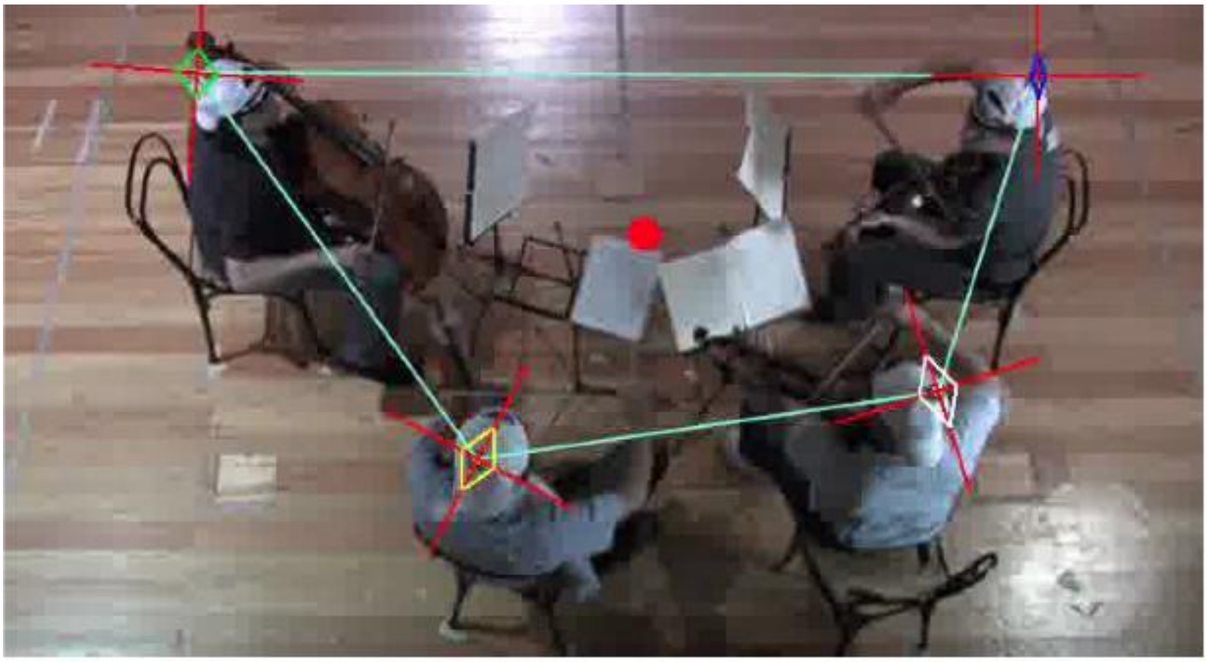
## 3.4 Series of experiments on synchronization and leadership

| |
|---|
| 3.4 Cues explaining synchronization and leadership in live performances of string quartets |
| **UNIGE** |

**Abstract:** A component of leadership, dominance, was investigated in the string quartet through the application of the Multi-Scale Fuzzy Entropy method. Results obtained on previous recording of the quartetto di Cremone revealed that dominance over others may be achieved through the regulation of individual and group's behavior complexity.

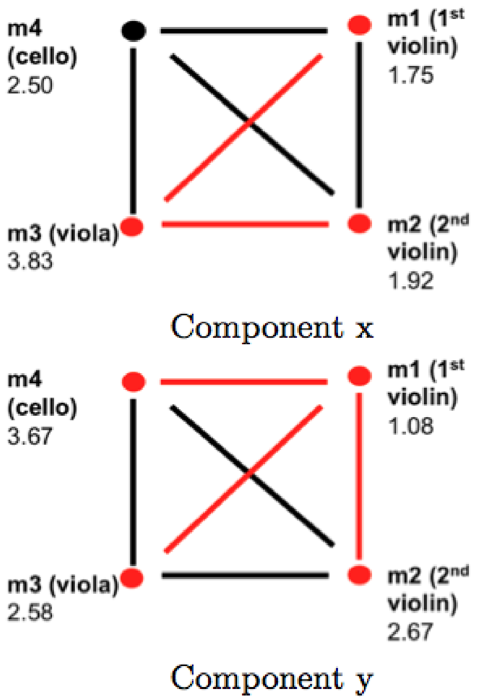
**Method**: This expanded ideas that were at an early stage in D2.1. The analysis concerns the time series data of the musicians' heads movements collected in a preliminary dataset of the Quartetto di Cremona while assessing the new data archive (see section 3.3.).

The heads of the musicians in the SQ form the four corners of the polygon (see Figure below). Thus, the algorithm for identifying the dominant member of the group took input data from the four vectors which denote the speeds at which the musicians' heads move, and also from one other vector, which denotes the speed at which the center of gravity of the polygon moves. The latter gives an approximate value for the group activity overall. Analysis of dominance was based on the theoretical framework of Multi-Scale Entropy (MSE), a non-linear technique initially developed by [1] to quantify the behavior complexity, i.e., the information expressed by the body movement dynamics over multiple time scales. A real-time implementation of the method was developed in the Eyesweb platform (see figure below, and demo at ftp://ftp.infomus.org/Pub/ftp-user-root/Siempre/quartetcomplexity.avi ).



Visualization of the polygon relating the four musicians' heads and the Complexity Index along the anteriorposterior (AP) and mediolateral (ML) directions (rhomb above each musicians'head).

**Evaluation**: Results confirm the hypothesis that the musician whose movements yield the lowest Complexity Index and the highest correlation with the group activity is, correspondingly, the dominant player. These two behavioral characteristics indicating dominance in a string quartet might be understood as follows: low complexity in the first violinist's behavior may increase its predictability for others. This also implies that real-time musical decisions, such as impulsive anterior-posterior (front-back) head movements to cue simultaneous full-ensemble entries attacks, for example, will be highly salient. The low complexity of the movements of the second violinist should also highlight his candidacy as an alternative dominant player. From this perspective, these results might be seen to confirm observations made by [2] in which leadership is assumed by the first violin in association with the second violin (as deputy leader). Correlations between data observed for the first violin and those observed for the polygon representing the SQ group activity, highlight that dominance requires the ability to regulate the activity of others [3].




Each node represents one of the four musicians and shows the sample average rank of its Complexity index (CI) for the three consistent conditions of performance: training, normal, overexpressive. The figures on the top and the bottom show the values for the x and y components of head movement respectively. Red lines indicate significant differences between the musicians (nodes).

These results have been published: Glowinski, D., Mancini, M., Rukavishnikova, N., Khomenko, V. and Antonio Camurri " Analysis of Dominance in Small Music Ensemble" in *Proceedings of the AFFINE satellite workshop of the ACM ICMI 2011 Conference*, Alicante Spain, 17-11 2011.

**Reference**

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

## 3.5 Pilot of an orchestra section

| 3.5.1 orchestra section pilot study: **Entrainment and leadership (conductor and musicians)** |
|---|
| **IIT** |

**Abstract:**
Non-verbal communication enables efficient transfer of information among people. In this context, classic orchestras are a remarkable instance of interaction and communication aimed at a common aesthetic goal: musicians train for years in order to acquire and share a non-linguistic framework for sensorimotor communication. To this end, we recorded violinists' and conductors' movement kinematics during execution of Mozart pieces, searching for causal relationships among musicians by using the Granger Causality method (GC). We show that the increase of conductor-to-musicians causal influence, together with the reduction of musician-to-musician coordination (an index of successful leadership) affects quality of execution, as assessed by musical experts judgments. Rigorous quantification of non-verbal communication efficacy has always been complicated and affected by rather vague qualitative methodologies. Here we demonstrate that the analysis of motor behavior provides an effective tool to quantify the rather intangible concept of aesthetic quality of music and non-verbal communication efficacy.

Method was as described in D2.1. No changes were done respect to D2.1.

A technical description of results has been given in a full paper submitted for publication in a journal.

**Evaluation:** In deliverable 2.1 we were depicting a roadmap for analyses and expected results. In the time being we completed the analyses and in September we submitted a full paper for consideration in a high impact journal.

| 3.5.2 orchestra section pilot study: **Modelling the orchestra behaviour for studying synchronisation and leadership (conductor and musicians)** |
|---|
| **UNIGE** |

**Abstract**:
Orchestra performance is particularly captivating cooperative phenomenon in social systems. The relationships among players and between players and conductor may be extremely useful in studying symmetrical (among peers) and asymmetrical (with a hierarchy) social interactions. This study proposes and validates a mathematical model of orchestra by assuming each player as a rhythmic unit and the whole orchestra as a network of rhythmic interacting units. This model will enable the study of social interaction with respect to (i) synchronisation, and (ii) functional role of the conductor.

**Experimental set-up and data set**
This study used a subset of the data set analysed in 3.5.1. Data refers to repeated orchestra performances conducted by both the regular orchestra conductor (OWN) and by another professional conductor (NEW). Data was segmented following the musical score and suitably pre-processed.

**Modelling the orchestra dynamics**
Playing in an orchestra requires cooperation due to a shared goal, division of roles and monitoring of progresses. This clearly implies coordination of movement making a collective rhythm emerge. Each player has her own rhythm, but through the interaction with the other players belonging to her section or to other sections, or the conductor, she modifies this rhythm toward the collective rhythm. Phase synchronisation is the more common and the simplest detectable form of entrainment among two or more signals, and in particular motion signal. The presence of rhythmic interacting units is an issue that was already faced by using the Kuramoto model.

The Kuramoto model is a mathematical model broadly used to study the emergence of synchronous behaviour in a population of coupled phase oscillators. The original model describes the dynamics of every oscillator taking into account both the characteristics of the single oscillator itself and the influence of the other oscillators linked by some connective arrangements with it.
The dynamics of each oscillator is generally governed by:

$$\dot{\theta}_i(t) = \omega_i(t) + \frac{1}{N}\sum_{j=1}^{N} K_{ij}(t) sin(\theta_j(t) - \theta_i(t)) \quad i = 1, ...N$$

where and $\omega_i$ and $\dot{\theta}$ are the rate of change of phase oscillator and the natural frequency of the $i^{th}$ oscillator, respectively. $N$ is the number of oscillators, $K_{ij}$ and ($\theta_j$ - $\theta_i$) are the coupling strength and the difference in phase between the $i^{th}$ and $j^{th}$ oscillators. The main assumptions of Kuramoto are: nearly identical oscillators, mean-field and time-independent coupling ($K_{ij} = K \ \forall \ i, j$), a very large number of oscillators, and unimodality and symmetry of the natural frequencies distribution $g(\omega)$.
This model was already successfully exploited in many different contexts ranging from physics to neurology. However, in order to make the Kuramoto model more suitable to study synchronisation in realistic situations such as, for example, neuronal networks or human being groups, some generalisation are needed.

**Fitting the model**

The phases of the z-component were computed by applying the analytic signal approach. As regarding the natural frequency of each player, due to the noisiness of the power spectrum, a fundamental frequency did not emerged clearly. Therefore, the tempo of music piece expressed as *rad/s* was selected as natural frequency. Indeed, the oscillatory movement of the bow follows the tempo of the music piece. However, to take into account the inter-personal variability, the natural frequency of each player was randomly extracted from a Gaussian distribution ($\sigma$ = 1) centred on the tempo of music.
For each group of players and for each music piece, some of the assumptions of the classical Kuramoto model were not satisfied: although the violin players can be deemed nearly identical, the mean-field cannot be assumed because the coupling strength among the players is not identical. The low number of players does not affect the results as argued by theoretical studies and previous experiments.
A first fitting of the model was done by removing the mean-field assumption and by applying a regression through the origin. However, the goodness-of-fit ($R^2$) was poor that is about 10%. In order to improve the model, a more complex coupling periodic function was chosen: the first term of Fourier series with unitary coefficients. A multiple regression through the origin was applied to the following sets of equations:

$$\dot{\theta_1} = \omega + \frac{1}{N_{g1}}[C_{12}(sin(\theta_2 - \theta_1) + cos(\theta_2 - \theta_1)) + C_{13}(sin(\theta_3 - \theta_1) + cos(\theta_3 - \theta_1)) + C_{14}(sin(\theta_4 - \theta_1) + cos(\theta_4 - \theta_1))]$$

$$\dot{\theta_2} = \omega + \frac{1}{N_{g1}}[C_{21}(sin(\theta_1 - \theta_2) + cos(\theta_1 - \theta_2)) + C_{23}(sin(\theta_3 - \theta_2) + cos(\theta_3 - \theta_2)) + C_{24}(sin(\theta_4 - \theta_2) + cos(\theta_4 - \theta_2))]$$

$$\dot{\theta_3} = \omega + \frac{1}{N_{g1}}[C_{31}(sin(\theta_1 - \theta_3) + cos(\theta_1 - \theta_3)) + C_{32}(sin(\theta_2 - \theta_3) + cos(\theta_2 - \theta_3)) + C_{34}(sin(\theta_4 - \theta_3) + cos(\theta_4 - \theta_3))]$$

$$\dot{\theta_4} = \omega + \frac{1}{N_{g1}}[C_{41}(sin(\theta_1 - \theta_4) + cos(\theta_1 - \theta_4)) + C_{42}(sin(\theta_2 - \theta_4) + cos(\theta_2 - \theta_4)) + C_{43}(sin(\theta_3 - \theta_4) + cos(\theta_3 - \theta_4))]$$

$$\dot{\theta_5} = \omega + \frac{1}{N_{g2}}[C_{56}(sin(\theta_6 - \theta_5) + cos(\theta_7 - \theta_5)) + C_{57}(sin(\theta_7 - \theta_5) + cos(\theta_7 - \theta_5)) + C_{85}(sin(\theta_8 - \theta_5) + cos(\theta_8 - \theta_5))]$$

$$\dot{\theta_6} = \omega + \frac{1}{N_{g2}}[C_{65}(sin(\theta_5 - \theta_6) + cos(\theta_5 - \theta_6)) + C_{67}(sin(\theta_7 - \theta_6) + cos(\theta_7 - \theta_6)) + C_{68}(sin(\theta_8 - \theta_6) + cos(\theta_8 - \theta_6))]$$

$$\dot{\theta_7} = \omega + \frac{1}{N_{g2}}[C_{75}(sin(\theta_5 - \theta_7) + cos(\theta_5 - \theta_7)) + C_{76}(sin(\theta_6 - \theta_7) + cos(\theta_6 - \theta_7)) + C_{78}(sin(\theta_8 - \theta_7) + cos(\theta_8 - \theta_7))]$$

$$\dot{\theta_8} = \omega + \frac{1}{N_{g2}}[C_{85}(sin(\theta_1 - \theta_8) + cos(\theta_5 - \theta_8)) + C_{86}(sin(\theta_6 - \theta_8) + cos(\theta_6 - \theta_8)) + C_{87}(sin(\theta_7 - \theta_8) + cos(\theta_7 - \theta_8))]$$

where $C_{ij}$ takes into account the connectivity arrangement of the complex network and the coupling strength.
The matrix of the regression coefficients was well-conditioned for every music piece, for every segment, for every group. Further, also the ratio between the Euclidean norm of residual and the output of the model is low. Therefore, the model fits the data. Also in this case the goodness-of-fit was computed. The following tables summarise the conditioning numbers, the ratio between the Euclidean norm of residual and the output of the model, and the adjusted $R^2$ for Piece 1 (Take 02 and Take 03) and for Piece 5 (Takes 13, 14, and 15) respectively. The tables on the right refer to the OWN conductor; those on the left to the NEW conductor.

| **Take 02** | $adjusted - R^2gr_1$ | $adjusted - R^2gr_2$ | $k(gr_1)$ | $k(gr_2)$ |
|---|---|---|---|---|
| Segm 1 | 0.46 | 0.63 | 2.56 | 2.75 |
| Segm 2 | 0.59 | 0.44 | 3.30 | 2.58 |
| Segm 3 | 0.78 | 0.74 | 7.30 | 6.42 |
| Segm 4 | 0.70 | 0.65 | 4.15 | 4.32 |
| **Take 03** | $adjusted - R^2gr_1$ | $adjusted - R^2gr_2$ | $k(gr_1)$ | $k(gr_2)$ |
| Segm 2 | 0.44 | 0.47 | 2.70 | 3.90 |
| Segm 3 | 0.78 | 0.70 | 7.41 | 6.60 |
| Segm 4 | 0.68 | 0.67 | 4.50 | 5.97 |

| **Take 02** | $adjusted - R^2gr_1$ | $adjusted - R^2gr_2$ | $k(gr_1)$ | $k(gr_2)$ |
|---|---|---|---|---|
| Segm 1 | 0.40 | 0.58 | 2.54 | 2.60 |
| Segm 2 | 0.60 | 0.53 | 5.50 | 3.60 |
| Segm 3 | 0.77 | 0.88 | 5.45 | 7.70 |
| Segm 4 | 0.59 | 0.53 | 3.82 | 3.52 |
| **Take 03** | $adjusted - R^2gr_1$ | $adjusted - R^2gr_2$ | $k(gr_1)$ | $k(gr_2)$ |
| Segm 2 | 0.64 | 0.44 | 4.28 | 3.42 |
| Segm 3 | 0.74 | 0.83 | 5.14 | 10.7 |
| Segm 4 | 0.62 | 0.57 | 4.10 | 3.80 |

| **Take 13** | $adjusted - R^2gr_1$ | $adjusted - R^2gr_2$ | $k(gr_1)$ | $k(gr_2)$ |
|---|---|---|---|---|
| Segm 2 | 0.48 | 0.44 | 3.27 | 2.66 |
| **Take 14** | $adjusted - R^2gr_1$ | $adjusted - R^2gr_2$ | $k(gr_1)$ | $k(gr_2)$ |
| Segm 1 | 0.58 | 0.54 | 4.80 | 4.78 |
| Segm 2 | 0.43 | 0.58 | 2.97 | 3.43 |
| Take 15 | $adjusted - R^2gr_1$ | $adjusted - R^2gr_2$ | $k(gr_1)$ | $k(gr_2)$ |
| Segm 1 | 0.66 | 0.61 | 6.02 | 5.02 |
| Segm 2 | 0.66 | 0.34 | 6.02 | 1.75 |

| **Take 13** | $adjusted - R^2gr_1$ | $adjusted - R^2gr_2$ | $k(gr_1)$ | $k(gr_2)$ |
|---|---|---|---|---|
| Segm 2 | 0.42 | 0.32 | 2.60 | 2.10 |
| **Take 14** | $adjusted - R^2gr_1$ | $adjusted - R^2gr_2$ | $k(gr_1)$ | $k(gr_2)$ |
| Segm 1 | 0.94 | 0.87 | 11 | 8.71 |
| Segm 2 | 0.42 | 0.52 | 2.98 | 3.41 |
| **Take 15** | $adjusted - R^2gr_1$ | $adjusted - R^2gr_2$ | $k(gr_1)$ | $k(gr_2)$ |
| Segm 1 | 0.69 | 0.61 | 3.89 | 3.58 |
| Segm 2 | 0.41 | 0.47 | 3.27 | 3.52 |

The model provides overall satisfactory results and it is suitable for probing the subtle phenomenon of social interaction among peers and when a hierarchy is evident.

## 3.6 Entrainment and leadership

| 3.6 Orchestra section study: **Entrainment and leadership (conductor and musicians)** |
|---|
| **IIT-UNIGE** |

**Abstract:** In experiment 3.6 we want to scale our previous pilot experiment (3.5) up to the final experimental set-up. In this sense we need to record better kinematics data including right and left arm, upper torso and head tracking. At the same time we'll record also muscle data for at least a limited number of musicians as piloted in 3.1 and 3.3. Data acquisition is scheduled later this year and the design is being discussed among Siempre partners. We'll use a String Orchestra from Music Conservatory of Genoa, including at least 4 first violins, 4 second violins, 2 viole, 2 celli and 1 contrabbasso and 2 conductors. They will play few short musical pieces (2/3 pieces no longer than 2 minutes each). Music pieces will be chosen as characterized by specific musical structure (e.g., isorhythmic and phrasing features) in order to facilitate the verification of hypotheses.

Method was briefly described in D2.1 and since data acquisition is scheduled later this year, the design is still being discussed among Siempre partners.

**Evaluation:** The kind of data we intend to record will be one of the major output from the consortium to the external research community. In fact, it will be the first time that multimodal (audio, kinematics, emg, physiological sensors) synchronized data of this kind has ever been recorded and made public. However, such data acquisition poses a series of very complex challenges. For these reasons we're trying to approach the problem in a step by step manner, and thus solving one problem at a time. The quartet scenario seemed technically more simple than the orchestra scenario and thus we focused on that first (see 3.3). Having learned a lot from the issues emerged with the quartet we're confident that in few months we'll be able to record also the orchestra.

# 4. THERMOGRAPHY EXPERIMENTS IN THE SIEMPRE PROJECT

One of the main pillars of the SIEMPRE project is the recording of the audience emotional entrainment. As shown in deliverable 1.1 there are several well-established methods that can measure such kind of data. In fact, the consortium will effectively use these methods.

However, during the fruitful brainstorming we had during the kick-off meeting a novel idea, proposed by IIT, emerged. We realized that we could record large audiences emotional entrainment via thermographic images. Such approach is potentially very powerful although there are several computational and technical issues with no solution yet.

In fact, 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.

Therefore, we decided to invest part of the first year in exploring the use of thermography in these contexts. The results of such research may prove extremely influential for both the advancement of the SIEMPRE project agenda and also a larger community exploring the use of thermography in affective neurosciences. Here follows a list of on going research projects about these issues.

## 4.1 Pilot of thermographic measures of large audiences

| 4.1 Pilot of thermographic measures of large audiences<br>Synchronization of the audience in terms of thermographic responses to music in an ecological scenario |
|---|
| **IIT** |

**Abstract:** This pilot recording was intended to test whether thermographic recordings of large audiences were feasible. Also, the goal was to find all possible technical problems of these kind of recordings and develop tools for data analyses. In fact, a great part of the problem is to remove artifactual temperature changes.

**Method**:

1. Every second N frames are sampled and their average frame is computed
2. The average frame is segmented into M ROIs. ROIs can be blind squares or ellipses circumscribing the subject faces.
3. Given two average frames the distributed temperature change (DTC) is extracted by computing the normalized difference between the No. of blocks that had an increase of temperature and the No. of blocks that had a decrease in temperature. The DTC indicates the amount of synchronous temperature changes.
4. Compute the integral of DTC.

Method is now obsolete and the method described in 4.4 is now used.

**Technical description of the results:** We developed a series of methods to remove artifacts from the data and extract global temperature changes as well as inter-subjects temperature synchronization.

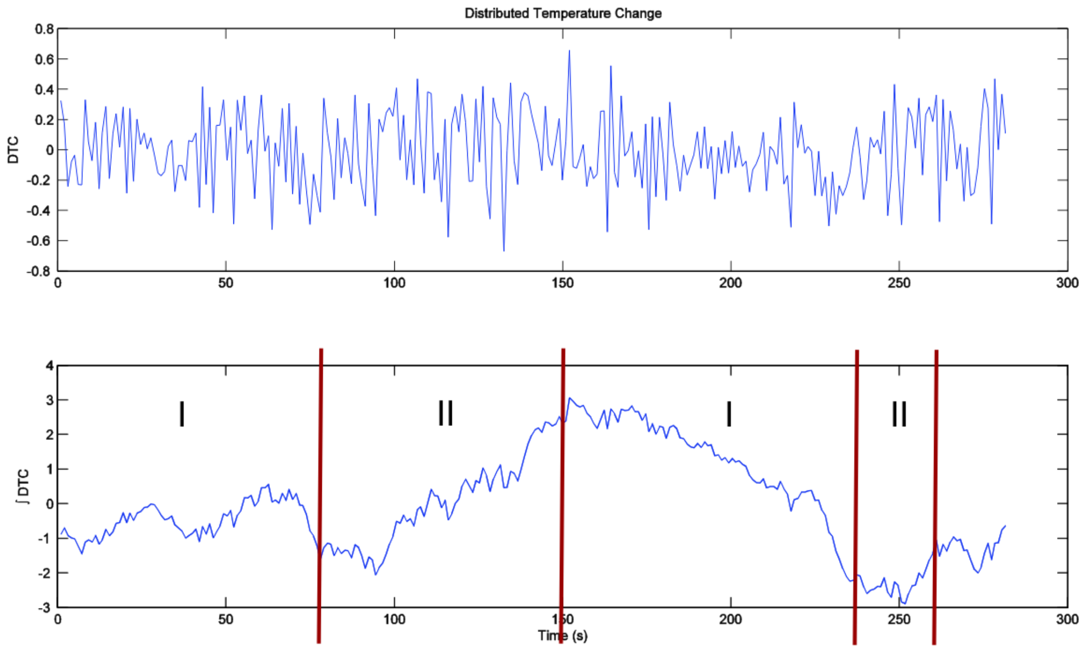


*Figure 4.1a: the upper panel shows the distributed temperature change during the execution of a single piece of music. The lower panel shows the integral of the distributed temperature change. Red lines represent the time points when the first and second musical pieces were started. It is clearly visible that the two themes induced a different global pattern of temperature in the audience.*

**Evaluation**: Results show that although technically challenging, thermographic recording of large audiences is possible. We were able to describe all possible technical problems and developed techniques and computational tools to solve all of them satisfactorily.

## 4.2 Pilot of thermographic measures of small audience

| |
|---|
| **4. 2 Testing of audience reactions using thermographic patterns in faces in relation to different types of musical expressiveness** |
| **UNIGE-CH** |

**Abstract:** As mentioned in the part 4.1 we are using thertmographic recordings in order to investigate the physiological reactions during music listening. We developed a method (Jarlier, Grandjean et al., 2011) to analyze thermographic patterns of small audience taking into account the face of each subject. In the first experiment described below and in the D2.1 we analyzed thermographic patterns of 10 participants during the listening of 2 different kinds of musical expressiveness (academic and emphatic). The statistical analysis of the group revealed significant increase of the facial temperature for the left and for the upper parts of the face.

**Method**: Method was as described in D2.1 and in the published paper Jarlier, Grandjean et al., 2011. For this experiment we defined the region of interest (ROIs) according to the prior hypotheses: i) increase of the facial temperature for the emphatic condition compared to the academic condition in the left side (according to the results in EMG and emotional facial expression showing that the left part of the face is more expressive compared to the right side, ii) increase of the temperature for the upper part of the face for the emphatic condition compared to the academic condition (accordingly with the results in EMG and facial expressions showing that the lower part of the face is more under voluntary control (especially for speech production) while the upper part is less under voluntary control and then more sensitive to be modulated by emotions.

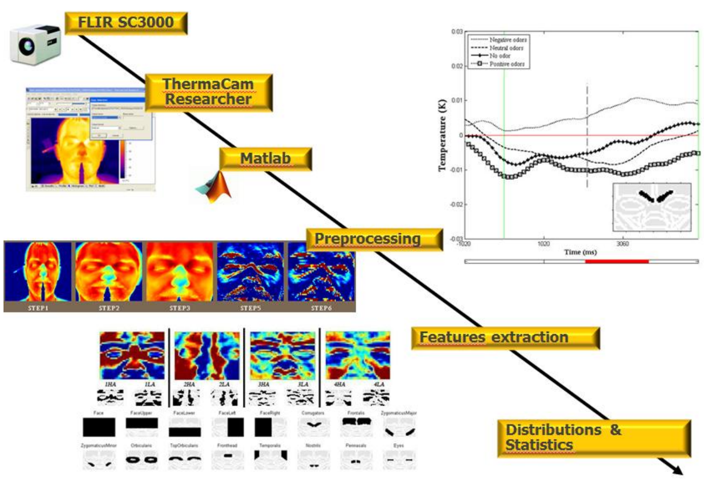



*Figure 1: left panel: extraction of the face for a small audience (only the visible faces were extracted). Right panel: methodology of the experiment including i) data acquisition, ii) pre-processing and normalization, iii) features extraction and iv) statistical analysis.*

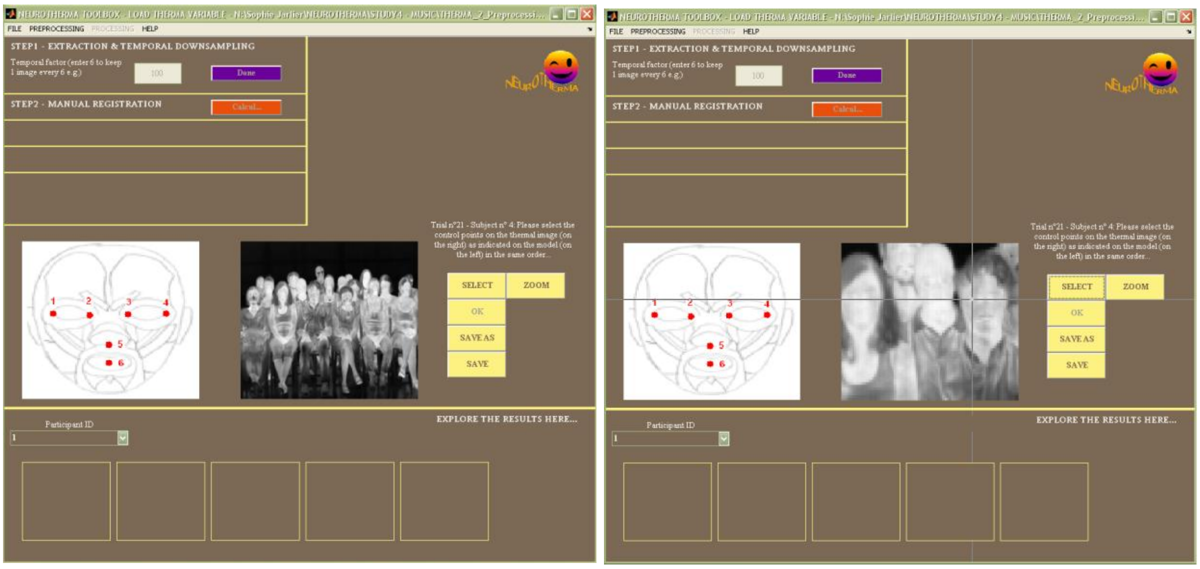



*Figure 2: Screen captures of the toolbox NeuroTherma during the pre-processing and the normalization of each face (red points: landmarks for the normalization).*

**Results:**

The analysis revealed a significant increase of temperature of the left part of the face for the emphatic condition compared to the academic condition from ~50 seconds after the onset of the musical performance to 200 seconds (see Figure 3 and 4 below).

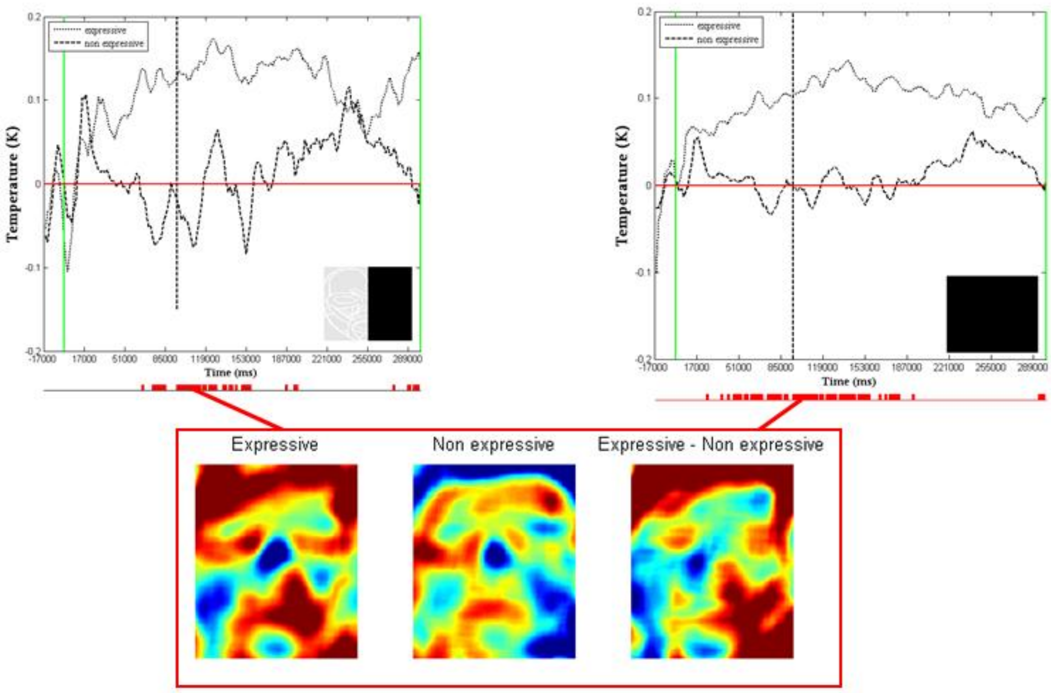



*Figure 3: Left upper panel: significant increase of temperature for the emphatic condition compared to the academic condition for the left part of the faces (the red dots under the graphic illustrate the significance of the t-tests performed in time). Right upper panel: significant increase of temperature for the emphatic condition compared to the academic condition for the entire faces. Lower panel:*

*average of the pattern of the temperature for the two experimental conditions and for the comparison emphatic condition (expressive) minus academic (non-expressive) condition.*

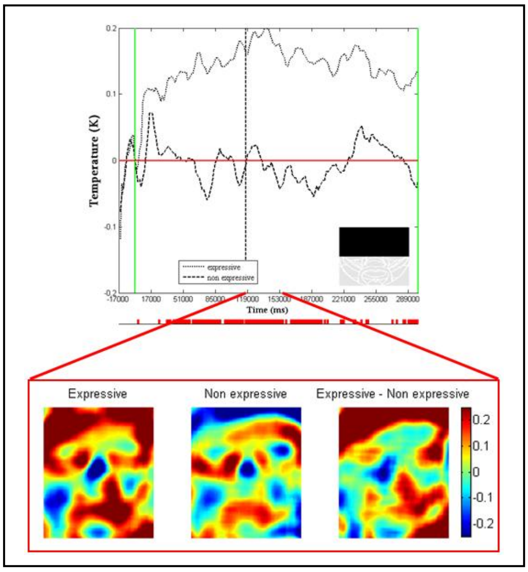



*Figure 4: Graphic of the average of the temperature showing a significant increase of temperature for the emphatic condition compared to the academic condition for the upper part of the faces. Lower panel: average of the pattern of the temperature for the two experimental conditions and for the comparison emphatic condition (expressive) minus academic (non-expressive) condition.*

**Evaluation:**

The final analyses are in progress, a paper will be submitted during the next months.

## 4.3 Thermographic recordings with two different cameras

| 4.2 Pilot of thermographic recordings with two different cameras |
| --- |
| IIT, UNIGE-CH |

**Abstract:** This pilot recording was intended to test whether the two thermocameras in the consortium differ in their ability to measure emotion-induced temperature changes. This test was motivated by the fact that the two cameras actually measure different infrared bands and no one has ever tested which IR band is more effective in these kind of research.

**Method**: Method was as described in D2.1 with no changes.

**Technical description of the results:** A fair comparison of two thermographic signals requires that the two signals refer to the same scene.Unfortunately the two cameras have quite different fields of view (see Figure 4.3a) so some kind of normalization needs to be applied. The normalization we applied consisted in the extraction of the same Regions of Interests (ROIs, which correspond to ellipses circumscribing the subject faces in this case), so that the thermographic signals that were compared referred to the same (unnormalized) parts of scene.
Figure 4.3b shows the average temperature of the ROIs per each thermographic camera over time. Apart from a difference in the temporal mean the two temperatures are almost identical, almost all fast and slow changes occurs in both temperatures at the same interval times. Small differences between the two temperatures can be due to fact the two cameras have a different special resolution and that the ROIs were not normalized. Given this striking similarity of the two signals was considered unnecessary
Comparison between the two cameras was also carried out using measures of temperature changes that are much more robust to artifacts do to movement (see section 4.4). As before, the behavior of the two cameras resulted almost identical.

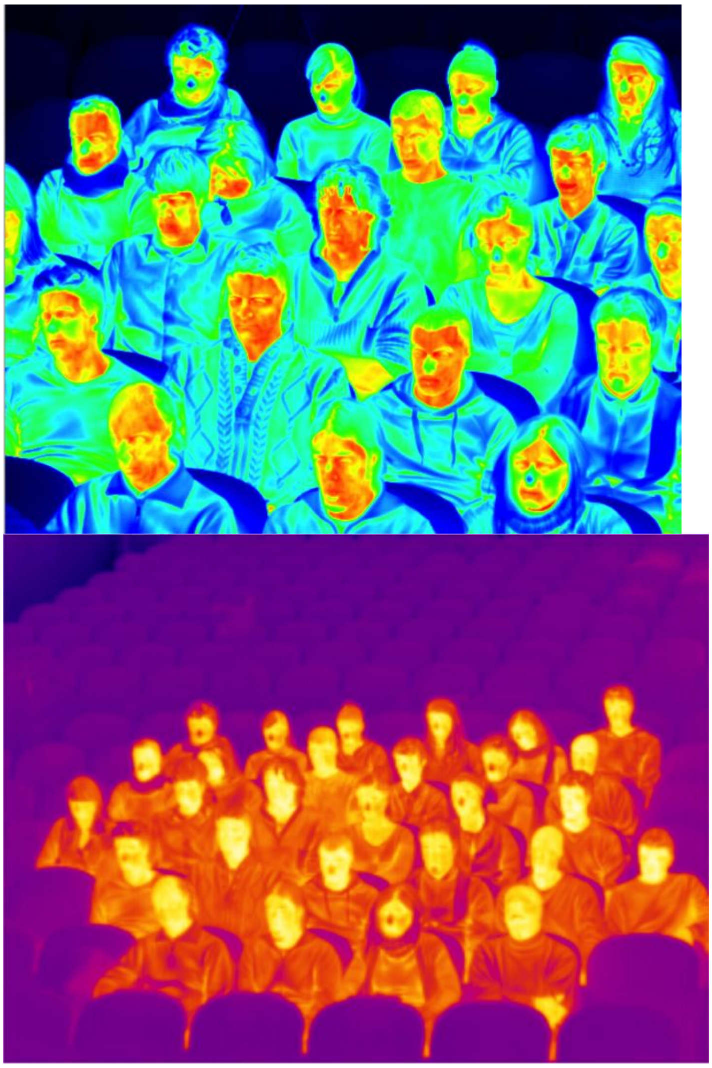
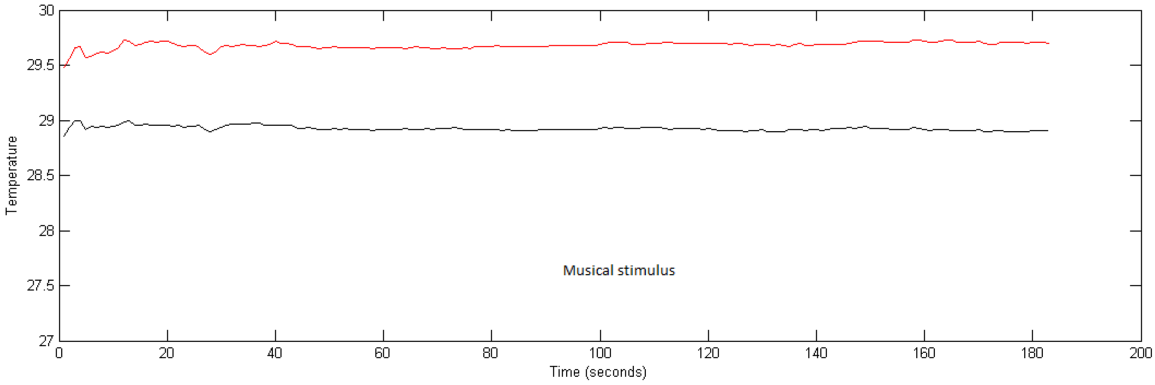


*Figure 4.3a. Two snapshots taken from the two cameras. Note the different resolution and field of view.*



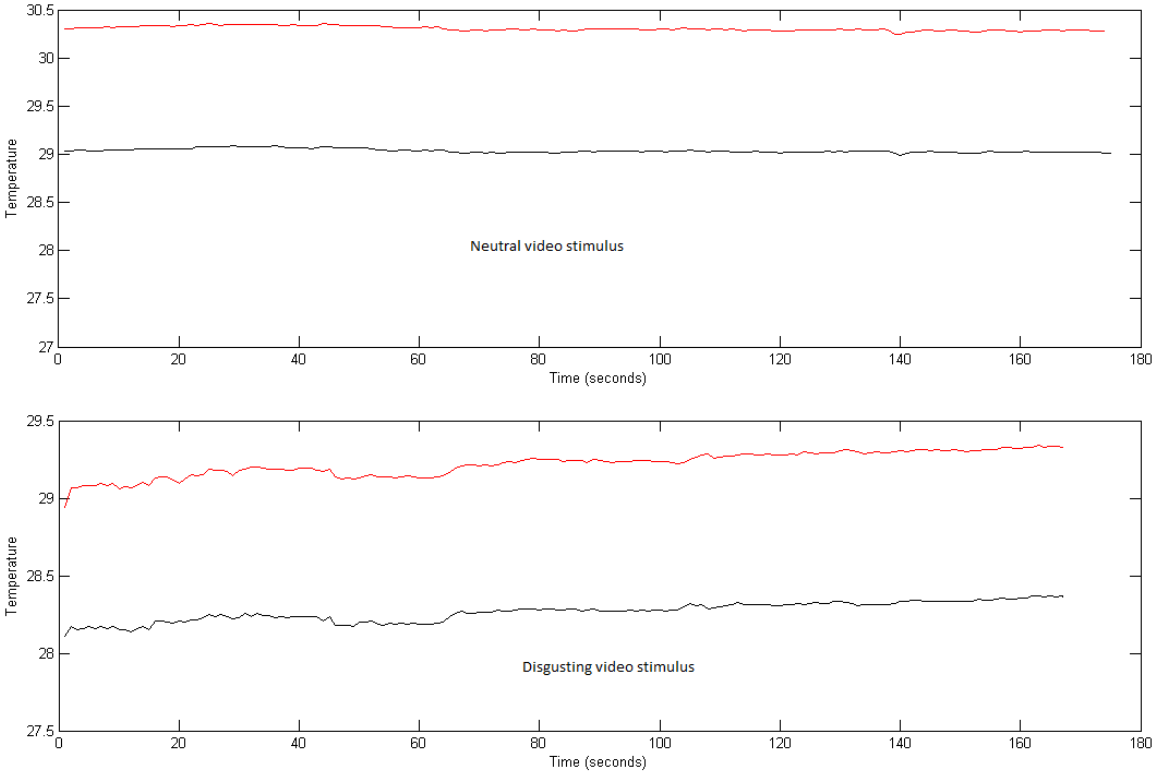



*Figure4.3b. Initial results comparing the performance of the two cameras.*

**Evaluation**: Results show that both thermographic cameras are equally appropriate and comparable for the experiments IIT and UNIGE-CH have being carrying out. This is an important outcome as it allows the two partners to run independent set-ups and compare their respective results using the same "measurement of unit".

## 4.4 Thermographic recordings of large audiences

| |
|---|
| **4.4** Thermographic measures of large audiences<br>Synchronization of the audience in terms of thermographic responses to music in an ecological scenario |
| **IIT** |

**Abstract:** One of the three main scenarios in the Siempre project is the Audience. The goal is to measure the emotional responses triggered by listening to music. However, measuring from a large number of subjects the typical emotional related physiological signals, like galvanic responses or hear beat rate, is quite challenging and intrusive. In parallel, we are testing the feasibility of a quite different approach. In fact, we want to verify whether facial temperature changes are associated to emotional contagion and entrainment while listening to music. In order to do so we are using distal thermographic recordings. We performed two main experiments one at the Carlo Felice theater and another at Casa Paganini. The first one involved around 30 participants listening to a piano concert. The second one instead involved 15 subjects listening to selected musical pieces and watching standardized movies able to elicit strong positive or negative emotions. In the second experiment we also measured state and trait personality data to be correlated to the temperature data.

**Method**: Method was as described in D2.1, with the following adjustments.
A novel measure of temperature change (henceforth, Synchronized Temperature Change, STC) was devised in order to: a) any possible artifact in temperature change due to movements (e.g., a subject moving her head, a subject covering her face with one of her hands, etc…); b) capture synchronized temperature changes in the audience.
First the ROIs (ellipses circumscribing the subject faces) were manually identified using a randomly chosen frame. Subsequently the thermographic signal was subsampled and the temperature change over two consecutive frames was computed per each pixel. Pixel exhibiting temperature and/or temperature changes not compatible with human physiology were excluded. Then the mean temperature change of each ROI was computed. The final measure of temperature change was computed by multiplying the a) integral of the ROIs mean temperature changes over an observation windows and the b) the ROIs inter-correlation in temperature change (i.e., the mean of the pairwise Pearson correlations between ROIs temperature changes ) computed using the same observation window.

**Technical description of the results:**
From an inspection of the recorded thermographic videos and the corresponding STCs the STC resulted to be a measure very robust to artifacts due to subject movements.
An example of STC is given in figure 4.4a. From an informal analysis of STC for all stimuli (plus their respective baselines) it turned out that significantly large positive peaks (often followed by negative peaks) occurred during the most intense scenes (i.e., scenes with highest arousal) when videos inducing negative emotions (disgust and sadness) where presented to the subject.
**Evaluation:**
The correlation between large positive peaks and most intense scenes needs still to be validated using large scale subjective classification of the intensity of the scenes in the video.

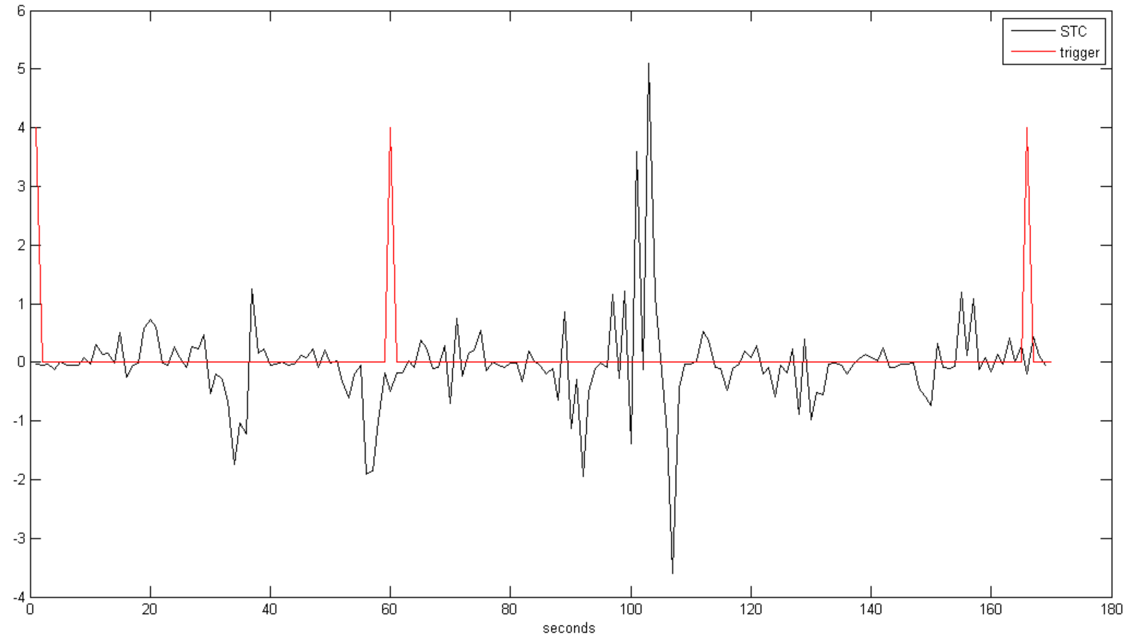



*Figure4.4a. An example of (normalized) synchronized temperature changes. The first trigger peak indicates the baseline stimulus onset, the second peak the stimulus onset and the last peak the stimulus offset.*

**Relationship between Personality traits, emotional regulation and affectivity in the music listening**

UNIGE investigated whether personality traits may affect the way emotions are regulated, felt and perceived when listening to music.

**Method**

A mixed ANCOVA was performed on the participants' ratings of three GEMS second order factor scores (*sublimity*, *vitality* and *uneasyness*), with type of music emotion and expressivity of the performance as within-subject factors and Big Five personality traits scores and positive and negative affectiveness scores as covariate.

**Evaluation**

A significant effect of track, expressivity and their interaction was found in a general linear model (GLM) in which personality traits were not taken into account as covariates. Covariates were subsequently included in the GLM, and a backward selection was used to exclude, one at the time, those covariated whose effect was not significant. In the final model *Extraversion*, *Conscentiousness* and *Openess* had a significant multivariate effect which could account for variation in scores over the effect of factors, which were no longer significant.
Even if replication studies are needed, these results suggest that personality traits may explain the GEMS scores regardless of the type of track or of the method of execution.

# 5. NEUROPHYSIOLOGICAL STUDIES ON RHYTHM ENTRAINMENT

## 5.1 Human Intracranial Local Field potential recordings during percussion listening paradigm (Intracranial I)

| 5.1 Intracranial EEG recording of brain activity during a percussion listening paradigm (Intracranial I): how different metrics and different tempi entrain brain areas during passive listening |
|---|
| UNIGE-CH |

**Abstract**

The goal of this study was to investigate how different brain areas are entrained by percussion beats that vary in terms of tempo (fast/slow) and metrical structure (simple/complex) in an epileptic patient with intracranial electrodes.

**Method**

Here, a young epileptic patient with intracranial electrodes in the supplementary motor area (SMA), amygdala, orbitofrontal cortex, anterior cingulate cortex, hippocampus and temporal cortex presented herself at the CMU for pre-operatory monitoring of seizures.

Eighteen beat tracks were composed for this study consisting in:

- 7 meters: 2/4, 3/4, 4/4, 5/4, 5/8, 7/8 and 9/8
- + 1 metronome condition, that is 1/1 time
- + 1 scrambled condition, no metric

All beats were played at two speeds:

- 90bpm (1.5 Hz) and
- 124bpm (~2.067Hz)

The design can be either thought of as: 7 levels of metric * 2 speed levels

| Metric | 1/1 | | 2/4 | | 3/4 | | 4/4 | | 5/4 | | 5/8 | | 7/8 | | 9/8 | | None | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Speed | 90 | 124 | 90 | 124 | 90 | 124 | 90 | 124 | 90 | 124 | 90 | 124 | 90 | 124 | 90 | 124 | 90 | 124 |

Or: 2 levels of difficulty (4 metric in each) * 2 speed levels

| Difficulty | Simple | Simple | Simple | Simple | Simple | Simple | Simple | Simple |
|---|---|---|---|---|---|---|---|---|
| Metric | 1/1 | 1/1 | 2/4 | 2/4 | 3/4 | 3/4 | 4/4 | 4/4 |
| Speed | Slow | Fast | Slow | Fast | Slow | Fast | Slow | Fast |

| Difficulty | Complex | Complex | Complex | Complex | Complex | Complex | Complex | Complex | Random | Random |
|---|---|---|---|---|---|---|---|---|---|---|
| Metric | 5/4 | 5/4 | 5/8 | 5/8 | 7/8 | 7/8 | 9/8 | 9/8 | None | None |
| Speed | Slow | Fast | Slow | Fast | Slow | Fast | Slow | Fast | Slow | Fast |

The scramble condition induces no sense of metric because it consists in a random juxtaposition of bars of the previous conditions. The patient listened passively to each track only once since it consisted in 25 identical bars of the same meter. After each track she tapped the “rhythm” of the track using the space bar of a laptop computer (tempo tapping was programmed with E-Prime 2, Psychology Software Tools Inc., Pittsburgh, PA). At the end of the experiment she answered the same 12-item explicit Entrainment Questionnaire that we have been using for the Capuçon II pilot study and the Geneva Emotional Music Scale (Zentner, Grandjean & Scherer, 2008).

**Results**

The analyses of the HR and the local field potentials recordings are in progress.

# 6. SCIENTIFIC PUBLICATIONS ATTACHMENT

In this appendix, scientific publications on some of the SIEMPRE experiment listed above and related to D 4.1 are included.

D.Carioti D. Glowinski and C.Chiorri, Personality traits, emotional regulation and affectivity in the music listening, Intl ISSSM2011, University of Jyvaskyla (Finland), 8-18 August 2011

Glowinski, D., Mancini, M., Rukavishnikova, N., Khomenko, V. and Antonio Camurri " Analysis of Dominance in Small Music Ensemble" in Proceedings of the AFFINE satellite workshop of the ACM ICMI 2011 Conference, Alicante Spain, 17-11 2011.

Panagiotis Papiotis, Esteban Maestre, Marco Marchini, Alfonso Peréz – Synchronization of intonation adjustments in violin duets: towards an objective evaluation of musical interaction. In Proc. of the 14th Int. Conference on Digital Audio Effects (DAFx-11), Paris France, September 2011.

Panagiotis Papiotis, Marco Marchini, Esteban Maestre and Alfonso Peréz – Measuring ensemble synchrony through violin performance parameters: a preliminary progress report. In Proc. of the 4th International ICST Conference on Intelligent Technologies for Interactive Entertainment (INTETAIN 2011), Genova Italy, May 2011.

R. Rovegno, M. Gagliardi, D. Glowinski, S. Ghisio, P. Coletta, and A. Camurri, Using the Qualisys MoCap System for Multimodal analysis of string quartet performance: a feasibility study, Intl ISSSM2011, University of Jyvaskyla (Finland), 8-18 August 2011.

# 7. OTHER ATTACHMENTS

## 7.1 Appendix 1: Quality of Experience Questionnaire

**Instructions**

When providing your ratings, please describe how the music you listen to makes you *feel* (e.g. this music makes me *feel* sad). Do not describe the music (e.g., this music is sad) or what the music may be expressive of (e.g. this music expresses sadness). Keep in mind that sometimes a piece of music can be sad or can sound sad without making you feel sad. Please rate the intensity with which you *felt* each of the following feelings on a scale ranging from 0 (*not at all*) to 4 (*very much*). Note that there may be specific labels for each value for some questions (these will be clearly marked). If you should find that an important label is missing from the list to describe what you feel, please add it at the end and rate it.

| **0** | **1** | **2** | **3** | **4** |
|---|---|---|---|---|
| Not at all | Somewhat | Moderately | Quite a lot | Very Much |

1a. Were you aware of any increases in physiological arousal of your nervous system (faster heartbeat, pulse and breathing rate, more perspiration) _____

1b. Were you aware of any decreases in physiological arousal of the your nervous system (slower heartbeat, pulse and breathing rate) ______

2a. Were any emotions or feelings you experience of a pleasant/positive nature? ____

2b. Were any emotions or feelings you experienced of a unpleasant/negative nature? ___

3. Did the music and performance make you feel happy? _____

4. Did the music and performance make you feel angry? _____

5. Did the music and performance make you feel sad? _____

6. Did the music make and performance you feel sublime (inspiring deep veneration or awe ; feelings of wonder, nostalgia, peacefulness or tenderness) _____

7. Did the music and performance make you feel increased levels of vitality? (physical or mental vigour/energy, power, activation, joy, wish to move body/dance) _____

8. Did the music make you feel uneasy (agitated, sad, irritated) _____

9. Did the music and performance make you feel tense? _____

10. Did the music and performance make you feel powerful? _____

11. Did the music and performance make you feel dominant? _____

12. Did the music and performance make you feel physically strong? _____

15. Did you experience any feelings of transcendence (a feeling of existing beyond or outside of the physical world)? _____

13. Did you feel any type of connection with other audience members / listeners? _____

14. Did you feel any type of connection with the music? _____

16. Did you feel any type of connection with performer? _____

17. Did you feel an empathy with the performer? _____

18. Did the performer show emotional engagement with the music? _____

19. Did the performer show a clear intention to communicate with the audience? _____

20. How much did the performance match your expectation of how it would be performed? _____

21. How much did the music itself match your expectation of how it would sound? _____

22. How much did you enjoy the aspects of the performance that were familiar to you/ that you expected?

_____

23. How much did you enjoy the aspects of the performance that were unfamiliar to you/ that you were not expecting? _____

23. How engaged did you find yourself with the music and performance? _____

24. How much was your concentration held on the music and the performance? _____

25. How aware were you of your own hunger, thirst and need to urinate during the music and performance?

_____

26. How aware were you of your heartbeat and other bodily sensations? _____

27. How much did you find yourself paying attention to the room and space around the performance area? (not the performance area itself) _____

28. How much did you find yourself paying attention to other audience members/listeners? _____

29. How aware were you of other audience members/listeners? _____

30. How much did you find yourself paying attention to the length of the performance? _____

31. How aware of time did you feel? _____

32. Did you feel the performance was shorter or longer than it actually was? _____

33. Did you feel self-conscious at any point during the performance? _____

34. Did you find yourself paying attention to your own thoughts during the performance? _____

35. Did you feel that the performer responded to the mood of the audience during the performance? (either through audience interaction or musically) _____

36a. After the performance do you feel any different than you did before in the categories below?

World is more beautiful _____ Life is better _____

Problems seem less than before _____ Less stressed _____

36b. After the performance did you feel as though you had more energy than before? _____

37. _____ Do you consciously experience any of the following?

Changes in heart rate _____ Changes in posture _____

Changes in breathing rate _____ Increased movement _____

38. Did you ever feel as though you were “floating” in the music? _____

39. Did you experience a loss of awareness of the physical world? _____

40. Did you feel as though you were part of/within the music? ____

41. Did you ever feel a sense of movement with the music? ____

42. Did you feel part of the performance? _____

43. Do you find it hard to become absorbed in movies/plays/books? _____

44. Do you find it hard to pay constant attention to movies/plays/books? _____

45. Did you experience any chills, thrills, tingling or goosebumps? _____

46. Can you recall specific incidents from the performance easily? _____

47, Did you find yourself physically marking the rhythm (foot tapping, head nodding, clapping, swaying etc.)

_____

48. Did you notice other audience members physically marking the rhythm? _____

49. Did you find yourself reproducing the music in any way? (mouthing or singing the harmony/lyrics, also closing eyes) ______

50. Did you notice others reproducing the music in any way? _____

(Note for comparison that a set of items related to the last 5 were used in 1.4a to measure ‘Explicit Entrainment’, i.e. self-reported subjective feelings of being entrained by the beat of the music).

## 7.2 Appendix 2 post-performance questionnaire

Submitted to each musician to investigate the quality of their performance and the level of interactions between them.

| | |
|---|---|
| **1. How did you feel satisfied by your performance?** | Not at all ① ② ③ ④ ⑤ ⑥ ⑦ Completely |
| **2. How much did you feel followed / heard by others ?** | |
| **From first violin** | Not at all ① ② ③ ④ ⑤ ⑥ ⑦ Completely |
| **From second violin** | Not at all ① ② ③ ④ ⑤ ⑥ ⑦ Completely |
| **From viola** | Not at all ① ② ③ ④ ⑤ ⑥ ⑦ Completely |
| **From cello** | Not at all ① ② ③ ④ ⑤ ⑥ ⑦ Completely |
| **3.How much did you follow / hear the others?** | |
| **From first violin** | Not at all ① ② ③ ④ ⑤ ⑥ ⑦ Completely |
| **From second violin** | Not at all ① ② ③ ④ ⑤ ⑥ ⑦ Completely |
| **From viola** | Not at all ① ② ③ ④ ⑤ ⑥ ⑦ Completely |
| **From cello** | Not at all ① ② ③ ④ ⑤ ⑥ ⑦ Completely |
| **4. How much did the expressive content of the music piece follow?** | Not at all ① ② ③ ④ ⑤ ⑥ ⑦ Completely |
| **5. How much did you feel in sync with the other musicians?** | Not at all ① ② ③ ④ ⑤ ⑥ ⑦ Completely |