KEER2010, PARIS | MARCH 2-4 2010
INTERNATIONAL CONFERENCE ON KANSEI ENGINEERING AND EMOTION RESEARCH 2010

# COMPUTATIONAL MODEL OF ENTRAINMENT WITHIN SMALL GROUPS OF PEOPLE:

## TOWARD NOVEL APPROACHES TO KANSEI INFORMATION PROCESSING

**Antonio Camurri*[a], Giovanna Varni[a] and Gualtiero Volpe[b]**

[a] *InfoMus Lab, DIST, Università di Genova, Italy*

## ABSTRACT

The study of human intended and unintended interpersonal co-ordination is one of the more interesting and challenging topics in the psychological and behavioral sciences, and is receiving an increasing interest from the ICT research communities. The objective here is to develop more natural and intelligent interfaces, with a focus on non-verbal communication, embodiment, and enaction [1], and, from another perspective, on KANSEI Information Processing [2], [3]. In natural sciences the co-ordination phenomenon is better-known as entrainment. In this paper we present research scenario to study and measure entrainment in small groups of people from the KANSEI perspective. In our approach, participants in the experiment are modeled as components of a complex system and entrainment is measured starting from the computation of an index of phase synchronization. The computational model for real-time extraction of such phase synchronization index from a group of people is also discussed. We adopted a string quartet musical performance as test-bed. We designed an experiment in which we recorded synchronized multimodal features (movement, audio, physiological signals) aiming at measuring entrainment in different conditions. Preliminary results from this research include a software library based on the EyesWeb XMI platform for real-time extraction of non-verbal social signals.

**Keywords:** *entrainment, synchronization, emotion, joint music activity, multimodal interaction.*

---

* **Corresponding author**: Antonio Camurri, InfoMus Lab, DIST, Università di Genova, Italy, Viale Causa, 13, I-16145-Genova (Italy), antonio.camurri@unige.it

## 1. INTRODUCTION

This paper presents early work from our research project aiming at developing novel theoretical and methodological frameworks, computational models, and algorithms for the analysis of creative communication within groups of people. Social interaction and its related competencies is one of the fundamental components of human life, widely studied in psychology, and of growing importance for ICT (e.g., social media, on-line and mobile communities and web 2.0 technologies).

“Social intelligence” or social competencies, understood as the ability to deal effectively in interpersonal contexts, is at the basis of any social group, small or large. This is true for humans as well as for animals, particularly for those characterized by rich social interactions. This ability was widely studied in psychology and more recently in neurophysiology, and in these last years is receiving a growing interest from the ICT communities. The attitude of a subject with respect to others can be studied by analyzing channels of non verbal communication with particular reference to behavioral measures (e.g., movement, gesture, posture), auditory production, physiological measures (e.g., electromyography, cardiac, skin conductance), and their expressive, emotional content. Complex aggregates of such verbal and non-verbal signals, expressing the attitude of a human being toward human or virtual communities are known as social signals. These include e.g., synchronization, empathy, attention, and dominance. Research in experimental psychology and neurosciences has shown that non verbal communication is one of the foundational aspects of social interaction. It is based on our capability of implicitly understanding the internal states of others, from simple motor behaviors to complex emotional reactions.

As referred by Vinciarelli [4], in the state of the art three main research directions are addressed to face social interaction: (i) the analysis of interactions in small groups, (ii) the recognition of functional roles, and (iii) the sensing of users attitudes towards computers interfaces. Pioneering studies toward the development of models and algorithms to analyze human social interaction have been performed, for example, by Alex Pentland and his group at MIT Media Lab. He developed three socially aware platforms to measure several non-linguistic vocalic social signals in salary negotiation and friendship scenarios [5], [6]. Research on automated detection of behavioral roles and on recognition of collective actions in small groups focused attention on detection of the most dominant person or of the *dominant clique* (i.e., a set of dominant people) in group-meeting scenario only [7] .

To date, in most cases, researchers face analysis of these signals relying on intensive observation sessions of videos, manual annotations, and questionnaire responses only. However, although trained psychologists can predict social behavioral outcomes from ‘thin slices’ of behavior with high accuracy [8], this approach results very time costly and unreliable when a continuous monitoring of large groups of people is required. Therefore, research is needed on the quantitative and qualitative measures of interpersonal communication and on the supporting models and tools.

One of the major elements of novelty of our research project is to focus on the study of creative social interaction. More specifically, our research is focused on ensemble musical performance, an ideal test-bed for the development of models and techniques for measuring creative social interaction in an ecologically valid framework. In this paper we present early results on the modeling and real-time analysis of entrainment occurring during implicit synchronization in the emotionally intense experiences of joint music performance in a string quartet. Interacting with music is one of the fundamental examples of non-verbal human activities that is above all interactive, creative and social. Music is widely regarded as the medium of emotional expression par excellence. In addition, ensemble performance is one of the most closely synchronized activities that human beings engage in (actions coordinated to within fractions of a second are considered routine even in amateur performance). Indeed, it is believed that this ability from individuals and groups to entrain to music is unique only to humans. Moreover, unlike speech, musical performance is one of the few expressive activities allowing simultaneous participation. As such, the potential of music to enable the communication and entrainment of emotion is unparalleled.

Real-time extracted multimodal features on movement, audio, and physiological data are used to extract expressive qualities, e.g., starting from research on multimodal expressive gesture [9] and quantitative emotional estimation [10]: we refer to such multi-level features as *expressive Movement, Audio, Physiological* (eMAP) features. eMAP features are extracted from participants (musicians) using real-time, synchronized, multi-modal feature extraction techniques briefly explained in this paper, and are the inputs to the computational models explaining the processes underlying interpersonal creative communication.

## 2. RESEARCH SCENARIO

String quartet is one of the most promising test-bed to investigate social interaction in a small group. We have strong consolidated previous experience in working with musicians and dancers. In particular we recently undertook feasibility and preliminary experiments on the measurement of string quartets. We investigated, in the last couple of years, in specific events at our research site, some internationally renowned concert quartets including the Cuarteto Casal, the Quartetto di Cremona, the Quartetto Prometeo.

Figures 1, 2, 3, and 4 show some details of the last experiment to measure entrainment. The figures also specify the eMAP signals chosen for this experiment.

**Figure 1**: the setup of the string quartet experiments at our research centre. Red circles show the two high-speed and high-resolution video-cameras used for movement analysis (frontal

view, and top view). Audio recording has been done with two environmental microphones, and audio analysis has been based on one contact microphone on each instrument, to obtain separate audio channels for each instrument**.**

**Figure 2:** set up of the string quartet experiments at our research centre: the picture shows other video-cameras for movement analysis, lighting systems for motion tracking, and the setup of physiological on-body sensors to musicians (in collaboration with Ben Knapp (SARC, University of Belfast).

**Figure 3:** The famous concert string quartet Quartetto di Cremona playing during the experiment: each musician wears a white hat including a green passive marker and a 3 axis

accelerometer. Further, a 3 axis accelerometer on the back, physiological sensors monitoring heart rate, breath depth, ocular movements, and face muscles are used.

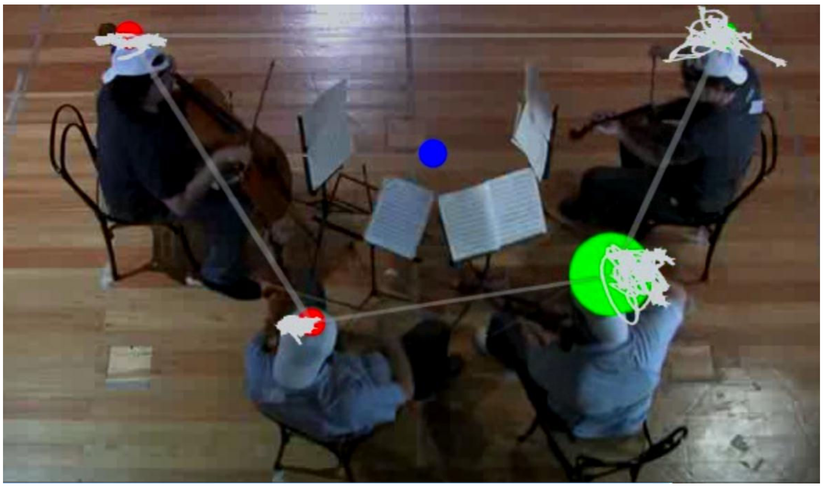

**Figure 4:** eMAP: movement cues - Visualization of some of the main individual and group movement cues used to measure synchronization and leadership in the quartet. Individual movement cues include those based on position and velocity of the head barycenter of each musician (e.g. [9], [11],[12]). Group cues include the dynamics of the polygon individuated by the heads of the musicians, and of its barycentre.

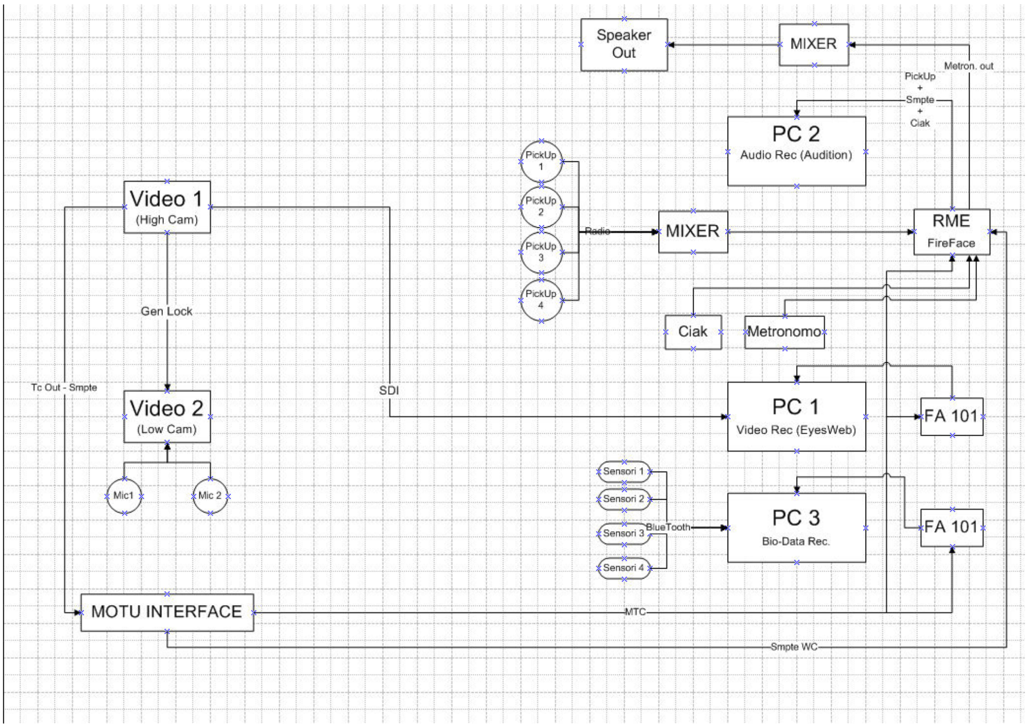

**Figure 5:** a sketch of the technical set-up used for the real-time eMAP multimodal recordings of the Quartetto di Cremona experiment.

A sketch summarizing the technical set-up used for the real-time eMAP multimodal recordings of the Quartetto di Cremona experiment is shown in Figure 5. PC1 is devoted to store the video recordings from the high-speed and high-resolution Video 1 Camera (top view camera in Figure1). Video 1 generates a Time Code signal which is sent to all other acquisition devices to synchronize the multimodal data. PC2 stores the audio signals from the contact microphones on the musical instruments. The synchronization with the audio signals is via SMPTE Time Code. Finally, PC3 stores the physiological data sent via BlueTooth from the sensors. Eight sensors and one ad hoc BlueTooth channel are used for each musician. Data from sensors are sent to PC3. The Video 2 Camera placed from the audience (frontal view camera in Figure 1) perspective is used to record video signals to be used for psychology experiments based on video and environmental audio. This video-camera is synchronized by a genlock signal and acquires also stereo audio signals from a pair of omnidirectional microphones (to be used for future experiments with subjects).

### 2.1. The Experiments

The quartet experiments have been carry out in a real-world, ecological environments: the concert hall which is part of our research centre.

The choice of the musical excerpts performed by musicians has been done carefully aiming at the best conditions for measuring entrainment: for example, in some cases the choice requires a prevalence of omo-rhythmic sections, i.e., the four musicians play sections sharing the same rhythmic patterns. In other cases, the musical excerpts are characterized by polyphonic phrasing with dialogic nature (e.g., a theme which runs from one musician to the other, answers etc.). These two choices can better lead to the measure of synchronization and leadership among musicians, respectively. Each music fragment lasts about two minutes: this is suitable for having sufficient time for both establishing synchronization (when occurs) and to measure other social signals among musicians.

Several different performance conditions have been individuated, e.g., to have the premises for obtaining different levels in synchronization among musicians. For example:

- **Condition 1 - Neutral**: a rehearsal-like performance, aimed at studying performance practice, more technique-oriented;
- **Condition 2 - Expressive**: how the quartet normally plays in a concert (with/without an audience);
- **Condition 3 - Emotion**: induction of a positive emotion before the performance on one or more musicians; or request to express/simulate a given emotion;
- **Condition 4 - Social**: e.g., soon before the recordings, a swap of the musical score of the first and second violin, unexpected by the musicians, is proposed: this is a strong perturbation, which already demonstrated to cause low entrainment among the musicians, less careful to social interaction, and more careful to their specific task (to read an unexpected musical score and to play a different role in the group).

After each performance a questionnaire about its quality is filled up separately by each musician.

### 2.2. Open research questions

The following research questions are of interest for our research:

- Do the expressive, emotional and social contexts (Conditions 2, 3, 4) affect the interpersonal entrainment and emotional contagion observed across the various modalities or in a combination of the modalities?

- In the case of changes in the context, is the group of musicians able to find strategies in order to re-establish the feeling of synchronization and group cohesion? How does this affect emotional contagion as indicated by the autonomic physiological signals?

- Does entrainment observed in one modality influence entrainment in the other modalities? E.g., does breath cycle entrain behavioral movement cues ?

- In the case the leading music score, usually played by the first violin, is switched to the second violin, is it still possible to determine the leadership in the group?

- Can the motoric component alone - see Scherer's "Component Process Model" [13] explain the entrainment or does autonomic activity (as effected by emotion) play a role?

These are ambitious research questions for the future, and in the following sections we only present a small contribute toward the modeling and real time analysis of entrainment in small group of people.

## 3. MEASURING ENTRAINMENT AS PHASE SYNCHRONIZATION

We start from the hypothesis that entrainment, in certain conditions, can be measured by the phase synchronization of peculiar eMAP features, and in particular in our experiment, in non-functional movement features: head and trunk movement. We propose to use methods originally conceived for complex systems, like neural populations, mechanical oscillators, and electronic circuits to face social interaction analysis. Our approach considers each single participant involved in the interaction as a component of a complex system, and describes his/her dynamical behavior by means of the time evolution of a *n*-dimensional state vector of features in her phase space. The state vector components of behavioral features may include, for example, coordinates and velocity of joints or of the other relevant points (e.g., center of mass of head or trunk), amount of movement, fluentness, impulsivity of head movement, as well as other multimodal features extracted from audio and biometric data. We take into account the temporal dynamics of each behavioral feature. This is crucial to a better understanding of the mechanisms underlying group cohesion and synchronization processes: in most existing approaches only global, static measures are considered.

The single interacting components (the human beings in our case) of a complex system exhibit global properties that are not obvious from their individual dynamics. Phase Synchronisation (PS) is one of these emerging properties and it can be generally defined as an "adjustment of rhythms of oscillating objects due to their weak interaction" [14]. PS is chosen as one of the baseline low-level social signals to indirectly measure more complex phenomena like empathy and dominance in small groups of subjects. This work addresses PS exploiting the concepts of Recurrence [15], Recurrence Plots (RP) [16] and Cross-Recurrence Plot (CRP), and their quantification by means of Recurrence Quantification Analysis (RQA) [17], [18]. RP/CRP and RQA give qualitative and quantitative information on systems' dynamics and their interrelations in terms of trajectories in the chosen phase space, respectively. RP is a time-time colorimetric plot of a binary matrix (the recurrence matrix) displaying all times instants at which the components of state vector take again almost (that is with a tolerance) the same numerical values. CRP extends the concept of recurrence taking into account together two systems evolving in the same phase space, and expresses graphically all the *simultaneous* occurrences of similar state of the two systems.

Analysis of phase synchronization is an initial step toward measuring of social signals such as empathy and dominance. Many definitions of empathy and dominance can be found in the literature. Our work refers to empathy as emotional reaction of an observer perceiving that another one is experiencing or is about to experience an emotion [19]. Dominance is referred as "expressive, relationally based communicative acts by which power is exerted and influence achieved" [20]. Changes in the number of occurrences and strength of PS among users are considered cues toward evaluation of empathy. These changes in PS are ascribed to changes in the motor expression of the users due to emotion, according to the Scherer's Component Process Model of Emotion [13]. Dominance is computed from the detection of the direction of synchronization. Using an offline PS approach, several results emerged: for example, in the case of a music duo performance, it was possible to evaluate how the visual and acoustic channels affect the exchange of expressive information during the performance [21], [22] and lay foundations to define a criterion to distinguish between parallel and reactive empathic outcomes. Furthermore, by measuring the direction of PS a confirmation of the hypothesis on egalitarian distribution of dominance in a duo performance was provided.

### 3.1. Preliminary Results: the EyesWeb XMI Social Signal Processing software library

We developed the EyesWeb XMI Social Signal Processing Library, a collection of software modules implementing the following algorithms: (i) the RP and the CRP modules extract and display Recurrence/Cross-Recurrence Plots directly from their operative definition [18], (ii) the CPR module computes the Correlation of Probability of Recurrence by cross-correlation of the normalized probabilities of recurrence, (iii) the Leadership Index module computes an index of dominance exploiting the combined approach Recurrence/Cross-Recurrence-ES as described in [22].

Inputs of the RP, CRP and CPR modules can be eMAP features, depending on the components of the state vector describing the user.

The real time measures introduced above are important to perform experiments and for a faster iterative evaluation and validation of the algorithms. Studies aimed at verifying the real time applicability of the algorithms were performed. The automated real time extraction and quantification of empathy and dominance is the final goal that goes beyond the scope of these simple experiments.

## ACKOWLEDGMENTS

This work is partially supported by the European (EU 7th Framework Programme ICT) Project 215749 SAME (www.sameproject.eu).

## REFERENCES

[1] Camurri A., Frisoli A. (Guest Eds.), Special Issue of Virtual Reality Journal on Multisensory Interaction in Virtual Environments, Vol. 10 , No.1, Springer, 2006.

[2] Camurri A. (Ed.), *Proceedings of the International Workshop on KANSEI: The technology of emotion*, Genova, Italy, 1997.

[3] Hashimoto, S. Kansei as the third target of information processing and related topics in japan, In Proceedings of the International Workshop on KANSEI: The technology of emotion, pp. 101–104, 1997.

[4] Vinciarelli A., Pantic M., Bourland H., and Pentland A., Social signals, their function, and automatic analysis: a survey, *In Proc. of 10th International Conference on Multimodal Interfaces, ICMI08,* pp. 61-68. ACM, October 2008.

[5] Pentland A., Social dynamics: signals and behaviour, In *Proceedings of IEEE International Conference on Development Learning, ICDL'04*, October 2004.

[6] Pentland A., Socially aware and computation and communication, *IEEE Computer*, Vol. 38, No. 3, pp. 33–40, 2005.

[7] Hung H., Jayagopi D., Yeo C., Friedland G., Ba S., Odobez J.-M., Ramchandran K., Mirghafori N., and Gatica-Perez N., Using audio and video features to classify the most dominant person in a group meeting, In *Proceedings of the 15th International Conference on Multimedia*, pp. 835-838. ACM, September 2007.

[8] Ambady N., and Rosenthal R., Thin slices of expressive behaviour as predictors of interpersonal consequences: a meta-analysis. *Psychological Bullettin*, Vol. 111, No. 2, pp. 256-274, 1992.

[9] Camurri A., De Poli G., Leman M., Volpe G., Toward Communicating Expressiveness and Affect in Multimodal Interactive Systems for Performing Art and Cultural Applications, *IEEE Multimedia Magazine*, Vol. 12, No. 1, pp. 43-53, IEEE CS Press, 2005.

[10] Jaimovich J. and Knapp R.B., Pattern Recognition of Emotional States During Musical Performance from Physiological Signal. In *Proceedings of the 2009 International Computer Music Conference.* Montreal, Canada, pp. 461-4, August 2009.

[11] Wallbott H.G., Bodily expression of emotion, *Eu. J. of Social Psychology*, Vol. 28, pp.879-896, 1998.

[12] Davidson J.W., What type of information is conveyed in the body movements of solo musician performer?, *Journal of Human Movements Studies*, Vol. 6, pp. 279-301, 1994.

[13] Scherer K., *Approaches to emotions*. Hillsdale NJ: Erlbraum, 1984, ch. On the nature and function of emotion: A component process approach.

[14] Pikovsky A., Rosenblum M., and Kurths J.,*Synchronisation: a Universal Concept in Nonlinear Sciences*, ser. Cambridge Nonlinear Science Series. Cambridge Univ. Press, Cambridge, 2001, vol. 12.

[15] Poincaré H., Sur la probleme des trois corps et les equations de la dynamique, *Acta Math*. Vol. 13, pp. 1-271, 1890.

[16] Eckmann J.P., Kamphorst S. O., and Ruelle D.. Recurrence plots of dynamical system, *Europh. Lett*., 5, pp. 973-977, 1987.

[17] Zbilut J. and Webber Jr C.L., Embeddings and delays as derived from quantification of recurrence plots, *Phys. Lett. A*, 5, pp. 199-203,1992.

[18] Marwan N., Romano M.C., Thiel M., and Kurths J.. Recurrence plots for the analysis of complex systems, *Physics Reports*, Vol. 438, pp. 237–329, 2007.

[19] Stotland E., Matthews K., Sherman S., Hanson R., and Richardson B., *Empathy. Fantasy and Helping*, Beverly Hills: Sage,1978.

[20] Dunbar N.E., Burgoon J., Perceptions of power and interactional dominance in interpersonal relationships, *Journal of Social and Personal Relationships*, Vol. 22, No. 2, pp. 207–233, 2005.

[21] Varni G., Camurri A., Coletta P., and Volpe G. Emotional entrainment in music performance.In *Proceedings of 8th International Conference on Automatic Face and Gesture Recognition, IEEE FG2008*, Amsterdam, The Netherlands, 17-19 September 2008.

[22] Varni G., Camurri A., Coletta P., and Volpe G., Toward Real-time Automated Measure of Empathy and Dominance In *Proceedings of the 2009 IEEE Intl. Conf. on Social Computing*, Vancouver, Canada, August 2009.