

## **Scientific Presentations - Abstracts list**

### **Tuesday, September 2**

#### **Session 1 – Chairman: Alessandro Vinciarelli**

**Title** Social Signal Processing (SSPNet)  
**Speaker** Alessandro Vinciarelli (Idiap Research Institute)  
**Schedule** 10:50 – 11:20  
**Abstract** The ability to understand and manage social signals of a person we are communicating with is the core of social intelligence. Social intelligence is a facet of human intelligence that has been argued to be indispensable and perhaps the most important for success in life. In spite of recent advances in machine analysis of relevant behavioural cues like blinks, smiles, crossed arms, laughter, and similar, design and development of automated systems for Social Signal Processing (SSP) are rather difficult. This paper surveys the past efforts in solving these problems by a computer, it summarizes the relevant findings in social psychology, and it proposes a set of recommendations for enabling the development of the next generation of socially-aware computing.

**Title** Short-term emotion assessment in a recall paradigm.  
**Speaker** Guillaume Chanel (UniGE)  
**Schedule** 11:20 – 11:50  
**Abstract** The work presented in this paper aims at assessing human emotions using peripheral as well as electroencephalographic (EEG) physiological signals on short-time periods. Three specific areas of the valence-arousal emotional space are defined, corresponding to negatively excited, positively excited, and calm-neutral states. An acquisition protocol based on the recall of past emotional events has been designed to acquire data from both peripheral and EEG signals. Pattern classification is used to distinguish between the three areas of the valence-arousal space. The performance of several classifiers has been evaluated on ten participants and different feature sets: peripheral features, EEG time-frequency features, EEG pairwise mutual information features. Comparison of results obtained using either peripheral or EEG signals confirms the interest of using EEG's to assess valence and arousal in emotion recall conditions. The obtained accuracy for the three emotional classes are of 63% using EEG time-frequency features which is better than the results obtained from previous studies using EEG and similar classes. Fusion of the different feature sets at the decision level using a summation rule also showed to improve accuracy to 70%. Furthermore, the rejection of non confident samples finally led to a classification accuracy of 80% for the three classes.

**Title** Structural Analysis of Dynamic Action Patterns associated with Affective States Attribution.  
**Speaker** Stéphane With (UniGE - AS)  
**Schedule** 11:50 – 12:20  
**Abstract** To a large extent, progress in the field of automated facial analysis depends on computer vision engineers accessing high quality video recording of spontaneous emotional expressions that have been manually coded for quantitative analysis. I will present such a corpora created at the Geneva Emotion research Group in the course of my PHD dissertation. Even though behavioural scientists traditionally show a strong bias in favour of functional analysis of human behaviour, it becomes increasingly clear for those interested in emotional communication that we need to conduct more structural analysis research on dynamic expressive displays as they occur in naturalistic settings if we are to develop sound hypothesis regarding the function of

emotional displays. I will present preliminary results of my own research on dynamic multimodal action patterns found to be associated with affective states attribution during judgments studies. 450 dynamic clips of spontaneous emotional expressions from 10 encoders participating in an emotional episode sharing task have been comprehensively coded with the FACS system. Concurrently to FACS action unit's codes, propositional content of participants statements have been annotated with a speech analysis system developed in our lab. Using the T-pattern detection algorithm developed by Magnus Magnusson (2000 ), I will illustrate how the temporal organisation of action patterns rather than the mere absence or presence of specific behavioural code account for judges impression formation. I will also illustrate how speech content acts as a mediating factor between expressive facial displays and their interpretation.

## **Session 2 – Chairman: Jean-Marc Odohez**

**Title** Contextual recognition of Visual focus of attention in meetings  
**Speaker** Jean-Marc Odohez (Idiap Research Institute)  
**Schedule** 10:30 – 11:00  
**Abstract** The presentation addresses the automatic recognition of people gaze (or visual focus of attention, VFOA) in meetings. In group conversation, gaze is one of the most important non-verbal communication cue with functions such as establishing relationships (through mutual gaze), regulating the course of interaction, expressing intimacy, and exercising social control. Thus, recognizing the gaze of people in meeting could increase our understanding of multi-party conversation, enhance technology based tools used for remote communication or for supporting meeting efficiency, or be used as an interesting cue to identify the role or status of people in meetings. However, estimating gaze directly is very difficult, especially if one looks for approaches which do not interfere with natural conversation. In our approach, we use as a surrogate the head pose information estimated from video as the basic cue to recognize the VFOA. However, as the same head pose can be used to gaze at different targets, the VFOA estimation can be improved by modelling the relationship between people's VFOA, the structure of the conversation represented by some conversational events, and other contextual cues related to the group activity.

**Title** Automatic estimation of two facets of social verticality in group meetings from nonverbal cues.  
**Speaker** Daniel Gatica-Perez (Idiap Research Institute)  
**Schedule** 11:00 – 11:30  
**Abstract** I will discuss an initial investigation on the automatic estimation of two aspects of social verticality (status and dominance) in small-group meetings using nonverbal cues. The correlation of nonverbal behaviour with these social constructs have been documented in social psychology, but their value for computational models is, in many cases, still unknown. I will present a study of automatically extracted cues - including vocalic, visual activity, and visual attention cues – and investigate their relative effectiveness to predict the most dominant person and the high-status project manager in four-person meetings from relatively short observations. The results suggest that, although dominance and role-based status are related concepts, they are not equivalent and are thus not equally explained by the same nonverbal cues. Furthermore, the best cues can correctly predict the person with highest dominance or role-based status with an accuracy of around 70%.

**Title** Social appraisal, emotion regulation, and attachment style  
**Speaker** Pascal Vrtička (UniGE - AS)  
**Schedule** 11:30 – 12:00  
**Abstract** Facial expressions play a major role in non-verbal social communication among humans and other primates, because faces do not only provide rapid access to information about the identity, but also about the internal states and intentions of others. However, the interpretation of such information is not always trivial, as the example of Mona Lisa's "mystic smile" nicely shows. My PhD thesis project at the NCCR Affective Sciences in Geneva is thus aimed at investigating how context can change the interpretation of emotional facial expressions.

Thereby, I am looking at the influence of external factors like the social context within which a face is perceived, but also at internal factors like individual personality traits. Using functional magnetic resonance imaging (fMRI) in combination with psychological questionnaires, I can not only grasp the neural signature of context-specific encoding of emotional facial expressions, but also see how the latter is modulated by the personality of scanned participants. During my talk, I will present fMRI data showing that specific expressions in faces are processed differently in the human brain depending on the social context where the faces are perceived, as well as the personality of the individual –as measured by adult attachment style. I will extend these findings by means of behavioural data showing that influences of adult attachment style on appraisal of visual stimuli are especially prominent for social as opposed to non-social pictures. Finally, I will briefly mention preliminary fMRI results suggesting that adult attachment style also influences the employment of different emotion regulation strategies, and this once more predominantly during appraisal of social as opposed to non-social emotional scenes.

**Title** Investigating neural activity of the production of facial expression with electroencephalography (EEG) and electromyography (EMG)

**Speaker** Sebastian Korb (UniGE - AS)

**Schedule** 12:00 – 12:30

**Abstract** The clinical dissociation of emotional and volitional facial movement suggests that voluntary and spontaneous facial expressions may be produced by partly distinct neural bases in the human brain. Voluntary suppression of one's emotional facial expression is a widely used emotion regulation strategy, which however has yet received little attention from cognitive neuroscience. The readiness potential (RP), a slow EEG-potential with negative polarity, has often been reported prior to voluntary movements of the limbs. In a first experiment we have found a RP before posed, self-paced smiles. In a second experiment we investigated the neurophysiological underpinnings of spontaneous smiles (emotional), and suppressed smiles, focusing on the RP. Preliminary results will be presented.

### **Session 3 – Chairman: Andrei Popescu-Belis**

**Title** Towards an automatic content linking device: online document retrieval and display during meetings.

**Speaker** Andrei Popescu-Belis (Idiap Research Institute)

**Schedule** 10:30 – 11:00

**Abstract** A just-in-time document retrieval system for meeting environments is a system that listens to a meeting and displays information about the documents from the participants' history that are most relevant to what is being said. A proof-of-concept prototype developed in IM2.HMI and AMIDA will be demonstrated. The prototype starts by indexing meeting-related documents and segments of previous meetings. During a meeting, the prototype continually retrieves the documents that are most relevant to keywords found automatically using the current meeting speech (online or offline ASR). The various software components communicate using the Hub, a client/server architecture for annotation exchange and storage in real-time developed within IM2.DMA.

**Title** Multimodal authentication: impact of blind dimensionality reduction.

**Speakers** Sviatoslav Voloshynovskiy and Oleksiy Koval (UniGE)

**Schedule** 11:00 – 11:30

**Abstract** In this talk, we consider an authentication framework for independent modalities based on binary hypothesis testing in random projections domain. We formulate a generic authentication problem taking into account several possible decision strategies. Then we review main unimodal authentication architectures based on channel and source coding and highlight the need for random projections for complexity, memory storage and privacy. The authentication performance analysis is accomplished in the scope of Neyman-Pearson framework as well as for an average probability of error for both direct and random projections domains. The upper bounds on the performance are given based on Chernoff distance for generic distribution and

data dimensionality. The random projection performance is also compared to those based on optimal dimensionality reduction derived with the knowledge of statistical models and the gap in performance is evaluated. The impact of noisy modality fusion on the performance is also confirmed by the results of computer simulation.

**Title** Feature selection for audio-visual speech recognition  
**Speakers** Mihai Gurban and Jean-Philippe Thiran (EPFL)  
**Schedule** 11:30 – 12:00  
**Abstract** We present a method for selecting features for audio-visual speech recognition, with an emphasis on the visual modality. Our algorithm is based on information theoretic measures and finds the features with the highest amount of information relevant for the recognition task, while at the same time minimizing redundancy. While our feature selection method is general and can be applied to any classification task, our focus is on multimodal signal processing and on showing how we can use information theoretic measures to quantitatively assess the relevance of features from different modalities.

**Title** Multimodal scene analysis from highly noisy video images: combining gaze and object detection to analyze human-machine interaction  
**Speakers** Basilio Noris (UniGE - AS), François Fleuret (Idiap Research Institute), Aude Billard (EPFL)  
**Schedule** 12:00 – 12:30  
**Abstract** We present a method whereby we track the focus of attention of a speaker during a meeting. The speaker wears a portable camera-hat that gives a first-person view of the scene. Automatic extraction of the direction of the gaze of the speaker is compared with automatic recognition of human faces and of objects of interest in the scene to determine whom or what the person looks at. This is a joint work between IDIAP (F. Fleuret) and EPFL.