Idiap 20th Anniversary   **IM2 Summer Institute 2011**

Scientific talks - Abstracts list
Thursday, September 1st

MOBILE MULTIMEDIA
Chairman: Dr. Daniel Gatica-Perez

Title	Person authentication using face and voice on a mobile phone: from research to reality
Speaker	Dr. Sébastien Marcel (Idiap Research Institute)
Schedule	09:30 – 10:00
Abstract	Since more and more portable devices are now equipped with a microphone and a video camera (while very few devices are equipped with fingerprint or iris scanners), face and voice biometrics are therefore natural choices for mobile biometric authentication. However, mobile biometric authentication is a challenging task since the biometric system should be robust to various realistic conditions. It is also complex since the biometric system should be running entirely on a mobile device with limited resources and without draining the battery.

This talk will present the outcome of the EU FP7 project "Mobile Biometry" (MOBIO http://www.mobioproject.org) that aimed at developing and evaluating core algorithms for mobile face and voice biometric authentication. We will discuss on the lessons learned during the MOBIO project and we will draw guidelines to the light of findings.

Title	Mobile multimedia services based on user attention profiles
Speaker	Francesca De Simone (EPFL)
Schedule	10:00 – 10:30
Abstract	The deployment of third generation mobile communication systems has paved the way for the delivery of mobile multimedia applications. Due to the ubiquity of mobile devices, it can be expected that the utilization and the portfolio of such mobile multimedia applications will continue to increase in the coming years. Clearly, this trend puts significant demands on the enabling technologies, as the related applications require increasingly high data rates that have to operate with rather limited bandwidth. Incorporating user behaviour and related cognition processes into multimedia system design and control will offer options for significantly improving spectral efficiency and user multimedia experience.

In this talk, we present a study of user attention during mobile video consumption. Specifically, the degree of attention given by users to mobile multimedia applications that run on mobile devices adapts to changes in the environmental context. For example, the user behaviour of watching a mobile video in a static environment, such as a residential home, and the slightly more dynamic scenario of watching it while waiting at a bus station may be rather different. Knowledge about user behaviour, in terms of periods when focus is given to the mobile device and when distractions by the environment attract the user, could be used in attention-aware video delivery systems to save bandwidth and reduce energy consumption in the mobile device.

SOCIAL MEDIA
Chairman: Dr. Alessandro Vinciarelli

Title	Vlogcast yourself: exploring nonverbal behavior in social media
Speaker	Joan-Isaac Biel Tres (Idiap Research Institute)
Schedule	11:00 – 11:30

Abstract

Video blogging (vlogging) has evolved from its "chat from your bedroom" initial format to a highly creative form of expression and communication, and represents one of the most popular types of user-generated content on sites like YouTube. Recent research in computational social media, including mining of blogs and online social networks, has made much progress on automatically analyzing text sources. However, human communication is more than the words we write: the nonverbal channel - gaze, facial expressions, body gestures and postures, prosody - plays a key role in the formation, maintenance, and evolution of a number of fundamental social constructs in face-to-face and remote communication settings.

In this talk, I will argue that the nonverbal channel available in vlogging opens several promising research lines in social media, and will present ongoing work towards automatic vlogger analysis from the nonverbal perspective. First, vlogging is multimodal in nature, and I will present audio and visual processing methods to characterize vloggers' nonverbal communicative behaviour from cameras and microphones. Then, I will examine connections between vlogging behaviour and social attention. Finally, I will present work on personality impressions in vlogging and their relation to nonverbal behaviour, and discuss other open issues related to this novel form of interaction.

Title	Driving media engagement
Speaker	Dr. Roelof van Zwol (Yahoo! Research)
Schedule	11:30 – 12:00

Abstract

The true challenge for multimedia is to find a balance between relevancy, freshness, quality, interestingness and diversity in order to provide an engaging rich media experience to the user.

Finding similar and relevant media content given a user query or sample image has been at the core of the multimedia retrieval community for a long time. In this talk, I will identify and address multimedia challenges that play a role at Yahoo!, and which go beyond relevancy of images and video to a given multimedia retrieval task. In particular I will address the role of images in context of search recommendations, and demonstrate how computer vision, and user-click feedback can be used to drive the user engagement with this component on the Yahoo! search engine.

LARGE SCALE MULTIMEDIA INDEXING AND COLLABORATIVE DESIGN
Chairman: Dr. Ronan Collobert

Title	Large-scale multimedia retrieval: distributing multimodal interactive learning
Speaker	Dr. Stéphane Marchand-Maillet (Viper - Geneva - IP1)
Schedule	13:00 – 13:30

Abstract   Along IM2, we have demonstrated that multimedia retrieval could largely be achieved thanks to user feedback interpretation. Machine learning strategies such as Boosting can be designed to help in performing information fusion to gather and exploit every piece of knowledge the user is providing to the system.

As a natural extension of these working mechanisms, we have more recently attacked the problem of scalability in multimedia retrieval. We look at how computation and information access may be scheduled over a network of computers with distributed storage to preserve the usability and usefulness of our tools when applied over large collections of items bearing multimodal information.

In this presentation, we therefore emphasise and review both aspects of retrieval performance and robustness against the increase in the scale of the dataset and in the complexity of the data. We summarise our achievements that have already resulted into concrete developments.

Title	Large scale image annotation: learning to rank with joint word-image embeddings
Speaker	Dr. Samy Bengio (Google, USA)
Schedule	13:30 – 14:00

Abstract   Image annotation datasets are becoming larger and larger, with tens of millions of images and tens of thousands of possible annotations. In this talk, I'll discuss a strongly performing method that scales to such datasets by simultaneously learning to optimize precision at k of the ranked list of annotations for a given image and learning a low-dimensional joint embedding space for both images and annotations. Our method both outperforms several baseline methods and, in comparison to them, is faster and consumes less memory. We also demonstrate how our method learns an interpretable model, where annotations with alternate spellings or even languages are close in the embedding space. Hence, even when our model does not predict the exact annotation given by a human labeller, it often predicts similar annotations.

Title	Automatic content linking: a speech-based real-time retrieval system using semantic search
Speaker	Dr. Andrei Popescu-Belis (Idiap Research Institute)
Schedule	14:00 – 14:30

Abstract   The Automatic Content Linking Device (ACLD) is a just-in-time retrieval system that monitors an ongoing conversation or a monologue and enriches it with suggestions for related documents. The suggestions are found using a semantic similarity measure between the words obtained from automatic speech recognition (ASR) and the documents in a repository -- a method that improves over keyword-based search. The ACLD can also be demonstrated simply with speech from the person giving the presentation.

INTERFACES AND INTERACTIONS
Chairman: Dr. Andrei Popescu-Belis

Title   Cross lingual speaker adaptation

Speaker   Phil Garner, Dr. John Dines, and Hui LIANG (Idiap Research Institute)

Schedule   15:00 – 15:30

Abstract   Recent advances in text to speech synthesis (TTS) have brought the underlying technology closer to that of automatic speech recognition (ASR). This in turn has allowed peripheral techniques from ASR to be used in TTS. One such technique is speaker adaptation. This allows a great variety of synthetic voices to be created from a single underlying model, and with comparatively little training data. During the EMIME project (http://www.emime.org) we developed techniques to extend this adaptation, allowing synthetic voices in languages other than that of the adaptation data. Coupled with ASR and translation, this allows a given speaker to appear to be speaking a different language. We will give an overview of the techniques, augmented and illustrated with audio examples.

Title   Epoch-based analysis of speech and its applications

Speaker   Prof. Bayya Yegnanarayana (International Institute of IT Gachibowli, Hyderabad, India)

Schedule   15:30 – 16:00

Abstract   Speech analysis is traditionally performed using short-time analysis to extract features in time and frequency domains, using fixed window size for the analysis. But speech in its primary mode of excitation is produced due to impulse-like excitation in each glottal cycle. Anchoring the speech analysis around the glottal closure instants (GCIs or epochs) yields significant benefits for speech analysis. Epoch-based analysis of speech helps not only to segment the speech signals based on speech production characteristics, but also helps in extracting important acoustic-phonetic features such as glottal vibrations, formants, instantaneous fundamental frequency, etc.

Epoch sequence is useful to manipulate prosody in speech synthesis applications. Accurate estimation of epochs helps in characterizing voice quality features. Epoch extraction also helps in speech enhancement and multispeaker separation. In this talk methods to extract the epoch information are reviewed, and some applications of epoch extraction from speech signals are demonstrated.

Title   Measuring gaze behavior in real environments: applications to research and industry

Speaker   Dr. Basilio Noris (EPFL)

Schedule   16:00 – 16:30

Abstract   Instruments for the measurement of gaze have been used in a number of different fields for several decades, from psychophysics to Human-Machine Interfaces. In its most common form, these devices allow to track gaze in static environments (e.g. in front of a monitor, integrated in a larger machine), but some wearable system have appeared that allow studying visual behaviour in the real world. This has created opportunities for research as well as for industrial applications. We will present the main techniques behind gaze tracking and some of their implementation from high-end commercial solutions to low-budget open source projects. We will put an emphasis on wearable systems and their applications in the domains of psychology, assistive technologies and commercial and industrial applications.

EVALUATION AND USABILITY
Chairman: Dr. Denis Lalanne

Title	The importance of evaluation for multilingual information access
Speaker	Prof. Carol Ann Peters (Institute for Information Science and Technologies, Italy)
Schedule	16:30 – 17:00

Abstract The aim of large-scale evaluation campaigns is to build a strong research community and to stimulate advances in the state-of-the-art in a given domain. The main goal of the Cross Language Evaluation Forum has been to sustain the growth of excellence in language processing and multilingual information access (MLIA) systems and to encourage the creation of a consolidated research community in this multidisciplinary area.

The talk will thus trace the main steps taken by CLEF over the years to encourage the development of multimodal multilingual retrieval systems and will present the most significant results. We will describe how in recent years attention has been given not only to advances in research but also to technology transfer with the organization of tasks designed to meet the needs of specific application communities. Finally, we will discuss new factors and trends influencing the field. Since CLEF began the associated technologies, services and users of multilingual IR systems have been in continual evolution. The expectations and habits of users are constantly changing, together with the ways in which they interact with content and services, often creating new and original ways of exploiting them. The evaluation campaigns of the future must be able to successfully scale up and embrace new communities and technological paradigms.

Title	What is beautiful is usable - aesthetics and other factors that influence the outcomes of usability tests
Speaker	Dr. Andreas Sonderegger and Prof. Juergen Sauer (UNIFR)
Schedule	17:00 – 17:30

Abstract System usability is a central product element during the design of technical systems. System usability is very often evaluated by means of a usability test. In a typical usability test, a test user completes different tasks with the system to be tested. A number of studies have shown that aspects that are unrelated to system usability may influence measures of system usability. It is therefore important for usability practitioners and system developers to take these influencing factors into account when interpreting results of usability tests during the development of a new system.

For example, our work indicated that the laboratory set-up (e.g., presence of observers or cameras) has an influence on how users behave and feel during the usability test. Further studies showed that product aesthetics also plays an important role in the evaluation of system usability in that aesthetically pleasing systems are also considered to be more usable by test users. Interestingly, it emerged that the influence of such factors may wane over time if a longitudinal approach is adopted in usability testing. Based on the findings of our empirical work, a number of recommendations are made to increase the effectiveness of usability testing.