RESEARCH INSTITUTE

SCIENTIFIC REPORT

March 2018

Contents

List of Figures

1 Idiap Structure and Research Areas

1.1 Research Areas

Idiap's research activities span five broad areas: perceptual and cognitive systems, human and social behavior, information and presentation interfaces, biometrics, and machine learning, which are summarized, with a set of descriptive keywords, in Figure 1 below.

In addition to these transverse research themes, new activities have been recently developed, towards bio-medical applications (through the new group in Computational Bioimaging) and environmental modeling (through the new group in Uncertainty Quantification and Optimal Design).

Idiap maintains a policy of continuous growth and diversification in new and complementary research directions, while building upon its long-standing institutional theme of "Human and Media Computing".

Research Areas	***Keywords***
Perceptual and cognitive systems	*Speech and audio processing, computer vision, document processing, robotics, natural language processing, machine translation, computational cognitive science*
Human and social behavior	*Social media, social sensing, verbal and nonverbal communication analysis, smartphone sensing, computational social science*
Information and presentation interfaces	*Multimedia information systems, user interfaces, personalization, system evaluation, mobile HCI using big data, data driven services*
Biometrics Security and Privacy	*Face recognition, speaker recognition, vein recognition, multimodal fusion, soft-biometrics, remore photoplethysmography, presentation attack detection (anti-spoofing), template protection, mobile and wearable biometrics*
Machine learning	*Statistical and neural network based ML, computational efficiency, online learning, multiple sensor processing, big data*

Figure 1: Overview of Idiap research areas. These research areas are currently covered by nine research groups, listed in Section 1.3 below, and are designed to be as pro-active as possible in covering the applications areas presented in Figure 2.

1.2 Application Areas

Idiap's application areas are listed in Figure 2, together with keywords and pointers to relevant projects. Idiap contributes to a broad range of applications areas, which reflect concrete, ongoing projects.

We would like to draw attention to the difference between our *research* and our *application* areas. While Idiap's research areas (Figure 1) are quite generic and have a strong transverse nature, the application areas are more targeted and take place along well-defined application axes. Although we do not necessarily address at Idiap all aspects within a given application area, we are often in the best position to provide the necessary enabling Information and Communication Technology components.

Application Areas	***Keywords***
Human-human & human-machine interaction	*Voice and gesture controlled devices and robots, hand-free control, spoken language systems, translation systems, social robotics, user profiling*
Exploitation of rich multimedia archives (audio, video, text)	*Hyper-events, semantic indexing, keyword spotting, object detection and recognition, image bank browsing, audio-video content filtering (summarization and recommendation), broadcast data analysis, scanned document analysis, analysis of cultural heritage media*
Collaborative and creative systems	*Remote meeting assistance, smart meeting room, video-conferencing, multimedia indexing and access, cross-lingual collaboration, interaction analysis, dynamics of negotiation*
Healthcare and bio-engineering	*Smart management of patient data, prosthesis (hearing aids, artificial body parts, voice banking), bio-systems modeling, interfaces for impaired users (speech and other inputs), behavior-based health diagnosis, bio-medical document processing, smartphone platforms for health, bio-imaging, microscopy*
Entertainment	*Multilingual gaming, remote family games, togetherness*
Mobile computing	*Signal processing for mobile platforms, mobile social networks, participatory sensing*
Security and risk management	*Biometric security, access control, mobile biometry, multi-sensor fusion, speaker identification, video monitoring of areas/activities, natural risk modeling, intrusion detection, crowd management*
Home automation (domotics)	*Multi-sensor activity analysis, adaptation to users' behavior, efficient use of energy, home safety and security*
Energy	*Energy grids, multiple sensor and smart meter networks, large-scale sensor data integration, modeling of behaviors to anticipate demand, safer, cheaper, and cleaner energy production*
Smart cities	*Ecology, environment management, reduction in pollution, traffic and noise, better use of roads*

Figure 2: Idiap application areas with several examples for each of them.

1.3 Internal Structure

1.3.1 Overview

The main research themes, briefly presented in Section 1.1 above, are currently covered by nine research groups, presented in the following section. The resulting structure is presented in Figure 3 along with the overall structure of Idiap. The leaders of the groups, who are all PIs of research projects, are explicitly indicated.

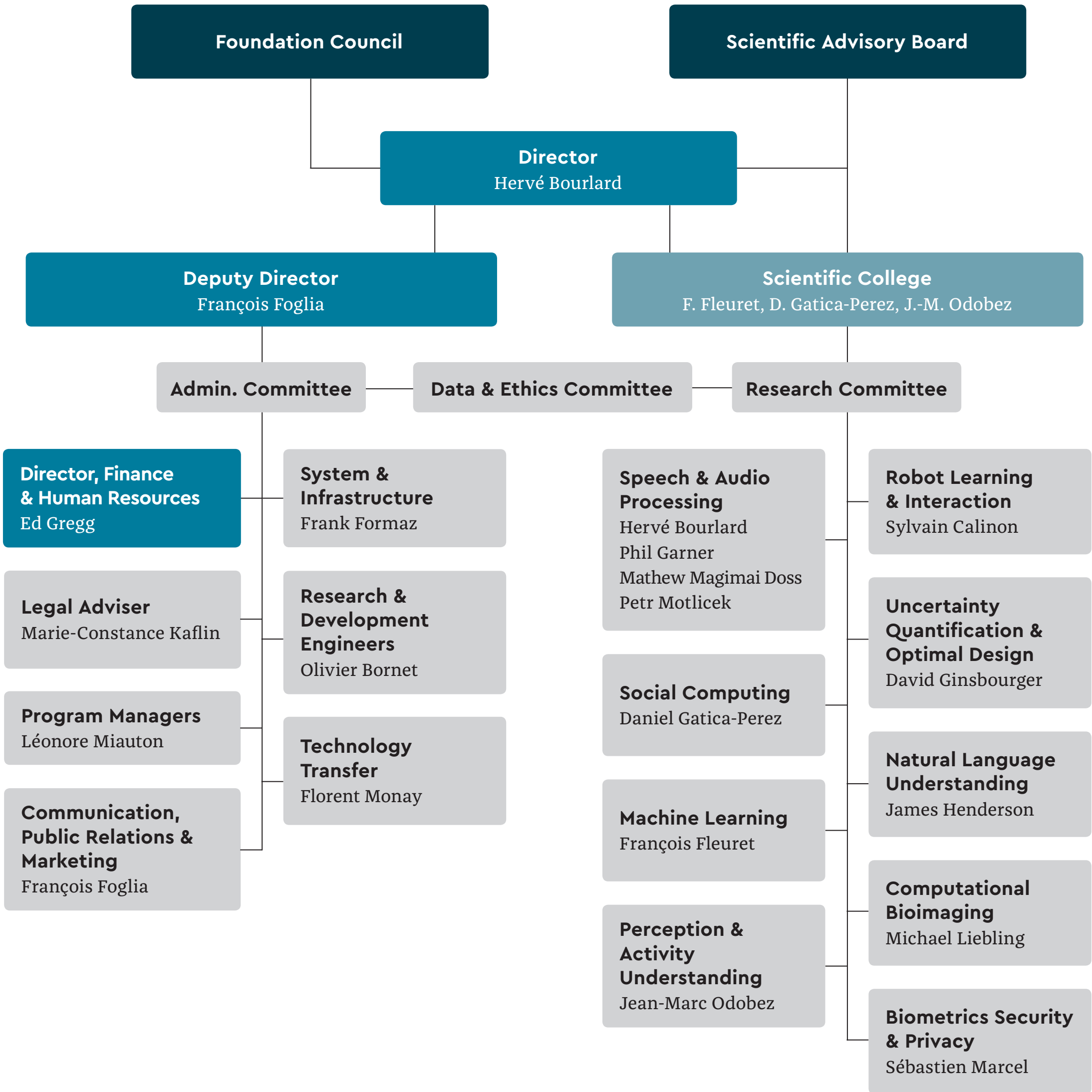

Figure 3: Idiap management and operational structure, including the main research and administrative responsibilities. For the research block, all names in white are key researchers, with clear supervision roles, and are all project PIs. On the admin side, the names in italics refer to positions that should still be filled, depending on needs and funding availability. In 2017, a legal adviser and a "Data & Ethics Committee" were added to the organisation chart.

1.3.2 Research Groups

In 2017, the research areas presented in Figure 1 were covered by nine research groups. The activity of each research group can be summarized as follows:

1. **Speech and Audio Processing (Prof. Hervé Bourlard, Dr. Phil Garner, Dr. Mathew Magimai-Doss, & Dr. Petr Motlicek)**

   Speech processing has been one of the mainstays of Idiap's research portfolio for many years. Today it is still the largest group within the institute, and Idiap continues to be recognized as a leading proponent in the field. The expertise of the group encompasses multilingual speech recognition, multilingual text-to-speech conversion, and generic audio processing – covering sound source localization, microphone arrays, speaker diarization, audio indexing, very low bit-rate speech coding, and perceptual background noise analysis for telecommunication systems.

   See Section 2.1 for the 2017 progress report.

2. **Machine Learning (Dr. François Fleuret)**

   The goal of the Machine Learning group is the development of new statistical learning techniques with a particular interest in their computational properties. Our application domain encompasses any processing of real-world signals, such as image understanding, detection of persons and biological structures, signal synthesis, or decision from low-level industrial sensors.

   See Section 2.2 for the 2017 progress report.

3. **Social Computing (Prof. Daniel Gatica-Perez)**

   Social computing is an interdisciplinary domain that integrates theories and models from ubiquitous computing, multimedia, machine learning, and social sciences, to sense, analyze, and interpret human and social behavior in everyday life, and to create systems that support social interaction. Current lines of research include social media analytics and mobile crowdsourcing for cities and health, analysis of ubiquitous social interaction, and analysis of cultural heritage media.

   See Section 2.3 for the 2017 progress report.

4. **Perception and Activity Understanding (Dr. Jean-Marc Odobez)**

   The group investigates models from machine learning, computer vision, multimodal signal processing, or social sciences, to address the understanding of activities from real-world signals, with an emphasis on those related to humans. Detection, tracking, pose estimation, recognition and analysis of non-verbal behaviors or the temporal interpretation of all this information in forms of gestures, activities behavior or social relationships are examples of studied tasks. Application domains encompasses surveillance, traffic and human behavior analysis, human-robot interactions, and multimedia content analysis.

   See Section 2.4 for the 2017 progress report.

5. **Robot Learning and Interaction (Dr. Sylvain Calinon)**

   This group focuses on human-centric robotic applications in which the robots can learn new skills by interacting with the end-users. From a machine learning perspective, the challenge is to acquire skills from only few demonstrations and interactions, with strong generalization demands. It requires the development of intuitive active learning interfaces to acquire meaningful demonstrations, the development of models that can exploit the structure and geometry of the acquired data in an efficient way, and the development of adaptive control techniques that can exploit the learned task variations and coordination patterns.

   See Section 2.5 for the 2017 progress report.

6. **Uncertainty Quantification and Optimal Design (Dr. David Ginsbourger)**

   The Uncertainty Quantification and Optimal Design group focuses on quantifying and reducing uncertainties in the context of natural and artificial complex systems. Application domains notably include energy and geosciences, with a number of collaborations ranging from safety engineering to hydrology and climate sciences. In all these fields the study of complex systems often relies on expensive data acquisition and model runs, calling for adapted experimental design strategies.

   See Section 2.6 for the 2017 progress report.

7. **Computational Bioimaging (Dr. Michael Liebling)**

   This group focuses on research in computational imaging and analysis of biomedical images. This includes developing algorithms for image deconvolution and super-resolution in optical microscopy, three-dimensional tomography reconstruction from projections, and, more generally, combining unusual sensing devices and approaches with computational methods to produce images ideally-suited for the observation and quantification of complex and live biological systems.

   See Section 2.7 for the 2017 progress report.

8. **Biometrics Security and Privacy (Dr. Sébastien Marcel)**

   Biometrics refers to the automatic recognition of individuals based on their behavioral and biological characteristics. The Biometrics Security and Privacy group investigates and develops novel image-processing and pattern-recognition algorithms for face recognition (2D, 3D, and near-infrared), speaker recognition, anti-spoofing (presentation attack detection), and emerging biometric modes (EEG and vein). The group is geared toward reproducible research and technology transfer, using its own signal-processing and machine-learning toolbox.

   See Section 2.8 for the 2017 progress report.

9. **Natural Language Processing (Dr. Andrei Popescu-Belis)**

   The Natural Language Processing group studies how semantic and pragmatic analyses of texts can improve the execution of two important tasks – machine translation and information retrieval. The group also studies how the search for information over networked data, including multimedia, can be improved by using semantic information and information from the network itself.

   See Section 2.9 for the 2017 progress report.

10. **Natural Language Understanding (Dr. James Henderson)**

    The Natural Language Understanding group (created September 2017) studies deep learning for natural language processing tasks, focusing on models with learned representations of the meaning of text. It continues the NLP group's work on neural machine translation and information retrieval, and extends to neural network structured prediction and representation learning for modelling the syntax and semantics of text and speech, including modelling abstraction (textual entailment) and summarisation.

    See Section 2.10 for the 2017 progress report.

1.3.3 Administration and Services

At the administration and services level, there are seven groups: finances and human resources; communication, public relations and marketing; system and infrastructure; technology transfer; development engineers; program managers; and legal advisor. The resulting admin organization is presented in Figure 3, page 3. The main responsibilities of the seven admin groups can be summarized as follows:

1. **Finance and Human Resources (Ed Gregg)**: The Human Resources Department (HR) is integrated within the financial and accounting activities of Idiap and has taken on a greater importance in the past year. With employees from over 30 different countries, the finances and HR department is continually growing to meet the needs of each employee.
2. **Communication, Public Relations & Marketing (Dr. François Foglia)**: The mission of the communication, public relations and marketing department is to use all forms of media and communication to build, maintain, manage the reputation of the Institute, and to promote the Idiap services available for external institutions, such as EU project management, submission proposal tools, etc.
3. **System and Infrastructure (Frank Formaz)**:The main mission of the system and infrastructure group is to provide an optimal and efficient work environment for the Idiap collaborators. The tasks can be split into three main activities covering (1) centralized IT services for the whole Institute (network, storage, servers, workstations, high performance computing, identity management, data distribution), (2) support for collaborators (helpdesk, project specific tasks, web presence), and (3) Infrastructure (building, offices, equipments, central purchasing office).
4. **Technology Transfer (Dr Florent Monay)**: Technology transfer is one of the Idiap Research Institute's three core missions. One of the fundamental challenges is to facilitate the interface between the knowledge and the skills of the researcher and the needs of the industrial partner. Idiap resolves this by providing a dedicated multi-disciplinary team of developers and programmers which transfers pieces of software, algorithms, knowledge and expertise. This transfer of technology is usually done by granting rights on the commercial exploitation of this technology (through license).
5. **Program Managers (Léonore Miauton)**: The work of the program management team is divided into two types of activities. The first is the provision of services to researchers within the framework of European and Swiss projects. The second category includes activities ranging from event organization to database management, which are not directly linked to the management of research projects but facilitate the work of Idiap researchers.
6. **Development Engineers (Olivier Bornet)**: The mission of the development team is to provide support to Idiap researchers in the software development tasks. This is done in three areas. The first is to help on Idiap research by building prototypes, implement algorithms, design and run experiments, and manage legacy code. The second area is for all the technology transfer tasks. In the third activity, development engineers give daily support to Idiap researchers (software disclosures, showroom and internal demonstrators, development tools).
7. **Legal Adviser (Marie-Constance Kaiflin)**: The main missions of the legal adviser are to write, analyse, and negotiate project contracts (research, consortium, consultant agreements, NDA, Memorandum of Understanding, etc) or technology transfer contracts (patents, knowhow, licenses) with industries, universities or research institutions. The legal adviser deals also with all the legal aspects related to human resources (work contracts, staff regulations, rules) and data protection (ethics, databases collection and distribution).

2 Research Groups

2.1 Speech and Audio Processing

Overview

Heads: Prof. Hervé Bourlard (MS and PhD, Polytechnic University, Mons, Belgium, 1982 and 1992), Dr. Philip N. Garner (MEng, University of Southampton, UK, 1991; PhD, University of East Anglia, UK, 2011), Dr. Mathew Magimai-Doss (MS by Research, Indian Institute of Technology Madras, India, 1999; PhD, Ecole Polytechnique Fédérale de Lausanne, Switzerland, 2005), Dr. Petr Motlicek (MS and PhD, Brno University of Technology, Czech Republic, 1999 and 2003).

Group overview: Speech processing has been one of the mainstays of Idiap's research portfolio for many years, covering most of the aspects of speech processing such as multilingual automatic speech recognition (ASR), speech synthesis, speech coding, speaker identification, automatic speech intelligibility evaluation, or speech processing for classification of motor speech disorders. The expertise and activities of the group encompass statistical automatic speech recognition (based on Hidden Markov Models-HMM, or hybrid systems exploiting Deep Neural Networks-DNN and new deep learning architectures), text-to-speech (TTS), speaker recognition (with extensions towards text-dependent and forensics scenarios) and generic audio processing (covering sound source localization, microphone arrays, speaker diarization, audio indexing, very low bit-rate speech coding, perceptual background noise analysis for telecommunication systems) and, more recently, Compressive Sensing (CS) and Sparse Recovering theories applied to ASR.

The Speech and Audio Processing group in 2017 was composed of 1 head of group, 3 principal investigators, 3 research associates, 6 postdocs, 9 PhD students and 4 interns.

Key scientific outputs: Idiap has always significantly contributed to both Hidden Markov Model (HMM) and Deep Neural Network (DNN) based approaches applied in acoustic modelling for various speech processing tasks. Use of techniques from HMM and HMM-DNN based speech recognition in HMM and HMM-DNN based speech synthesis resulted in a unified approach to speech recognition and synthesis. The group was well placed to take full advantage of recent advances in new architectures of deep learning, studied in particular through *PyTorch* and other open source frameworks. Advances in Automatic Speech Recognition (ASR) are usually researched through *Kaldi* toolkit, now used by most of the international speech community.

In 2017, several key contributions were achieved by the group, including: (1) multilingual speech recognition, especially in cross-lingual adaptation, and speech recognition in low-resourced language conditions, (2) speaker recognition, through both text-independent and particularly text-dependent speaker verification scenarios and information fusion for large-scale speaker identification, (3) large scale media processing, including multilingual broadcast news recognition, and spoken query for spoken term detection, and (4) development of novel deep neural network (DNN) architectures for multilingual speech recognition.

The group has also started exploiting new Compressive Sensing (CS) and Sparse Recovering theories to automatic speech recognition, developing new theoretical links between CS and statistical/HMM-DNN approaches, resulting in improved ASR performance, as well as new spoken term query detection algorithms.

Additional information and a list of projects are available from www.idiap.ch/speech.

Automatic speech recognition

In recent years, our ASR research activities have been expanded from mono-lingual to cross-/multi-lingual processing. More specifically, in addition to focusing on "majority" languages other than English such as, French, German, Idiap is actively carrying research in several ASR directions, including:

- **Robust parametrisation and acoustic modelling**: We are still investigating new features (e.g., posterior-based features) and new acoustic models (new forms of hidden Markov models, such as KL-HMM, or artificial neural networks) that are more robust to noise and acoustic environments, as well as to speaker variability (e.g., accented speech, or dialect).
- **Cross-lingual and multi-lingual speech recognition**: The EC SUMMA project (Scalable Understanding of Multilingual Media), as illustrated in Figure 4, aims at integrating stream-based media processing tools (including speech recognition and machine translation) with deep language understanding capabilities (including named entity relation extraction and semantic parsing), for open-source applications and implemented in use cases at the BBC and Deutsche Welle. The US IARPA SARAL project (Summarization and domain-Adaptive Retrieval of Information Across Languages) aims at developing cross-lingual retrieval and summarization techniques that will work for any language in the world, given minimal resources to work with. In those contexts, we focus on investigating and exploiting fast acoustic model adaptation techniques in cross-lingual and multi-lingual scenarios. The resulting speech recogniser relies on a hybrid approach, where an artificial neural network acoustic model is boot-strapped using well-resourced data and adapted to the target language.

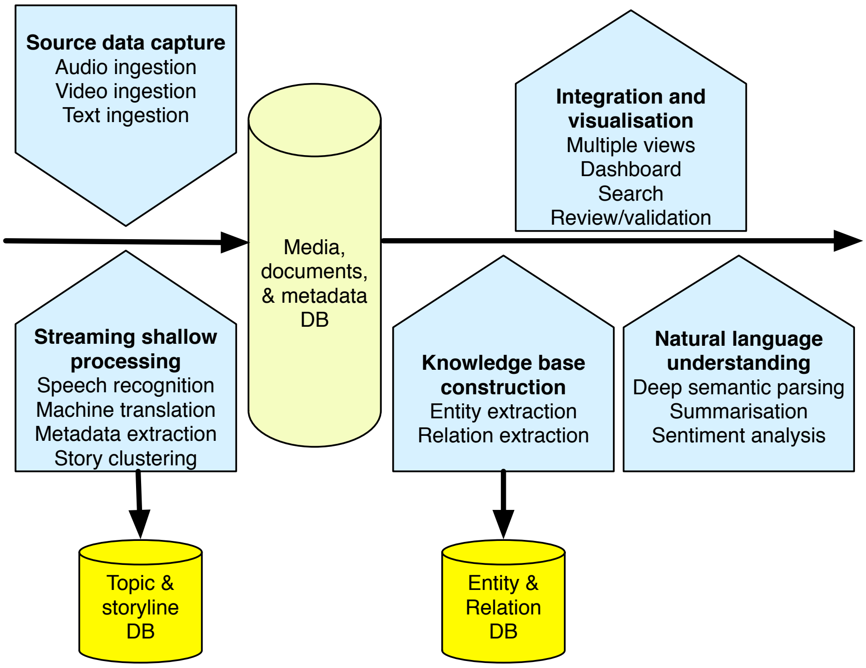

Figure 4: An overview of the EC H2020 SUMMA project, a current focus for multilingual speech recognition at Idiap.

- **Swiss languages**: We continuously improve our speech recognisers for Swiss German and Swiss French and also apply the most recent advances in speech technology employing deep neural networks. The recognisers are also used in commercial products of the Idiap spinoff *recapp IT AG*. Work on speech recognition for Walliserdeutsch one of the most difficult to understand of the Swiss dialects, was first published in 2014; the associated database is also available for download. Since 2015, we collaborate with *recapp IT AG* on a wider range of Swiss dialects

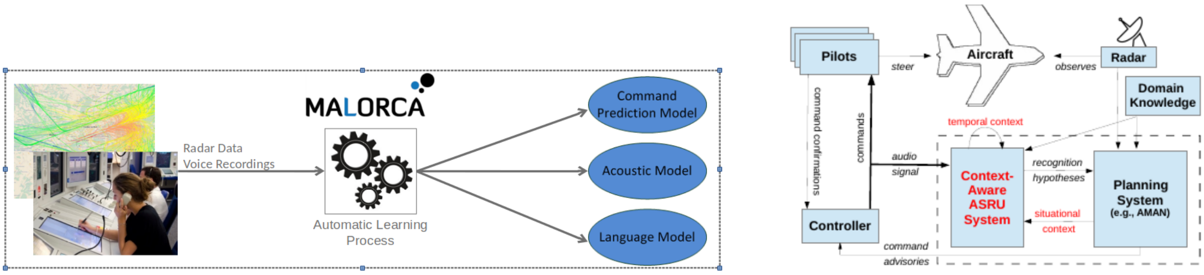

Figure 5: Left: Overview of the EC MALORCA project; semi-supervised adaptation of three models on the left side. Right: graphical representation of building blocks.

towards the first commercial product that performs Swiss German (dialect) speech recognition. A collaboration in 2017 with Swisscom enabled us to extend our Swiss German work to server-based large vocabulary situations, and in tandem with translation to English and other Swiss languages.

- **Template-based ASR**: Idiap continues to work on template-based or exemplar-based ASR using posterior-based features in the context of projects like SNSF projects A-MUSE, PHASER, and PHASER-QUAD, and CTI project MultiVEO[1]. In this context, there are currently two on-going efforts. The first focuses on development of "sparse" and "low-rank" HMM frameworks by establishing a link between statistical speech recognition formalism and the recent theories of sparse modelling and (probabilistic) principle component analysis (PCA). The framework integrates the advantages of template-based modelling within the HMM sequence matching scheme. The second effort focusses on development of an approach that seamlessly combines both statistical sequence model based ASR system and template-based ASR system into a single ASR system that can yield performance better than the best individual system.
- **Lexicon development**: Over the past six years, Idiap has been conducting research on the use of alternative subword units, specifically graphemes, and development of an ASR approach which is capable of handling both acoustic and lexical resource constraints by learning grapheme-to-phoneme relationship through the acoustic speech signal (funded through the SNSF-project FlexASR and HASLER Foundation project AddG2SU).
- **Semi-supervised learning of acoustic and language model parameters**: This problem has been conducted at Idiap over several past years, as it aims to develop a technology to (1) automatically select speech data which is acoustically highly informative, or (2) to automatically assess recognition output with high confidence. This type of technology is esential to allow semi-supervised learning of speech recognition models using unlabelled data. One of the specific architecture has been developed within the EC H2020 MALORCA project[2], combining both acoustic (i.e. speech) and situational (i.e. radar) information to build automatic speech recognition system for air-traffic controllers (i.e. controller-pilot communication). The MALORCA project, graphically represented in Figure 5 further focused on technologies to automatically adapt the developed system for new airports, while minimising human effort to manually transcribe speech data.
- **Punctuation prediction**: As our ASR output becomes an input for processing at higher semantic levels, for instance in the SUMMA project, other meta-data such as punctuation becomes necessary. ASR does not normally produce punctuation, but it is possible using similar techniques, notably language modelling and pause detection. More recently, this is influenced by the encoder-decoder approaches used in machine translation.

[1] https://www.idiap.ch/scientific-research/projects/multiveo
[2] http://www.malorca-project.de

- **Failure analysis of ASR and HSR decoding channels**: This is a novel strategy further developed in 2017 to identify the key sources of information loss in ASR and human speech recognition (HSR). This approach builds on the power of DNN in probabilistic characterization of the sub-word classes constituting a language. We cast ASR and HSR as separate channels decoding the sequence of sub-word components from their probabilistic representation. Information theoretic measures are developed to assess and quantify the information loss in acoustic modeling for ASR decoding using hidden Markov models.

Speech synthesis and coding

- **Text-to-speech synthesis (TTS)**: Although newer than ASR, TTS is now an established venture for the speech group at Idiap. TTS has been central to several projects, including: SIWIS[3] (Spoken Interaction With Interpretation in Switzerland), D-BOX[4] and SP2[5] (SCOPES project on speech prosody). The TTS work at Idiap was largely aligned with the statistical synthesis trend, which uses the same technology as ASR. However, the group has tracked the recent developments in deep learning which will dominate future research. SIWIS aimed to do speech to speech translation in Swiss languages, including the concept of accents. Current work under MASS (Multilingual Affective Speech Synthesis) brings the concept of emotion into the speech synthesis, particularly via modelling of prosody.
- **Speech coding**: Another research area requiring TTS is speech coding, where very low bit rates can be achieved by using a speech recogniser followed by a synthesiser. Previously, under the RECOD project funded by Armasuisse, the technology lends itself to operate at 200–300 bits per second. The solution relies on deep learning characterization of the phone attribute features dubbed as phonological posteriors. Recent advancement building on the findings of the SNSF project PHASER, led to increased efficiency of the architecture. Unique structures of the phonological posteriors are identified as the sparse pronunciation codes composing natural speech communication; a small size codebook is thus constructed and used for ultra low-bit-rate speech coding. Moreover, work on speech coding continued under the SP2 project, mainly on aspects concerning prosody. This led to the release of the "PhonVoc" toolkit, an end-to-end neural network based phonetic and phonological vocoder.

Speaker recognition and speech data analytics

In the context of the European SIIP project[6], illustrated by Figure 6, the Speech and Audio Processing group has significantly improved their capabilities in suspect identification applicable to very large scale data. The SIIP technology has successfully passed three proof-of-concept and field-test events in 2016 and 2017, demonstrating the performance of the developed technology among the key stakeholders. The developed suspect identification solution can analyse not only lawfully intercepted calls, but also multiple types of social-media channels. SIIP has also developed a framework allowing to combine evidences extracted by different types of engines (i.e. inter-task engines such as language/accent/gender/age identification, or keyword-spotting) to eventually improve speaker identification.

As discussed in Section 2.8, and as part of the SNSF Project UniTS,[7], the group is also contributing to the Biometric Person Recognition group, including the development of countermeasures to detect attacks on speaker verification systems through forged speech samples.

[3] http://www.idiap.ch/project/siwis/
[4] http://www.idiap.ch/scientific-research/projects/dbox
[5] http://www.idiap.ch/scientific-research/projects/sp2
[6] http://www.siip.eu
[7] https://www.idiap.ch/scientific-research/projects/units

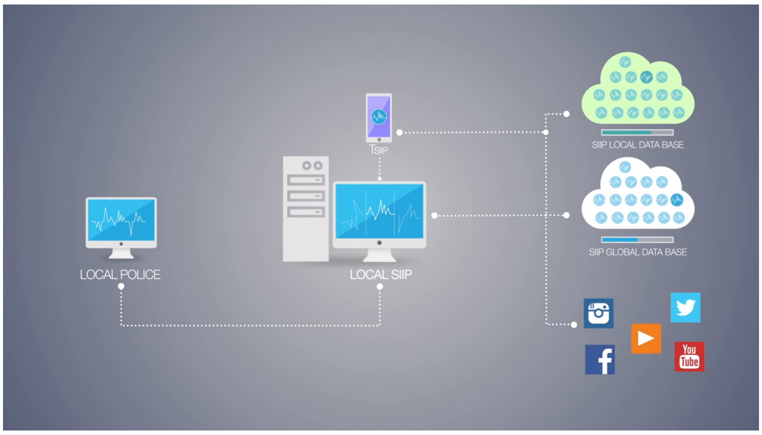

Figure 6: Illustration of SIIP speaker identification system exploiting lawfully intercepted calls, including multiple types of social-media information streams.

Forensic speaker verification and audio analysis

In 2017, Idiap continued its work in audio forensic sciences, aiming to extract relevant evidences from speech and audio signals that may be ultimately presented as admissible fact in a court of law. Besides a task of enhancement of speech recordings to improve the intelligibility, we have mostly focused on forensic voice comparison to determine the identity of the speaker associated with the strength of evidence. Idiap collaborates with various law enforcement agencies in Switzerland (e.g. Federal and cantonal Police) and abroad to provide direct support in analysing strength of evidences.

Large scale spoken query retrieval

- **Query-by-example spoken term detection (QbE-STD) based on subspace modelling and detection**: Mainly exploiting its knowhow in Deep Neural Networks (DNN) and sparse recovery modeling, Idiap continues its research efforts towards keyword spotting and spoken term detection with a focus on searching large audio archives using spoken queries. Over the past 12 months, Idiap has developed pioneering retrieval solutions relying on the characterization and detection of the low-dimensional subspace of the DNN phonetic posteriors.
- **Large scale spoken query indexing**: In 2017, Idiap further developed a powerful methodology for large-scale spoken query retrieval relying on hashing. Contribution of hashing is two-fold: (1) Compressing the information bearing data characteristic and (2) Identifying a data-driven symbolic space for effective search. Idiap hashing solution relies on DNN representation learning at sub-phonetic attribute level to facilitate cross-lingual applications.
- **Objective speech intelligibility assessment**
  Over the past two years, Idiap has been actively developing novel methods to objectively assess intelligibility of speech signal. This work is now exploited in the context of the pathological speech processing activities discussed below.

Pathological speech processing

In 2017, a new research track, focusing on the processing of pathological speech, was initiated at Idiap, mainly through two new projects:

- MoSpeeDi (Motor Speech Disorder) SNSF Sinergia project: aiming at characterizing phonetic speech planning and motor speech programming/execution and their impairments, in collaboration with Marina Laganaro, Faculty of Psychology and Educational Science, University of Geneva, Cécile Fougeron, Laboratoire de Phonétique et Phonologie, Paris 3-Sorbonne Nouvelle, and

Frédéric Assal, Neurology, Geneva University Hospitals and Faculty of Medicine, University of Geneva.

- EC TAPAS (Training Network on Automatic Processing of PAthological Speech): targeting three key research problems, (1) detection (develop speech processing techniques for early detection of conditions that impact on speech production), (2) therapy (using newly-emerging speech processing techniques to produce automated speech therapy tools), and (3) assisted Living (redesigning current speech technology so that it works well for people with speech impairments and also helps in making informed clinical choices.

Sign language recognition and assessment

In the context of SNSF Sinergia project SMILE,[8] Idiap has initiated research on sign language recognition and assessment. The consortium project coordinated by Idiap involves partners from HfH, Zurich and University of Surrey, UK. The end goal of the project is to develop a sign language assessment system that can assist Swiss German sign language learners as well as aid in standardizing a vocabulary production test that can be aligned with levels A1 and A2 of the Common European Framework of Reference for Languages (CEFR).

Sound localization and microphone array

In 2017, our research activities in the area of microphone array based speech processing and speaker diarization were mainly addressing the problem pf source localization and speech detection and also reconstruction through binary sparse coding framework. In 2017, the group has also contributed to the Perception and Activity Understanding group through EC H2020 MuMMER project, focusing on audio source localization applied in robotics.

Key publications

[1] P.-E. Honnet, B. Gerazov, A. Gjoreski, and P. N. Garner, “Intonation modelling using a muscle model and perceptually weighted matching pursuit”, *Speech Communication*, Vol. 97, 2018.

[2] D. Ram, A. Asaei, and H. Bourlard, “Sparse Subspace Modeling for Query by Example Spoken Term Detection,” to be published in *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, March 2018.

[3] A. Asaei, C. Milos, and H. Bourlard, “Perceptual Information Loss due to Impaired Speech Production”, *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, Issue: 99, August 2017.

[4]

[5] S. Tong, P. Garner, and Bourlard, “An Investigation of Deep Neural Networks for Multilingual Speech Recognition Training and Adaptation”, Proc. of Interspeech”, *Proceedings of Interspeech*, Stockholm, Sweden, 2017.

[6] S. Dey, P. Motlicek, S. Madikeri, and M. Ferras, “Exploiting Sequence Information for Text-Dependent Speaker Vefirication”, *Proc. of 2017 IEEE Intl.Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, New Orleans, pages 5370-5374, 2017.

[8] https://www.idiap.ch/scientific-research/projects/smile

2.2 Machine Learning

Overview

Head: Dr. François Fleuret (MS École Normale Supérieure and University of Paris VI, 1995; PhD, University of Paris VI, 2000; Habilitation, University of Paris XIII, 2006; EPFL MER)

Group overview: Machine learning encompasses computer techniques that aim at learning automatically representations and decision rules from data. It lies at the intersection between statistics, algorithmics, and signal processing. The main objective of the Machine Learning group is the development of novel machine-learning and statistical methods, with a strong emphasis on their algorithmic efficiency.

The application domain for the methods we develop includes image and video processing, but also industrial sensor analysis and sequence modeling.

The group was composed in 2017 of one head of group, two post-doctoral researchers, five PhD students, one intern, and it had strong interactions with the software development team. François Fleuret is also the supervisor of one phd student from the EPFL Space Engineering Center, and the co-supervisor of two phd students from the EPFL Computer Vision lab.

Key scientific outputs: In 2017, our work has resulted in contributions that improved the state-of-the-art for large-scale clustering, planetary surface reconstruction from stereo images, knowledge transfer between large neural networks, training of generative adversarial models, and modeling of sequences with recurrent neural networks.

Additional information and a list of projects are available from www.idiap.ch/ml.

Efficient Machine learning

- **Large-scale clustering:** The k-means algorithm is one of the cornerstones of machine learning for real-world applications, in particular for large-scale training. Over the recent years, several approaches have been developed based on the triangular inequality to avoid computing distances when simple bounds insure that it is not necessary.
  In the context of the MASH2 project funded by the Hasler foundation, we have first improved these exact bounds, and reduced the computational complexity of virtually all state-of-the-art methods, gaining up to a factor of two. We also have developed a novel exact algorithm which is the fastest in low dimension. In parallel, we have investigated the "batch" approaches, and proposed a new algorithm combining the triangular inequality and batches of increasing size. It resulted in a very powerful scheme that reuses computation already done over samples until statistical accuracy requires the use of additional data points.
- **Sub-linear hard sample extraction:** In the SNSF project DASH, we were interested in speeding-up "hard sample" extraction. Most of the state-of-the-art detection algorithms are trained in a sequential manner, improving iteratively the detector by training it with samples miss-classified by its previous version.
  In the ongoing ISUL SNSF project, we are investigating a novel approach that bypasses the need for prior knowledge, and instead relies on an efficient approximation of a sample's importance, combined with a sound estimate of the actual speed gain to expect. This allows us to activate our procedure only when the trade-off between speed and accuracy is beneficial.
- **High-dimension similarity measures:** The SNF WILDTRACK project was a collaboration between Idiap, EPFL and ETHZ around the design of a robust multi-camera pedestrian tracking system.
  One core component of such a system is a re-identification algorithm to automatically associate

a person leaving and later re-entering the camera field of view. This boils down to learning a metric, so that a small distance between two images is a reliable indicator that the person is the same on both, and large values is a reliable indicator these are different persons.
The technique we have developed casts this learning as a rank optimization problem with a regularizer that enforces the linear projection to remain well-conditioned. Our method allows very large scale learning and beats all existing state-of-the-art methods on standard benchmark data-sets.

Figure 7: Person detection obtained with a deep neural network, and the corresponding occupancy map estimates.

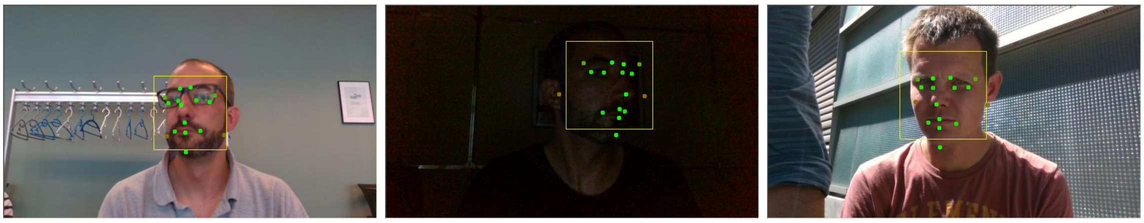

Figure 8: Face alignment examples with qualitative visibility estimation. The color from green to red indicates the visible confidence from high to low.

Figure 9: We have developed a novel strategy to find informative training examples in an image, that concentrates the computation on difficult images and parts of images to train an object detector (left), and on difficult and unusual examples to train an image classifier (right).

Deep learning

- **Depth estimation for planetary surface reconstruction:** In the framework of the NCCR PlanetS, and in collaboration with the EPFL eSpace center, we are developing a new algorithm to reconstruct the surface of Mars, given images provided by the ExoMars Trace Gas Orbiter.
State-of-the-art methods use machine-learning to automatically match corresponding pairs of images, to estimate the depth of the surface that has been imaged. We improve upon existing technique by relaxing the need for a supervised training set, and using instead crude prior knowledge about the general topography of the planet. This will allow in the future to leverage very large unlabeled training sets.

- **Deep learning for multi-camera detection:** We have developed in the MEMUDE project funded by the Hasler Foundation a new approach to adapt a monocular deep-learning detector to a multi-camera context. We first fine-tune it to the problem of person detection, and then fine-tune a Siamese network with one such monocular structure per view on a small multi-view data-set. This staging from a very large generic data-base to a small specific multi-view person data-set allows to beat existing state-of-the art multi-view methods.
- **Generative Adversarial Networks:** A very promising family of methods are the so called “Generative Adversarial Networks”, which rely on training jointly two models, one synthesizing realistic signals (images, sound, text) and another trying to discriminate the synthetic signal from the genuine one. Such techniques have demonstrated striking performance in many application domains, but involve a complex and unstable optimization problem.
  We have developed a new method that consists of training several such pairs in parallel, and maintaining carefully their statistical independence. This insures that their joint behavior has a good “covering” property, and we show experimentally that the resulting synthesis is less likely to miss sub-families of samples.
- **Manufacturing quality control:** High-end manufacturing requires accurate quality control, which results in a difficult computer vision problem: while “good” examples are plentiful and quite predictable, the failures are rare and diverse.
  We have developed a model able to learn from the correct examples a efficient representation, specific to their appearance. The poor performance of this representation when it encodes a defective example provide an efficient detection criterion.
- **Multi-modal accurate face localization for identification:** A standard pre-processing used to improve face recognition consists of localizing accurately face landmarks in order to register feature extraction with respect to them. In recent works, we have extended a state-of-art procedure to a multi-modal context where both a RGB image, near-infrared image, and a depth-map are available. Doing so improves performance drastically, in particular when dealing with poor lighting conditions.
- **Low-computation Drone Detection**
  In a collaboration with ArmaSuisse, we have investigated the development of efficient neural networks for drone detection in video feeds captured by in-flight drones. The main objective of this project was to adapt object detectors to the limited computation capabilities of embedded hardware, usually one order of magnitude less powerful than standard hardware for machine learning. Our approach consist in penalizing during training the use of large number of parameters, pushing the process toward the removal of certain sub-structures of the model identified as less useful in the inference.

Key publications

[1] J. Newling and F. Fleuret. K-Medoids For K-Means Seeding. In Proceedings of the international conference on Neural Information Processing Systems (NIPS), 2017.

[2] S. Tulyakov, A. Ivanov, and F. Fleuret. Weakly Supervised Learning of Deep Metrics for Stereo Reconstruction. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), pages 1348–1357, 2017.

[3] P. Baqué, F. Fleuret, and P. Fua. Deep Occlusion Reasoning for Multi-Camera Multi-Target Detection. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), pages 271–279, 2017.

[4] T. Chavdarova and F. Fleuret. Deep Multi-Camera People Detection. In Proceedings of the IEEE International Conference on Machine Learning and Applications (ICMLA), pages 848–853, 2017.

[5] J. Newling and F. Fleuret. A Sub-Quadratic Exact Medoid Algorithm. In Proceedings of the

international conference on Artificial Intelligence and Statistics (AISTATS), pages 185–193, 2017. (Best paper award).

2.3 Social Computing

Overview

Head: Prof. Daniel Gatica-Perez (PhD, University of Washington, USA, 2001; EPFL Adjunct Professor)

Social computing is an interdisciplinary domain that integrates theory and models from ubiquitous computing, social media, machine learning, and social sciences to analyze human and social behavior in everyday life, and to create devices and systems that support social interaction.

The Social Computing group in 2017 was composed of one group head, four postdoctoral researchers, three PhD students, one intern, and three EPFL master students. Research lines investigated in 2017 included: social media analytics and mobile crowdsourcing for cities and health, ubiquitous conversational interaction analysis, and analysis of Maya hieroglyphic media.

Key scientific outputs: Publications on (1) social media analytics and mobile and online crowdsourcing to understand youth nightlife patterns, urban perception in cities, and eating patterns in everyday life; (2) multimodal analysis of soft skills in face-to-face employment interviews and hospitality service encounters; and (3) Maya hieroglyphic visual analysis. 13 EPFL PhD students have graduated from the group since 2002.

Additional information and a list of projects are available from www.idiap.ch/socialcomputing.

Social media analytics and mobile crowdsourcing for cities and health

Our work in this domain spans several research lines. First, as part of the SNSF Youth@Night and Dusk2Dawn projects[9] (Multidisciplinary Study of Young People's Going Out and Drinking Behaviors, in collaboration with Addiction Switzerland and the University of Zurich), we investigated the use of mobile crowdsourcing and social media analytics to characterize urban phenomena, both in Switzerland and other world regions.[10] This included the automatic recognition of place ambiance from social media images using deep learning [1] (Figure 10), and automatic recognition of drinking activity from smartphone sensor data, for a population of 200 young volunteers in Switzerland.

Regarding mobile crowdsourcing and health, in the context of mobile the Bites-n-Bits project [11] (Understanding Eating Routines in Context, in collaboration with Nestle Research Center), we developed a smartphone sensing app to collect everyday life data about eating and drinking for a population of 120 college students, and used it to build an approach for recognition of eating occasions [2]. Furthermore, using visual social media (Instagram) as data source, we analyzed eating and drinking patterns in the context of the Social Food and People project[12]. (supported by EPFL Integrative Food Science and Nutrition Center.)

Regarding mobile crowdsourcing for social innovation, we continued our work using the Civique platform[13] (Mobile Data for Local Causes), which included its use for teaching students about humanitarian technologies [14], and to collect multimedia data from Swiss heritage speakers living in the USA [15] (in a collaboration with the University of Lausanne, supported by the UNIL-EPFL Program on Collaborative Research on Science and Society). Finally, our work in Latin America on mobile crowdsourcing

[9] https://www.youth-night.ch/
[10] https://www.youtube.com/watch?v=71ht15VAoLw
[11] https://www.bitesnbits.org/
[12] http://www.idiap.ch/en/allnews/what-does-instagram-reveal-about-food-consumption-in-switzerland
[13] https://www.civique.org
[14] http://www.idiap.ch/en/allnews/understanding-how-technology-can-revolutionize-humanitarian-work
[15] http://wp.unil.ch/sina/

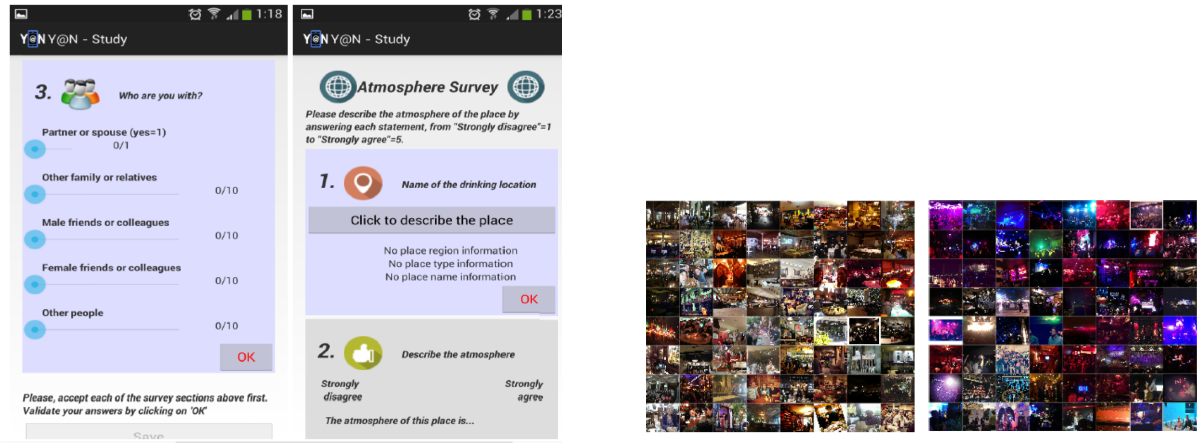

Figure 10: Left: Youth@Night survey logger app. Right: Social media images classified as Restaurant and Stage by convolutional neural network.

for youth engagement on local urban issues was expanded through a collaboration with institutions in Mexico and Colombia [3]. Other collaborations using the Civique platform are under development.

Ubiquitous interaction analytics

In the context of the SNSF UBImpressed project[16] (Ubiquitous Frist Impressions and Ubiquitous Awareness), we have developed methods to analyze dyadic interactions in the workplace using multiple sensors (cameras, Kinect devices, microphone arrays, smartphones, and wearables), and to infer variables like perceived hirability and performance (Figure 11). This is joint work with the University of Lausanne and Cornell University, in partnership with Vatel Switzerland Hotel Management School. Specifically, we found connections between automatically measured nonverbal behavior (including speaking activity, prosody, head activity, and body motion) and impressions in job interviews and hospitality service encounters [4]. Some of this work will be transfered to industry in the context of a new CTI project.

Figure 11: Sensors available in the social sensing lab. Kinect V2, Microcone, Q-Sensor, Android smartphone, Google Glass.

[16] http://www.idiap.ch/project/ubimpressed

Visual analysis of Maya hieroglyphs

In the context of the SNSF MAAYA project[17] (Multimedia Analysis and Access for Documentation and Decipherment of Maya Epigraphy), we developed methods for visual analysis of hieroglyphs in ancient codices [18]. Specifically, we designed a methodology that included crowdsourced glyph localization and segmentation by non-experts for data labeling, and deep learning for visual recognition [5] (Figure 12). This included visualization of the output of deep networks for interpretation of the learned visual representations. Our work was covered by Horizons Magazine, the scientific magazine of the SNSF and the Swiss Academies of Arts and Sciences [19].

Figure 12: Left: Convex hull of the segmentation ground truth for the target glyph (red line, blue filling). Center: Gray-scale image of aggregated segmentations generated by online crowdworkers. Right: Final aggregated segmentation.

Key publications

[1] Y. Benkhedda, D. Santani, and D. Gatica-Perez, “Venues in Social Media: Examining Ambiance Perception Through Scene Semantics,” in *Proc. ACM Int. Conf. on Multimedia (MM)*, Mountain View, Oct. 2017.

[2] J.-I. Biel, N. Martin, D. Labbe, and D. Gatica-Perez, “Bites’n’Bits: Eating Behavior in Context from Mobile Data,” *PACM on Interactive, Mobile, Wearable, and Ubiquitous Technologies (IMWUT)*, Vol. 1, No. 4, Dec. 2017.

[3] S. Ruiz-Correa, D. Santani, B. Ramirez Salazar, I. Ruiz Correa, F. Alba Rendon-Huerta, C. Olmos Carrillo, B. C. Sandoval Mexicano, A. H. Arcos Garcia, R. Hasimoto Beltran, and D. Gatica-Perez, “SenseCityVity: Mobile Sensing, Urban Awareness, and Collective Action in Mexico,” *IEEE Pervasive Computing, Special Issue on Smart Cities*, pp. 44-53, Apr.-Jun. 2017.

[4] S. Muralidhar, M. Schmid Mast, and D. Gatica-Perez, “How May I Help You? Behavior and Impressions in Hospitality Service Encounters,” in *Proc. ACM Int. Conf. on Multimodal Interaction (ICMI)*, Glasgow, Nov. 2017.

[5] G. Can, J.-M. Odobez, and D. Gatica-Perez, “Codical Maya Glyph Segmentation: A Crowdsourcing Approach,” *IEEE Trans. on Multimedia*, published online Sep. 2017, Vol. 20, No. 3, pp. pp. 711-725, Mar. 2018.

[17] http://www.idiap.ch/project/maaya/
[18] http://lab.idiap.ch/maaya/
[19] http://www.idiap.ch/en/allnews/maya-glyph-analysis-on-horizons-magazine

2.4 Perception and Activity Understanding

Overview

Head: Dr. Jean-Marc Odobez (Engineer degree, ENST Bretagne, 1990; Ms in Signal Processing, Rennes University, 1990; PhD, University of Rennes, France, 1994; EPFL MER)

Group overview: The Perception and Activity Understanding group conducts research in human activities analysis from multi-modal data. This entails the investigation of fundamental tasks like the detection and tracking of people, the estimation of their pose or the detection of non-verbal behaviors, and the temporal interpretation of this information in forms of gestures, activities, behavior or social relationships. These tasks are addressed through the design of principled algorithms extending models from computer vision, multimodal signal processing, and machine learning, in particular probabilistic graphical models and deep learning techniques. Surveillance, traffic and human behavior analysis, human-robot interactions, and multimedia content analysis are the main application domains.

In 2017, the group was composed of one group head, four post-doctoral researchers, six PhD students.

Key scientific outputs: The group is well known for its work on video sequence analysis, probabilistic tracking, non-verbal behavior extraction (in particular attention modeling), and temporal motif discovery. In 2007, the Klewel company was created using its OCR technology. In 2015 and 2016, the PAU team ranked first at the MediaEval Person discovery challenge. In 2016, its work on 3D face and gaze tracking from cheap RGB-Depth sensors has been patented and has led to the creation of the Eyeware start-up company. Besides the consolidation of his ground work, the group has investigated deep learning methods for several tasks like gesture recognition, audio-visual speaking activity modeling, gaze, joint audio localization and speech/non-speech detection, body landmark detection using depth data, and multimedia processing (cross-modal transfer learning, shape recognition, text localization). It is also working on the integration of its sensing technology and algorithms into real-time perceptual systems for human-robot interaction (Pepper platform, EU MuMMER project), or to companies (e.g. anti-tailgating detection system). During the period 2013-2017, the group published 17 journal papers and more than 40 conference papers.

Additional information and a list of projects are available from www.idiap.ch/perception.

Deep learning

Gesture Recognition. In [5] we studied multimodal Deep Dynamic Neural Networks (DDNN) for the segmentation and recognition of short spontaneous communicative gestures, as illustrated in Fig. 13a. We adopted a semi-supervised hierarchical approach based on statistical temporal models relying on high-level spatio-temporal gesture representations trained using deep neural networks suited to the input modalities (dynamics of skeletal joint information (Fig. 13b), fusion of batches of depth and RGB images (Fig. 13c).

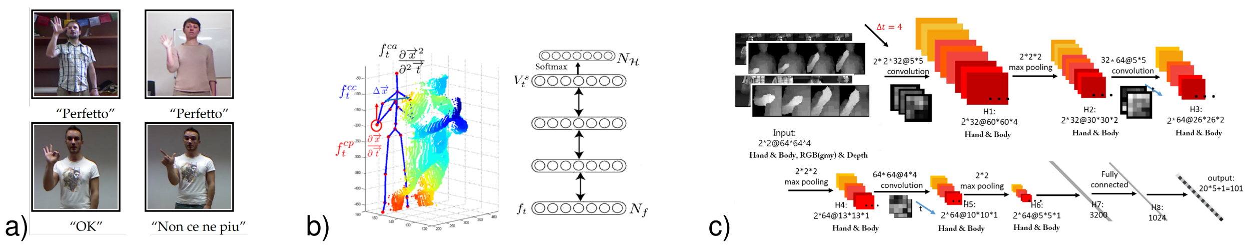

Figure 13: Multimodal gesture recognition from RGB-D data, using deep-learning methods. a) Spontaneous communicative gestures to be recognized. b) Encoding of the skeleton information with Restricted Bolzman Machines. c) CNN models of visual and depth hand data.

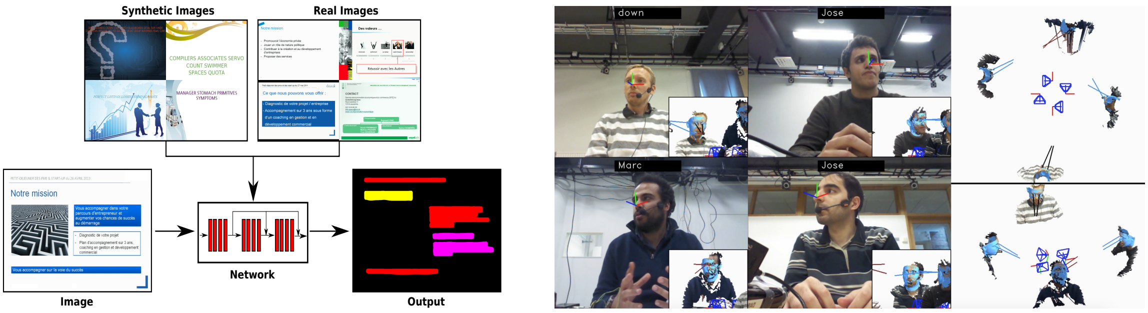

Figure 14: Left: Semantic Text Segmentation Network (STSN) trained with synthetic and real data, that apart from localizing text in images with high accuracy also classifies text regions in different semantic categories (title, bullets, standard text), and then further recognizes text (OCR). Right: Heads are tracked, eye localized and aligned before inferring gaze and attention towards people or other semantic labels (all information analysed in 3D).

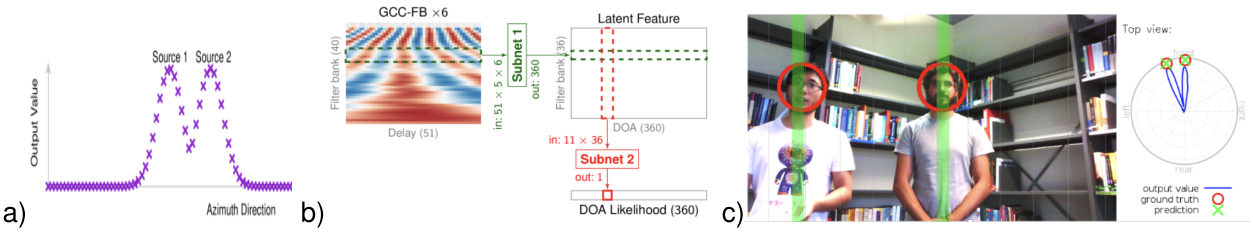

Figure 15: Multiple sound source localization for HRI. a) Encoding the network outpout as likelihood. b) Neural network architecture, mapping pairwise generalized-cross-correlation features to the direction of arrival. c) Sample result on real HRI data from the pepper robot.

Semantic text recognition (OCR). In the context of the CTI funded VIEW project, we investigated deep learning methods for the detection, segmentation, categorization and recognition of text content in presentations, allowing further semantic tagging (see Fig. 14). Models and systems are being transfered to the Klewel company, improving the internet referencing and visibiliy of the recorded and broadcast presentations of its client.

Audio analysis. In the context of the EU MuMMER project on social robotics, we have investigated the use of different DNN architectures for the localization of multiple speech sources, a situation that can often arise when the robot has to deal with multiple people. By encoding the potential sound source locations using a likelihood approach (Fig. 15a), exploiting a hierarchical approach (per band prediction, then fusion, Fig. 15b), the network can learn to simultaneously localize multiple sound sources (Fig. 15c). Methods were demonstrated to perform much better than state-of-the-art signal-processing based approaches (GCC or SRP-PHAT and other models or variants).

We also investigated deep-temporal representations for the modeling of audio-visual speaking activities, e.g. for differentiate dubbing from genuine talking situations for instance.

Shape representation of Maya glyphs. In the context of the SNSF MAYA project and in collaboration with the Social Computing group we investigated crowdsourcing, vizualisation and deep learning methods for the data labeling, shape representation, and glyph analysis in ancient Codices (see Fig. 12).

Multiple object tracking

Our previous works resulted in an enhanced Conditional Random Field tracking-by-detection framework with important characteristics making it successful for person or face tracking in challenging conditions: long-term pairwise links, pairwise cue measurement costs sensitive to the time interval between two detections, integration of multi-cue association reliability factors, and unsupervised learning.

Figure 16: Robust head pose tracking. a) Example of automatically reconstructed heads. b) Tracking results in adverse conditions (UBImpressed registration desk set-up). c) Tracking results in an HRI situation (image taken from the Pepper robot).

Gaze, non-nerbal behavior extraction, human-robot interactions

Our long-standing work on NVB perception has been extended in several ways. In the context of the Swiss UBImpressed project, In [1] we have worked on robust and accurate head pose tracking from RGB-D data. Combining the benefits of the online fitting of a 3D face morphable model with the online 3D reconstruction of the full head (Fig. 16a), the approach provides more support when handling extreme head poses, like for the analysis of difficult registration desk sequences (SNSF Sinergia UBImpressed project, Fig. 16b), or multiple interacting people in HRI (MuMMER project, Fig. 16c).

Our long-term research on gaze estimation from cheap RGB-D (depth) cameras like Kinect [4] has been pursued. We investigated online techniques leveraging social situations to compute parameters of a gaze bias correction model. This improved gaze coding from 56% accuracy to 85% on several dozen hours of data. Figure 14 illustrates the result of applying our tools for gaze coding in multiparty situations.

We continue improving our real-time multi-person perception system (Pepper robot in the EU MuMMER project), with features including tracking, pose estimation, re-identification, extraction of non-verbal cues (attention, head gestures), as well as speech sound localization.

Multimodal face and person diarization and naming

Identifying people appearing and speaking in multimedia data as we have done in the EUMSSI EU project allows to monitor who said what and when and is thus important for the design of search and fast browsing tools of or broadcast programs. On addition to face tracking, clustering, audio-visual association, and dubbing detection, we investigated this year domain cross-modal domain adaptation and transfer learning methods for improving the estimation of short utterance speaker embedding leveraging face embeddings [3].

Key publications

[1] Y. Yu, K. Funes and J.-M Odobez. HeadFusion: 360^{o}Head Pose tracking combining 3D Morphable Model and 3D Reconstruction *IEEE Trans. Pattern Anal. Mach. Intell.*, accepted for publication, 2017.

[2] G. Can, J.-M. Odobez, and D. Gatica-Perez. Codical Maya Glyph Segmentation: A Crowdsourcing Approach, *IEEE Trans. on Multimedia*, published online Sep. 2017, Vol. 20, No. 3, pp. pp. 711-725, Mar. 2018.

[3] N. Le and J.-M. Odobez. A Domain Adaptation Approach to Improve Speaker Turn Embedding Using Face Representation, In *Int. Conf. on Multimodal Interactions (ICMI)*, 2017.

[4] K. Funes and J.-M Odobez. Gaze estimation in the 3d space using rgb-d sensors: Towards head-pose and user invariance. *Int. Journal of Computer Vision*, 118(2):194–216, june 2016.

[5] D. Wu, L. Pigou, P.-J. Kindermans, N. Le, L. Shao, J. Dambre, and J.-M. Odobez. Deep dynamic neural networks for multimodal gesture segmentation and recognition, *IEEE Trans. Pattern Anal. Mach. Intell.*, 38(8):1583–1597, 2016.

2.5 Robot Learning and Interaction

Overview

Head: Dr. Sylvain Calinon (MS and PhD, EPFL, 2003 and 2007)

Group overview: The Robot Learning and Interaction group, created in 2014, focuses on human-centric robotic applications in which the robots can learn new skills by interacting with the end-users. From a machine learning perspective, the challenge is to acquire skills from only few demonstrations and interactions, with strong generalization demands. It requires the development of intuitive active learning interfaces to acquire meaningful demonstrations, the development of models that can exploit the structure and geometry of the acquired data in an efficient way, and the development of adaptive control techniques that can exploit the learned task variations and coordination patterns.

The Robot Learning and Interaction group in 2017 was composed of one postdoctoral fellow, four PhD students, three MSc students, five BSc students and one trainee.

Key scientific outputs: Development of robot learning and adaptive control algorithms that can be applied to a wide range of applications, for robots that are either close to us (assistive robots in I-DRESS), parts of us (prosthetic hands in TACT-HAND), or far away from us (manipulation skills in deep water in DexROV). Attentive to reproducible research, the group regularly releases open source codes accompanying its publications at www.idiap.ch/software/pbdlib/.

Additional information and a list of projects are available from www.idiap.ch/rli.

Probabilistic models of movements and skills

*Figure 17: Human-centric robotic applications. **Left:** Teleoperation assistance for the control of a bimanual robot (DexROV project), where the robot assists the teleoperator when repetitive or structured tasks are detected, as a form of human-robot collaboration. For example, if the robot observes that the task of drilling requires the drill to be perpendicular when it approaches a surface, the robot will then automatically orient the drill when it approaches another surface, letting the teleoperator concentrate on the position to drill while letting the robot maintain the correct orientation. **Center:** Personalized assistance in dressing (I-DRESS project), where the robot learns by demonstration the preferences and type of assistance required by each user. This is achieved by extending the movement primitives frequently employed in robotics to a wider repertoire of skills composed of reactive behaviors based on force, impedance, position and orientation. **Right:** Online learning and adaptive control of a prosthetic hand (TACT-HAND project), where electromyography and tactile sensing data are combined to control a prosthetic hand.*

In many robotics applications, demonstrations or experiences are sparse. In such situation, it is important to get as much information as possible from each demonstration. We explore approaches encoding demonstrations from the perspective of multiple coordinate systems. This is achieved by providing a list of observers that could potentially be relevant for the movement or the task to transfer. A statistical learning approach is then used to determine the variability and coordination patterns in the movement by considering different coordinate systems simultaneously. This approach allows the orchestration of multiple coordinate systems to reproduce movements in new situations (typically, to adapt a movement to new positions of objects).

The proposed *task-parameterized model* exploits the structure of the task, which can in many robotics problems be expressed in the form of coordinate systems or local projections. It was shown that such approach provides better generalization capability than conventional regression.

A distinctive view of optimal control

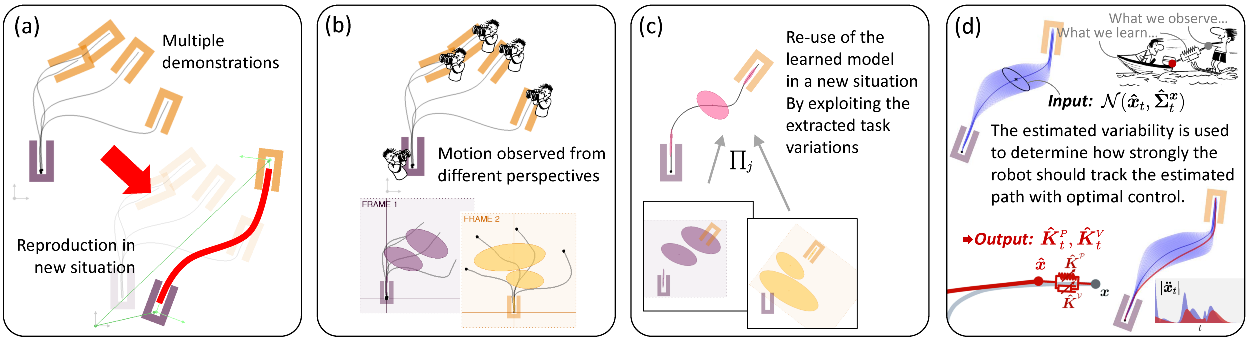

Figure 18: Task-parameterized model to synthesize movements in new situations. (a) *Observation in different situations to extract the underlying structure of the task.* (b) *Probabilistic encoding of the task in multiple coordinate systems.* (c) *The cross-situational observations are used to adapt the motion to new situations.* (d) *Model predictive control strategy to reproduce the movement by exploiting the retrieved variability and correlation information.*

Model predictive control (MPC) is ubiquitous in robot control, but the core formulation of this control problem and its associated algorithms can be extended to a wider range of problems, which has often been overlooked in robot learning. In particular, the most simple form of MPC (unconstrained and linear, with a homogeneous double integrator system) already has advantage for motion synthesis and planning problems, where it can be combined elegantly with probabilistic representations of movements.

This method allows the retrieval of smooth and natural trajectories analytically, by taking into account variation and coordination constraints. Instead of learning trajectories directly, the approach allows the learning of the underlying controllers to drive the robot. Namely, it learns to reject perturbations only in the directions that would affect task performance (minimal intervention control). This can typically be exploited with torque-controlled robots to regulate the tracking gain and compliance required to reproduce a task in an adaptive manner.

Geometry-aware statistical learning and control

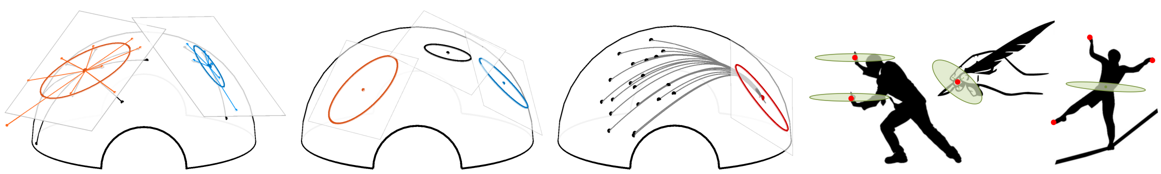

Figure 19: Statistics and control on Riemannian manifold. From left to right: Encoding as Gaussian mixture model, fusion of information with product of Gaussians, linear quadratic tracking, and example of application consisting of learning and tracking manipulability ellipsoids (symmetric positive definite matrices), which are used to find poses that a robot should adopt to face external perturbations.

The data encountered in robotics are characterized by simple but varied geometries, which are often underexploited when developing learning and control algorithms. Such data range from joint angles in revolving articulations, rigid body motions, orientations represented as unit quaternions, sensory data processed as spatial covariance features, or other forms of symmetric positive definite matrices such as inertia or manipulability ellipsoids. Moreover, many applications require these data to be handled altogether.

We exploit Riemannian manifold techniques to extend algorithms initially developed for Euclidean data, by efficiently taking into account prior knowledge about these manifolds and by modeling joint distributions among these heterogeneous data. The use of these differential geometry techniques allow us to treat data of various forms in a unified manner (including data in standard Euclidean spaces). It can typically be used to revisit common optimization problems in robotics formulated in standard Euclidean spaces, by treating them as unconstrained problems inherently taking into account the geometry of the data.

Key publications

[1] Pignat, E. and Calinon, S. (2017). Learning adaptive dressing assistance from human demonstration. *Robotics and Autonomous Systems*, 93, 61–75.

[2] Jaquier, N., Connan, M., Castellini, C. and Calinon, S. (2017). Combining Electromyography and Tactile Myography to Improve Hand and Wrist Activity Detection in Prostheses. *Technologies*, 5:4, Special Issue on Assistive Robotics.

[3] Zeestraten, M.J.A., Havoutis, I., Silvério, J., Calinon, S. and Caldwell, D.G. (2017). An Approach for Imitation Learning on Riemannian Manifolds. *IEEE Robotics and Automation Letters (RA-L)*, 2:3, 1240–1247.

[4] Calinon, S. (2016). A Tutorial on Task-Parameterized Movement Learning and Retrieval. *Intelligent Service Robotics*, 9:1, 1–29.

[5] Berio, D., Calinon, S. and Leymarie, F.F. (2017). Generating Calligraphic Trajectories with Model Predictive Control. In Proc. of the 43rd Conf. on *Graphics Interface*, pp. 132–139.

2.6 Uncertainty Quantification and Optimal Design

Overview

Head: Prof. David Ginsbourger (Ph.D. Mines Saint-Etienne 2009, Habilitation Universität Bern 2014, Titularprofessor Universität Bern 2018)

Group overview: The Uncertainty Quantification and Optimal Design group focuses on quantifying and reducing uncertainties in the context of natural and artificial complex systems. Application domains notably include energy and geosciences, with a number of collaborations ranging from safety engineering to hydrology and climate sciences. In all these fields the study of complex systems often relies on expensive data acquisition and model runs, calling for adapted experimental design strategies.

UQOD started at Idiap in September 2015, with members coming from and keeping strong academic ties to the Institute of Mathematical Statistics and Actuarial Science (IMSV) of the University of Bern (UniBE). During the year 2017, the UQOD group has been composed of a permanent senior researcher, an Idiap postdoctoral researcher (now at IDSIA), an intern, and two UniBE students occasionally visiting the group (one master student and a PhD student who was in “co-tutelle” with Marseille).

Key scientific outputs: Current contributions include efficient algorithms for Bayesian set estimation, notably for estimating and quantifying uncertainties on overcritical parameter regions with Gaussian Process (GP) models. In 2017, the project “Learning and visualizing dangerous regions in multivariate parameter spaces” funded by the Hasler foundation has been conducted within the UQOD group. Other recent results deal with the interplay between covariance kernels and properties of associated GPs, with implications in function prediction under structural constraints and beyond. Ongoing work also encompasses novel algorithms for non-stationary modelling of extremes with application in climate sciences, contributions to statistical modelling of hail, as well as collaborations with geoscientists in uncertainty quantification and simulation-based optimization.

Additional information and a list of projects are available from www.idiap.ch/uqod.

Bayesian optimization and emulation with Gaussian Process models

Bayesian global optimization relying on GPs has become a standard for optimizing prohibitively expensive to evaluate systems, e.g. with response(s) of interest stemming from heavy numerical simulations. It constitutes one of the core domains of expertise of the UQOD group, with recent contributions ranging from theoretical to methodological questions (such as parallelization, handling large data sets, coping with high-dimensional inputs, etc.) and applications. Notably, the group has been involved in a collaboration with researchers in hydrogeology from the universities of Lausanne and Neuchâtel, with the aim to investigate Bayesian optimization for contaminant source localization relying on flow simulations. The preprint “Contaminant source localization via Bayesian global optimization” (now available online as a Hydrology and Earth System Sciences discussion paper) illustrates some first optimization results based on a fixed geology: Bayesian optimization automatically selects the ground truth contamination source much more efficiently than brute force grid search. This ongoing work has also lead to generate objective functions that are publicly available for algorithm benchmarking.

The UQOD group also investigates GP emulation per se, with a particular focus on the incorporation of expert knowledge and the identification of structural properties of objective functions through the specification of covariance kernels and the estimation of their parameters.

Recent work includes non-stationary GP modelling using warping approaches and also connections to multiresolution analysis and wavelet modelling, see for instance [3]. Ongoing work (presented at the ISI world statistic congress 2017) also concerns the estimation of weights by maximum likelihood in multiple kernel Gaussian Process modelling.

Computer experiments for the quantification and the reduction of uncertainties

Besides global optimization, UQOD has also been focusing on strategies dedicated to other goals such as locating parameter regions leading to a response exceeding a given threshold, corresponding e.g. to an abnormal behaviour of the considered system.

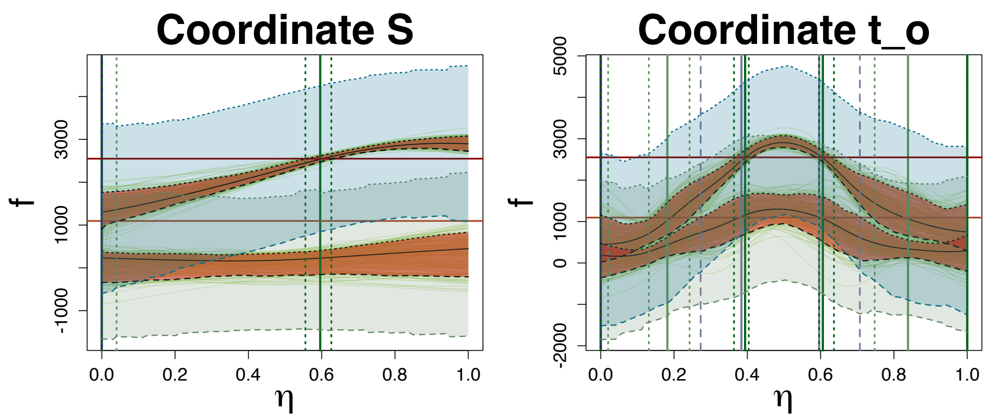

Figure 20: Estimating profile optima on a coastal flooding BRGM application

In the project “Learning and visualizing dangerous regions in multivariate parameter spaces”, novel approaches based on profile extrema have been developed for visualizing and quantifying uncertainties on excursion regions with application to a coastal flooding test case provided by colleagues from BRGM (Bureau de Recherches Géologiques et Minières, Orléans, France).

Figure 20 represents the kind of output delivered by the *developd* approach: for each individual parameter of interest, estimates of maximum/minimum response with respect to all remaining parameters are estimated (with confidence statements) in function the individual parameter in question.

More generally, the combination of GP modelling and stochastic simulation techniques have been investigated within UQOD for quantifying and reducing uncertainties on sets, with example applications in safety engineering, where the typical goal is to identify the set of dangerous and/or safe configurations of a complex system and if possible to provide some measures of confidence along with the estimate(s).

Motivated by the estimation of such sets, a research work has been conducted that lead to a class of algorithms called “asymmetric nested Monte Carlo”, that prove useful for computing conservative estimates of excursion sets and more generally for efficiently estimating orthant probabilities of high-dimensional Gaussian vectors [2].

These results have been used to derive conservative set estimates on a neutronic criticality safety test case (from IRSN, the French Institut de Radioprotection et de Sûreté Nucléaire) starting from a initial experiments, and also to derive sequential design strategies dedicated to this class of conservative set estimation problem, that were applied to this IRSN test case and also to another BRGM test case. This ongoing work, in collaboration with colleagues from Neuchâtel, CentraleSupélec and IRSN, is summarized in the preprint “Adaptive Design of Experiments for Conservative Estimation of Excursion Sets” now available online (arXiv:1611.07256).

UQOD has also been active in a collaboration with colleagues from CentraleSupélec and Toulouse University towards establishing convergence results (See arXiv:1608.01118) for some important classes of Sequential Uncertainty Reduction strategies, with links to Bayesian optimization.

Statistical analysis of non-stationary time series arising in climate sciences

Extreme value analysis is a key approach to a number of phenomena from nature, economic and industry activities. The theory of extreme value distributions has been studied for decades, and adequate statistical distributions are known for dealing with maxima and quantify how severe and with what probability extreme events may occur. However, many methods in use require a sample of independent identically distributed maxima in order for the underlying mathematical theory to be valid. This poses serious challenges to climate scientists and beyond, as phenomena exhibit dependencies and also change over time. In approaches that have been recently developed within UQOD and in collaboration

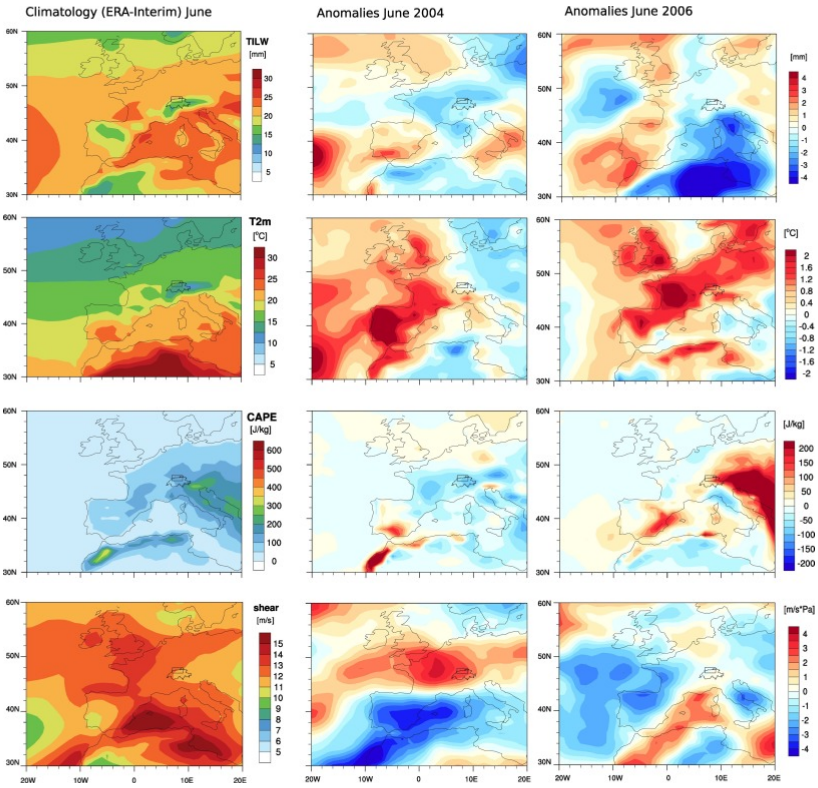

Figure 21: Environmental variables and their anomalies, used for hail prediction

with the Oeschger Center for Climate Change Research, e.g. within the Internship of Alan Maître, the main focus has been on extreme value analysis of non-stationary climatological times series.

Key publications

[1] A Poisson regression approach to model monthly hail occurrence in Northern Switzerland using large-scale environmental variables, Erica Madonna, David Ginsbourger and Olivia Martius, in: *Atmospheric Research*, 203:261-274, 2018.

[2] Estimating orthant probabilities of high dimensional Gaussian vectors with an application to set estimation, Dario Azzimonti and David Ginsbourger, in: *Journal of Computational and Graphical Statistics*, Accepted author version posted online: 03 Aug 2017.

[3] Non-parametric warping via local scale estimation for non-stationary Gaussian process modelling, Sébastien Marmin, Jean Baccou, Jacques Liandrat and David Ginsbourger, in: *Wavelets and Sparsity XVII*, International Society for Optics and Photonics, 2017.

[4] On uncertainty quantification in hydrogeology and hydrogeophysics, Niklas Linde, David Ginsbourger, James Iriving, Fabio Nobile and Arnaud Doucet, in: *Advances in Water Resources*, 110:166–181, 2017.

2.7 Computational Bioimaging

Overview

Head: Dr. Michael Liebling (MS, EPFL, 2000; PhD, EPFL 2004; postdoc, Caltech, 2004–2007; Assistant Prof (2007–2013), Associate Prof (2013–2017), Adjunct Prof (2017–) University of California Santa Barbara (UCSB)

Group overview: Research in the Computational Bioimaging Group focuses on developing image acquisition, reconstruction and analysis algorithms to study live biological systems. Practical tools aim at (i) extending the physical limits of imaging hardware via techniques including super-resolution and multi-view, space variant deconvolution, and (ii) quantitative analysis of complex biological systems: motion-based image analysis, cell tracking, microscopic fluid flow estimation, and integration of multi-modality images.

In 2017, the Computational Bioimaging Group was composed of the head of group and three PhD students.

Key scientific outputs: Recent milestones include the reconstruction of 3D volumes of the beating embryonic heart at frame rates above 1000 volumes per second, temporal super-resolution for sensitive fluorescence cameras, and observation and quantitation of heart development in animal models.

Additional information and a list of projects are available from www.idiap.ch/cbi.

Direct inversion algorithm for focal plane scanning optical projection tomography

To achieve approximately parallel projection geometry, traditional optical projection tomography (OPT) requires the use of low numerical aperture (NA) objectives, which have a long depth-of-field at the expense of poor lateral resolution. Particularly promising methods to improve spatial resolution include ad-hoc post-processing filters that limit the effect of the systems' MTF and focal-plane-scanning OPT (FPS-OPT), an alternative acquisition procedure that allows the use of higher NA objectives by limiting the effect of their shallow depth of field yet still assumes parallel projection rays during reconstruction. We have established the existence of a direct inversion formula for FPS-OPT and, based on this formula, we have proposed a point spread function-aware algorithm that is similar in form and complexity to traditional filtered backprojection (FBP). With simulations, we have demonstrated that the point-spread-function aware FBP for FPS-OPT leads to more accurate images than both traditional OPT with deconvolution and FPS-OPT with naive FBP reconstruction. We further illustrated the technique on experimental zebrafish data, which showed that our approach reduces out-of-focus blur compared to a direct FBP reconstruction with FPS-OPT.

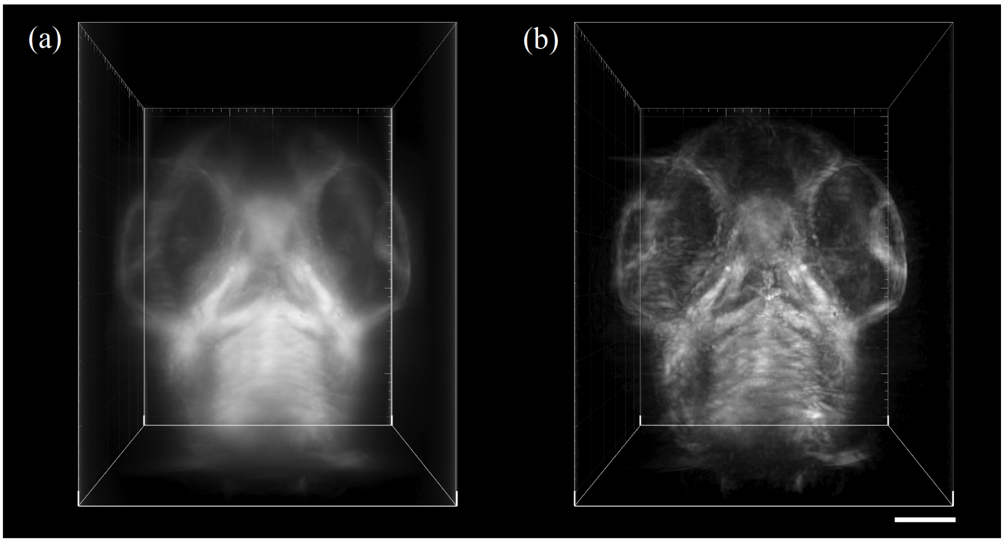

*Figure 22: We used focal plane scanning OPT (FPS-OPT) to image the head of a Tg(*fli1a*:eGFP) zebrafish in 3D fluorescence with a 10×/0.3NA air objective. Under such conditions, single-plane OPT would be unable to produce an acceptable reconstruction due to the large sample thickness and shallow depth-of-field. With FPS-OPT, we compare 3D reconstructions from (a) standard filtered backprojection (FBP) and (b) our proposed PSF-aware FBP. Our proposed PSF-aware FBP algorithm reconstructs an image with less out-of-focus blur. Scale bar is 100μm. Adapted from [1].*

Platform for Reproducible Acquisition, Processing, and Sharing of Dynamic, Multi-Modal Data

In this project, a joint effort between the Computational Bioimaging group, the Robotics and Interaction group and the Biometrics Security and Privacy group, we assembled a platform for reproducible acquisition, processing, and sharing of dynamic, multi-modal data. This modular platform consists of four layers: sensing, positioning, computing, and storage. It accepts a wide range of sensing devices and complements Idiap's already existing infrastructure for collecting data in the areas of computer vision, biometry, and speech processing. Concretely the platform consists of the following components (Figure 23):

- Two 7-axis backdrivable robotic arms (Panda, Franka Emika), computer-driven
- Pneumatic gripper system
- Two-color Open-SPIM light-sheet microscope (20 × water immersion microscope object, 488 et 561 nm laser excitation for green and red fluorescence imaging), computer-driven
- Dedicated data processing, storage and serving backend (4GPU NVidia K80 (19968 GPU cores), 8CPU Intel (112 CPU cores), 40TB storage)

The platform is connected to Idiap's computing and storage infrastructure, with platform-dedicated computing and storage nodes that were integrated within Idiap's computing facility. During setup and calibration, we have used the platform to acquire preliminary evaluation data. In particular, we adjusted the microscope's 3D sample positioning stage and acquired image series of fixed samples (Figs. 23). We foresee that the platform will enable new research projects both within Idiap and with outside research groups in Valais, Switzerland, and beyond.

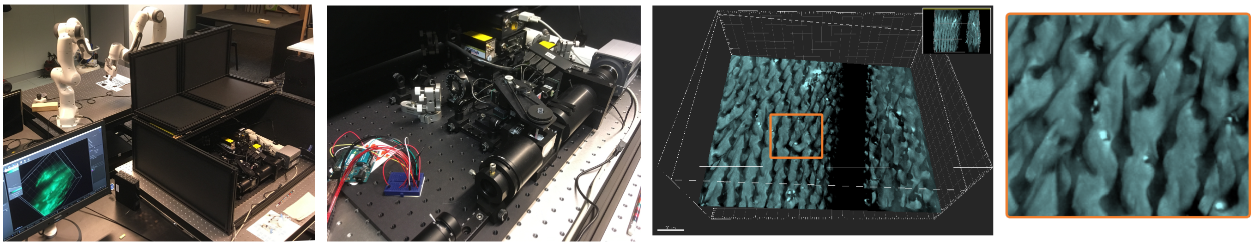

Figure 23: Photograph of the reproducible acquisition platform. Two-color Open-SPIM microscope in foreground, with robotic arms (Panda, Franka Emika, GmbH, Munich, Germany) in background; in addition to the data acquisition platform visible here, the platform relies on dedicated data processing, storage and serving backend. Right two frames, Demonstration data (Wing of a common housefly) acquired with microscopy platform. Width of detail rectangle is 100 μm.

Key publications

[1] Kevin G. Chan and Michael Liebling, "Direct inversion algorithm for focal plane scanning optical projection tomography," *Biomed. Optics Express*, Vol. 8, Issue 11, pp. 5349-5358 (2017).

2.8 Biometrics Security and Privacy

Overview

Head: Dr. Sébastien Marcel (PhD, University of Rennes, France, 2000; Visiting Professor, University of Cagliari, 2010; Lecturer, EPFL, 2013-)

Group overview:

Biometrics refers to the automatic recognition of individuals based on their physiological and/or behavioral characteristics. The Biometrics Security and Privacy (BSP) group at Idiap focuses on three main areas of research:

- Biometric recognition: We investigate and develop new biometrics based recognition algorithms, notably for face, voice, and vein biometric modalities.
- Presentation attack detection: We look for new and better ways of detecting presentation attacks on face, voice, and vein biometric recognition systems.
- Biometric template protection: We research effective methods of preserving both the security of biometric recognition systems and the privacy of their users by protecting the biometric models ("templates") that are employed by the system for recognition purposes.

The BSP group prioritizes reproducibility in research. This is important for ensuring that our work can be both verified and built upon by the wider research community. To enable reproducibility, we mainly make use of our Python-based signal-processing and machine-learning toolbox, Bob (http://www.idiap.ch/software/bob/), which we make freely available for academic purposes. The group also develops and maintains the BEAT platform (https://www.beat-eu.org/platform/) – a MLaaS platform compliant with Swiss and European data-security norms.

The group participates in several large-scale biometrics projects at Swiss (SNSF), European (H2020) or world-wide levels (eg. IARPA/DARPA) but also conducts projects directly with companies.

The BSP group provides also expertise to the Swiss Center for Biometrics Research and Testing, which, among other things, carries out tests and evaluations on commercial products related to biometrics.

The BSP group in 2017 was composed of 1 head of group, 4 research associates, 5 postdocs, 3 PhD students, 1 intern and 1 engineer.

Key scientific outputs: The BSP group has been pioneering the work on mobile biometrics (face and speaker recognition) and on PAD in face and speaker recognition by sharing the first open databases, organising the first International competitions and producing the first reproducible research studies in the domain. Regarding face PAD in mobile scenario, the group confirmed that the current trend using discriminant classifiers is prone to over-fitting hence resulting in a lack of generalisation on unseen presentation attacks. Regarding voice PAD we demonstrated that the existing methods generalise poorly when different databases or different types of attacks are used for training and testing. These results question the efficiency and practicality of the existing PAD systems, as well as, call for creation of databases with larger variety of realistic speech presentation attacks. The BSP group also investigated approaches for heterogeneous face recognition and vein recognition.

Additional information and a list of projects are available from www.idiap.ch/biometric.

Face and speaker recognition

We leveraged prior work on distribution modelling for part-based face recognition using session variability modelling techniques. Session variability modelling aims to explicitly model and suppress detrimental within-class (inter-session) variation. We examined two techniques to do this, inter-session variability modelling (ISV) and joint factor analysis (JFA), which were initially developed for speaker recognition. Next, we explored Total Variability modelling (TV), so called i-Vectors originally proposed for speaker recognition, for the task of face recognition. We also developed recently a scalable formulation of Probabilistic Linear Discriminant Analysis (PLDA). PLDA is a probabilistic model that has been shown to provide state-of-the-art performance for both face and speaker recognition.

Figure 24: Illustration of the heterogeneous face recognition problem: matching Visible spectra and Near-Infrared spectra (left), matching Visible spectra and sketch (middle), matching Visible spectra and Thermal spectra (right).

Heterogeneous face recognition

The task of Heterogeneous Face Recognition (Figure 24) consists in to match face images that were sensed in different modalities, such as sketches to photographs, thermal images to photographs or near infrared to photographs. We we proposed a novel and generic approach based on Inter-session Variability Modelling to handle this task.

Presentation attack detection

One important aspect of biometric systems is their reliability not only when assaulted by impostors, but also under different types of attacks. One possible security treat is presentation attacks (aka spoofing attacks): an action of outwitting a biometric sensor by presenting a counterfeit biometric evidence of a valid user (Figure 25). It is a direct attack to the sensory input of the biometric system and the attacker does not need previous knowledge about the recognition algorithm. Most of the biometric modalities are not resistant to presentation attacks: a biometric system is usually designed to only recognise identities without concern whether the sample comes from a live person or not. Despite the existence of very sophisticated biometric systems nowadays, the task of implementing presentation attack detection (PAD aka anti-spoofing) schemes for them has attracted much less attention.

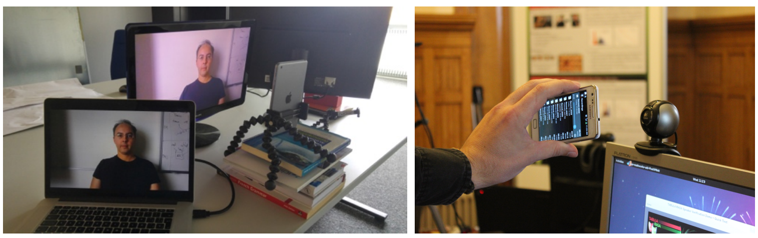

Figure 25: Illustration of video and audio presentation attacks.

Speaker recognition (SR) systems are highly vulnerable to presentation attacks (Figure 26) limiting their wide practical deployment. Therefore, to protect against such attacks, effective PAD techniques, need to be developed. We focused on the integration of PAD and SR systems, demonstrating the effect of score-based integration on recognition and attack detection performance.

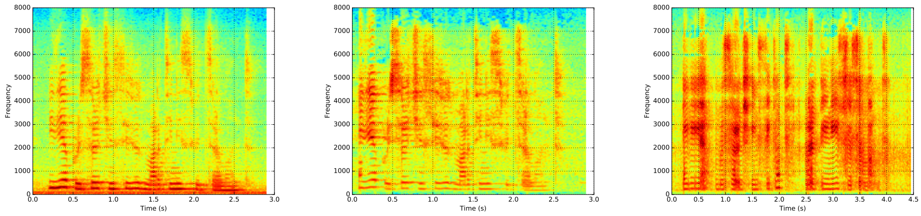

Figure 26: Spectrograms of genuine samples (left) vs. attack (middle and right) samples.

Remote photoplethysmography

Photoplethysmography (PPG) consists in measuring the variation in volume inside a tissue, using a light source. The aim of remote photoplethysmography (rPPG) is to measure the same variations, but using ambient light instead of structured light and widely available sensors such as a simple webcam (Figure 27).

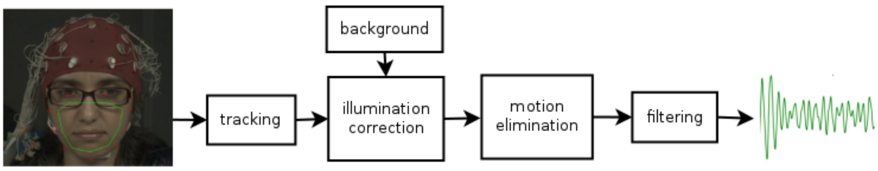

Figure 27: Illustration of remote photoplethysmography: colors from the video signal are filtered to produce an estimation of the heart beat signal.

We presented a new, publicly available database containing a relatively large number of subjects recorded under two different lighting conditions. Also, three state-of-the-art rPPG algorithms from the literature were selected, implemented and released as open source free software.

Vein biometrics

Vein recognition relies on the complex network of blood vessels located under the human skin. The vascular image of veins, that are located about 3 mm below the surface of the skin, is typically captured under near infra-red (NIR, wavelength between 700 and 1000 nm) illumination. The vein pattern can then be extracted with image pre-processing techniques (Figure 28) and used for recognition by any pattern recognition method.

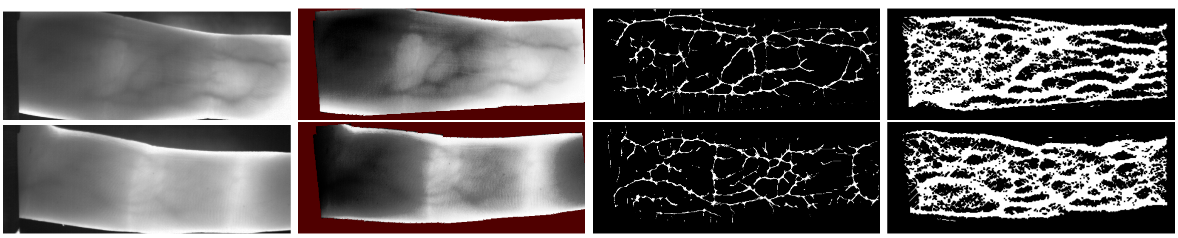

Figure 28: Illustration of image processing for vein biometrics (first row for data subject 1 and second row for data subject 2): the original image (left) is geometrically normalised (middle left) and binarised with maximum curvature (middle right) or repeated line tracking (right).

Swiss Centre for Biometrics Research and Testing

In 2014, the Idiap Research Institute launched the "Swiss Centre for Biometrics Research and Testing" (www.biometrics-center.ch), a competence centre within the Institute following recent successes in co-ordinating International research projects in Biometrics (MOBIO, TABULA RASA and BEAT). The aim of this centre is to serve as a legacy for these projects and to push for industry-driven research and testing in biometrics.

The centre attracted the attention of large companies (license, research and testing agreements) and led to many new projects (DARPA, IARPA, CTI). In 2017, the centre has developed over 3 directions:

- Improving the security of the BEAT platform: an external IT security company conducted an audit to identify vulnerabilities of the infrastructure, allowing our engineers to strengthen the platform.
- Enriching the BEAT platform: our researchers and engineers developed Deep Learning and speech processing technologies to enrich the BEAT platform with new content. More particularly our engineers developed a new feature to execute GPU simulations at the heart of modern AI tools.
- Testing: our researchers conducted the independent testing of biometrics products from SMEs (e.g., KeyLemon now acquired by AMS) and the evaluation of a prototype for a large corporation.

Key publications

[1] A. Mohammadi, S. Bhattacharjee and S. Marcel, "Deeply Vulnerable – a study of the robustness of face recognition to presentation attacks", *IET Biometrics*, 1–13, 2017.

[2] P. Korshunov and S. Marcel, "Impact of score fusion on voice biometrics and presentation attack detection in cross-database evaluations", *IEEE Journal of Selected Topics in Signal Processing*, 695–705, 2017.

[3] H. Muckenhirn, P. Korshunov, M. Magimai-Doss and S. Marcel, "Long-Term Spectral Statistics for Voice Presentation Attack Detection", *IEEE/ACM Transactions on Audio, Speech and Language Processing*, 2017.

[4] S. Bhattacharjee and S. Marcel, "What you can't see can help you: extended-range imaging for 3D-mask presentation attack detection", *IEEE International Conference of the Biometrics Special Interest Group (BIOSIG)*, 2017.

[5] A. Anjos, L. El-Shafey and S. Marcel, "BEAT: An Open-Science Web Platform", *International Conference on Machine Learning*, 2017.

2.9 Natural Language Processing

Overview

Head: Dr. Andrei Popescu-Belis (MEng, École Polytechnique, France, 1995; MSc, Université Pierre et Marie Curie, France, 1996; PhD, Université Paris-Sud, France, 1999)

Group overview: The Natural Language Processing group studies text analysis at the semantic and pragmatic levels for two main applications: machine translation (MT) and information retrieval (IR). The group aims to improve the state of the art on core semantic and pragmatic analysis problems, such as the disambiguation of nouns and pronouns, topic labeling, keyword extraction, or sentiment analysis – so that they improve in turn MT and IR. Regarding MT, we combine text-level processing techniques with statistical and neural MT systems to improve translation quality. Regarding IR, we design models of document content and sentiment to improve multimedia classification and recommendation.

In 2017, the NLP group had the following members: the head of the group, one postdoctoral student, two PhD students, and two interns. In September 2017, the head of the group started a professorship at HEIG-VD / HES-SO, while a new senior researcher was recruited in the same field. Therefore, the NLP group became the NLU group, which is covered in a different section of this report.

Key scientific outputs: In 2017, we have demonstrated that the semantic analysis of noun phrases is also beneficial to neural MT (in additional to statistical phrase-based MT). We have shown that word sense disambiguation could be combined with statistical and neural MT, and that coreference resolution was helpful to statistical MT, using source/target coreference similarity as a component of the objective function. The hierarchical neural networks with attention proposed for cross-lingual transfer on a document classification in 8 languages have met with considerable success, earning a best paper award at one of the main NLP conferences (IJCNLP 2017, Taiwan).

Document-level machine translation

We focus on the translation of words that depend on long-range relationships between sentences. The main examples are discourse connectives (which have been our main focus in 2010–2014),[20] verb tenses (studied in 2013–2014), and pronouns and noun phrases (our current focus).[21] The NLP group coordinates a consortium with teams from Geneva, Zürich and Utrecht, which has reached an end in 2017. Our main 2017 achievements in document-level MT were the following ones.

We designed a method to enforce the consistency of noun translation, as illustrated in Fig. 29. The method is based on a classifier that decides whether two occurrences of the same noun in a source text should be translated consistently. We combined these classifiers with phrase-based SMT for Chinese-to-English and German-to-English, in two ways: automatic post-editing of noun translations vs. re-ranking of MT hypotheses. Our method closes up to 50% of the gap in BLEU scores between the baseline and the oracle classifier.

In addition to noun consistency, we proposed two methods to improve the translation of pronouns based on coreference links. First, we built a *coreference model*, which captures the probabilistic connection between a pronoun and the features of its antecedent (gender, number and humanness) by learning from the output of an anaphora resolution system on parallel texts. When used in combination with a phrase-based MT decoder, our method improves pronoun translation for English-French and Spanish-English MT. Second, we generalized our studies and started to improve the translation of entity mentions (nouns or pronouns) with the help of coreference resolution. We implemented a coreference-

[20] Supported by the COMTIS SNSF Sinergia project, 2010-2013, see www.idiap.ch/project/comtis.
[21] Supported by the MODERN SNSF Sinergia project, 2013-2017, see www.idiap.ch/project/modern.

Example 1
Source: nach einführung dieser **politik** [...] die **politik** auf dem gebiet der informationstechnik [...]
Reference: once the **policy** is implemented [...] the information technology **policy** [...]
Baseline MT: after introduction of **policy** [...] the **politics** in the area of information technology [...]
Consistent MT: after introduction of **policy** [...] the **policy** in the area of information technology [...]

Example 2
Source: 欺诈性旅行或身份**证件**系指有下列情形之一的任何旅行或身份**证件**
Reference: Fraudulent travel or identity **document**; shall mean any travel or identity **document**
MT: 欺诈性 travel or identity **papers**. 系指 have under one condition; any travel, or identity **document**

Figure 29: Inconsistent translations of repeated nouns, in blue, from German (Ex. 1) and Chinese (Ex. 2) into English. Only Ex. 1 is truly mistaken, and is corrected by our MT system.

aware translation system that helps to disambiguate the translation of mentions by optimizing the similarity of mention-grouping in source and target documents, either by re-ranking entire sentence, or by post-editing mentions only. Again, this improved Spanish-English translation of pronominal mentions. Finally, using results from the DiscoMT 2015 shared task, we showed that a reference-based metric for the translation of pronouns correlates well with human judgments.

Text analysis for multilingual document representations

We develop methods for text and multimedia information retrieval, classification and recommendation, based on the analysis of content and sentiment, and on the networked structure of (social) media repositories. This work was supported by the SUMMA EU project[22].

Within a weighted multiple-instance framework, we showed how to jointly learn to focus on relevant parts of documents and to classify them into categories. The model can operate on various sentence and document representations, including intermediate hidden states of a neural network, and has state-of-the-art results for aspect rating prediction (i.e. multi-aspect sentiment analysis). Generalizing our previous work on sentiment analysis, we demonstrated the explanatory power of our model by comparing its predicted aspect saliency values with those assigned by humans, and we showed that the aspect saliencies benefit text segmentation and summarization. We also showed that this model is mathematically equivalent to attention mechanisms recently proposed for neural networks.

We developed a state-of-the-art architecture for multilingual document modeling, based on hierarchical neural networks. We explored 3 options for multilingual training over disjoint label spaces: sharing the parameters at each layer of the network, sharing the attention mechanisms at each layer, and sharing both. We demonstrated that these architectures are useful for transferring knowledge to low-resource languages, and that they also improve over monolingual models on the full-scale data. The models were applied to topic labeling in the SUMMA project, using 600k articles from Deutsche Welle, in 8 languages, with multiple topic labels. We thus delivered a deep multilingual tagging library for the SUMMA demonstrator, illustrated in Fig. 30 for one test article. This work received the Best Paper Award at the IJCNLP 2017 Conference in Taiwan.

Key publications

[1] Pappas N. & Popescu-Belis A. (2017) – Multilingual Hierarchical Attention Networks for Document Classification. *Proceedings of the 8th International Joint Conference on Natural Language Processing (IJCNLP)*, Taipei, Taiwan. Best paper award.

[2] Pu X., Pappas N. & Popescu-Belis A. (2017) – Sense-Aware Statistical Machine Translation using Adaptive Context-Dependent Clustering. *Proceedings of the 2nd Conference on Machine*

[22] See www.summa-project.eu.

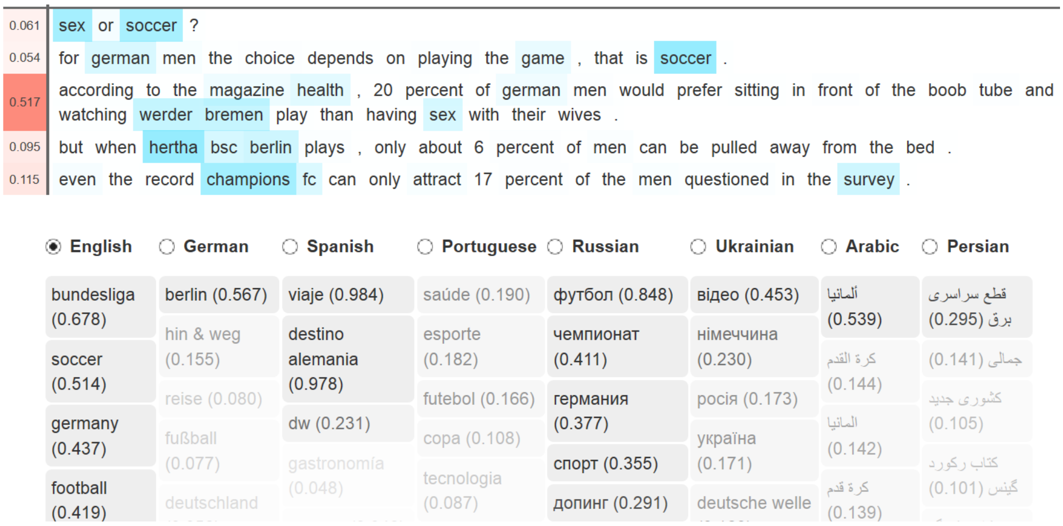

Figure 30: Multilingual topic labeling in 8 languages of a news article from Deutsche Welle.

Translation (WMT), Copenhagen, Denmark.

[3] Pu X., Mascarell L. & Popescu-Belis A. (2017) – Consistent Translation of Repeated Nouns using Syntactic and Semantic Cues. *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, Valencia, Spain.

[4] Miculicich Werlen L. & and Popescu-Belis A. (2017) – Using Coreference Links to Improve Spanish-to-English Machine Translation. *Proceedings of the EACL Workshop on Coreference Resolution Beyond OntoNotes (CORBON)*, Valencia, Spain.

[5] Pappas N. & Popescu-Belis A. (2017) – Explicit Document Modeling through Weighted Multiple-Instance Learning. *Journal of Artificial Intelligence Research (JAIR)*, vol. 58, p. 591-626.

2.10 Natural Language Understanding

Overview

Head: Dr. James Henderson (BSc, Massachusetts Inst. Technology, USA, 1987; MSE & PhD, Univ. Pennsylvania, USA, 1991,1994; MER & Chargé de Cours, Univ. Geneva, 2008–2012,2012–)

Group overview: The Natural Language Understanding group was created in September 2017, in part as a continuation of the previous Natural Language Processing group. The NLU group studies deep learning for natural language processing tasks, focusing on models with learned representations of the meaning of text. These tasks include machine translation, information retrieval, language modelling, syntactic and semantic parsing, sentiment analysis, text classification and entailment detection, applied to both text and speech. We focus on recurrent and attention-based neural network models, vector-space representations for entailment (rather than similarity), unsupervised learning of semantic representations, and multi-task learning.

From September until December 2017, the NLU group had the following members: the head of the group, one postdoctoral student, and two PhD students.

Key scientific outputs: From September to December 2017 the work of the NLU group has included both work started within the previous Natural Language Processing group and work developing new lines of research. Notable in the former is work showing improvements in neural machine translation by adding an attention mechanism over the previous target-side words, and label-aware text classification, where text descriptions of output classes are used in deep learning models to not only generalise to classes with no training data ("zero-shot" learning) but, for the first time, to improve accuracies on classes that have training data. The new lines of research build on recent results by the head of group on the unsupervised learning of word embeddings that predict entailment. Two grant proposals were submitted that develop this entailment-vectors framework for sentence-level textual entailment and its application in opinion summarisation and the analysis of the interpretation of official announcements.

Neural Network Architectures for NLP Tasks

The impressive initial results from using deep learning architectures in NLP tasks, such as for neural machine translation, are now being improved by modifying the architectures to better embody inductive biases that are important for NLP tasks. This is manifested in our work on self-attention in the decoder for neural machine translation. Instead of generating the translation with the standard LSTM decoder, we add an attention mechanism over the prefix of generated words. This better models non-local correlations in the output sentence. We find that this mechanism works best if the attention applies directly to the word embeddings, rather than the LSTM states at each word.

Another way to import inductive biases into a deep learning architecture is to explicitly model the similarity between output classes. This can be done with what we call label-aware models, where the text of the output class label is used to compute a vector for each output class, such that similarity between classes is represented as similarity between their vectors. Previously such models have generalised well in "zero-shot" learning, where the output classes at test time do not occur at all in the training data. With our proposed use of cross-entropy loss in our attention-based label-aware architecture for text classification, we showed for the first time that such models can also show improvement for the output classes which do occur in the training data, where previously non-label-aware classification has performed better. This effect is particularly marked with very large label sets, where our model also trains faster, in time independent of the label set size.

Vector-Space Models of Entailment

Entailment is the relation which reflects adding information or abstracting away from information, and is fundamental to many theories of semantics. But typically vector-space models, such as word embeddings and deep learning architectures, are based on similarity between vectors, not entailment. Prior to joining Idiap, the head of group has developed a vector-space framework for modelling entailment. Unprecedented results have been achieved on the unsupervised learning of word embeddings that predict entailment between words (hyponymy), using a novel entailment-based distributional semantic model. This framework extends to entailment between larger texts (textual entailment), and has particular relevance to large-scale summarisation tasks. Textual entailment is important for summarisation because the summary must abstract away from unimportant information but still contain as much information as possible from the text.

This line of work has been the basis of two grant proposals submitted at Idiap in 2017. The first grant proposal includes exploiting the entailment-vectors framework and the word embedding results to build models of the fundamental problem of textual entailment. In parallel, the project would apply textual entailment to developing models and resources for large-scale opinion summarisation. Opinion summarisation is where a large collection of texts must be summarised, including the popularity of different opinions. The second grant proposal is for an inter-disciplinary project investigating the interpretation of official announcements. The interpretation text is a form of summary of the announcement text, so textual entailment is again relevant. But typically interpretation involves more complex reasoning and background knowledge, making it a challenging testbed.

Key publications

[1] Lesly Miculicich Werlen, Nikolaos Pappas, Dhananjay Ram & Andrei Popescu-Belis (to appear) – Self-Attentive Residual Decoder for Neural Machine Translation. *Proceedings of the 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL 2018)*, New Orleans, Louisiana, USA.

3 List of Non-permanent Scientific Staff

3.1 Research Associates

Lastname	Firstname	Nationality	Supervisor	Start	Estimated End
BHATTACHARJEE	Sushil	CH	Marcel Sébastien	01.01.2018	31.12.2018
HEUSCH	Guillaume	CH	Marcel Sébastien	08.06.2015	
IMSENG	David	CH	Bourlard Hervé	01.12.2017	30.11.2018
KORSHUNOV	Pavel	EE	Marcel Sébastien	01.01.2018	31.12.2018
LABHART	Florian	CH	Gatica Perez Daniel	01.11.2017	31.12.2019
RABELLO DOS ANJOS	André	BR	Marcel Sébastien	01.01.2017	31.12.2018

Figure 31: List of research associates in alphabetical order.

3.2 Post-doctoral Scholars

Lastname	Firstname	Nationality	Supervisor	Start	Estimated End
ABROL	Vinayak	IN	Magimai Doss Mathew	01.01.2018	31.12.2018
AICHINGER	Ida	AT	Bourlard Hervé	01.03.2018	28.02.2019
CAO	Yuanzhouhan	CN	Odobez Jean-Marc	15.11.2017	15.11.2018
GEISSBUHLER	David	CH	Marcel Sébastien	01.06.2017	31.05.2019
GEORGE	Anjith	IN	Marcel Sébastien	21.08.2017	31.03.2019
KHONGLAH	Banriskhem	IN	Bourlard Hervé	01.03.2018	28.02.2019
KODRASI	Ina	AL	Bourlard Hervé	01.12.2017	30.11.2018
KOMATY	Alain	FR	Marcel Sébastien	01.05.2017	30.04.2018
KOTWAL	Ketan	IN	Marcel Sébastien	01.01.2018	31.12.2018
KRIVOKUCA	Vedrana	NZ	Marcel Sébastien	15.01.2017	15.01.2019
KUPCSIK	Andras	HU	Calinon Sylvain	03.04.2017	31.03.2019
LEPOITTEVIN	Yann	FR	Fleuret François	01.08.2016	30.04.2018
LIU	Gang	CN	Odobez Jean-Marc	18.09.2017	17.09.2018
MADIKERI	Srikanth	IN	Motlicek Petr	01.08.2017	31.07.2018
NGUYEN	Laurent	CH	Gatica-Perez Daniel	01.12.2017	15.03.2019
NIKISINS	Olegs	LV	Marcel Sébastien	01.05.2017	30.04.2019
PAOLILLO	Antonio	IT	Calinon Sylvain	15.01.2018	15.01.2019
PAPPAS	Nikolaos	GR	Henderson James	01.02.2017	31.01.2019
RAZAVI	Marzieh	IR	Magimai Doss Mathew	01.08.2017	31.07.2018
REKABSAZ	Navid	IR	Henderson James	01.03.2018	28.02.2019
SARFJOO	Saeed	IR	Marcel Sébastien	01.01.2018	31.12.2018
SHAKAS	Alexis	CY	Garner Phil	22.01.2018	22.01.2019
SRINIVASAMURTHY	Ajay	IN	Motlicek Petr	01.09.2016	28.02.2018
VILLAMIZAR	Michael	ES	Odobez Jean-Marc	01.11.2017	31.10.2018
VLASENKO	Bogdan	DE	Magimai Doss Mathew	01.07.2017	30.06.2018

Figure 32: List of post-doctoral scholars in alphabetical order.

3.3 PhD Students

Lastname	Firstname	Nationality	Supervisor	Start	Estimated End
AMINIAN	Bozorgmehr	CH	Odobez Jean-Marc	01.04.2018	31.03.2022
CAN	Gulcan	TR	Gatica-Perez Daniel	15.09.2013	31.01.2018
CHAVDAROVA	Tatjana	MK	Fleuret François	01.12.2014	30.11.2018
DE FREITAS PEREIRA	Tiago	BR	Marcel Sébastien	01.07.2014	28.02.2019
DESPRÈS	Nicolas	FR	Odobez Jean-Marc	01.02.2018	31.01.2022
DEY	Subhadeep	IN	Motlicek Petr	01.04.2014	31.12.2018
DIGHE	Pranay	IN	Bourlard Hervé	01.08.2014	31.07.2018
DUBAGUNTA	Pavankumar	IN	Magimai Doss Mathew	01.04.2017	31.03.2021
HE	Weipeng	CN	Odobez Jean-Marc	01.06.2016	31.05.2020
JANBAKHSHI	Parvaneh	IR	Bourlard Hervé	01.02.2018	31.01.2022
JAQUES	Christian	CH	Liebling Michael	01.04.2016	31.03.2020
JAQUIER	Noémie	CH	Calinon Sylvain	02.08.2016	01.08.2020
JOSE	Cijo	IN	Fleuret François	01.01.2014	31.05.2018
KABIL	Selen	TR	Magimai Doss Mathew	01.05.2018	30.04.2022
KATHAROPOULOS	Angelos	GR	Fleuret François	01.03.2017	28.02.2021
KULAK	Thibaut	FR	Calinon Sylvain	01.12.2017	30.11.2021
LE	Nam	VN	Odobez Jean-Marc	15.01.2015	15.01.2019
MARIANI	Olivia	CH	Liebling Michael	01.06.2016	31.05.2020
MARTINEZ GONZALEZ	Angel	MX	Odobez Jean-Marc	15.11.2016	15.11.2020
MICULICICH	Lesly	PE	Henderson James	01.03.2016	29.02.2020
MOHAMMADI	Amir	IR	Marcel Sébastien	15.01.2016	15.01.2020
MUCKENHIRN	Hannah	FR	Magimai Doss Mathew	13.07.2015	12.07.2019
MURALIDHAR	Skanda	IN	Gatica-Perez Daniel	01.12.2014	30.11.2018
NEWLING	James	GB	Fleuret François	15.09.2013	28.02.2018
PHAN	Trung	VN	Gatica-Perez Daniel	01.09.2015	31.08.2018
PIGNAT	Emmanuel	CH	Calinon Sylvain	01.03.2016	29.02.2020
PU	Xiao	CN	Henderson James	01.08.2014	31.07.2018
RAM	Dhananjay	IN	Bourlard Hervé	01.01.2015	31.12.2018
SCHNELL	Bastian	DE	Garner Phil	01.05.2017	30.04.2021
SEBASTIAN	Jilt	IN	Magimai Doss Mathew	01.09.2017	31.08.2018
SHAJKOFCI	Adrian	CH	Liebling Michael	20.09.2016	20.09.2020
SIEGFRIED	Rémy	CH	Odobez Jean-Marc	01.01.2017	31.12.2020
SRINIVAS	Suraj	IN	Fleuret François	01.03.2017	28.02.2021
TANWANI	Ajay	PK	Calinon Sylvain	12.01.2015	13.04.2018
TONG	Sibo	CN	Garner Phil	01.05.2016	30.04.2020
TORNAY	Sandrine	CH	Magimai Doss Mathew	01.08.2016	31.07.2020
YU	Yu	CN	Odobez Jean-Marc	15.07.2015	15.07.2019

Figure 33: List of PhD students in alphabetical order.

3.4 Interns

Lastname	Firstname	Nationality	Supervisor	Start	Estimated End
BOURGEOIS	Dylan	FR	Calinon Sylvain	23.12.2017	26.01.2018
CANDY	Romain	FR	Bourlard Hervé	01.03.2018	31.08.2018
CHAPUIS	Lucille	CH	Gatica-Perez Daniel	01.01.2018	31.05.2018
CHRISTEN	Stefano	CH	Gatica-Perez Daniel	01.02.2018	30.04.2018
COTTER	Audrey	CH	Marcel Sébastien	15.12.2017	31.01.2018
DELEZE	Laura	CH	Gatica-Perez Daniel	27.11.2017	30.04.2018
GAUTIER	Athénaïs	FR	Ginsbourger David	16.04.2018	16.10.2018
GIRGIN	Hakan	TR	Calinon Sylvain	01.03.2018	31.08.2018
KABIL	Selen	TR	Magimai Doss Mathew	19.09.2017	30.04.2018
LOUPI	Dimitra	GR	Garner Phil	26.02.2018	24.08.2018
MA	Wei	IN	Calinon Sylvain	19.09.2017	19.01.2018
MARELLI	François	BE	Bourlard Hervé	01.02.2018	31.07.2018
MCGARRITY	Marie-Noëlle	CH	Gatica-Perez Daniel	01.01.2018	31.05.2018
MEDINA RIOS	Luis Emmanuel	MX	Gatica-Perez Daniel	19.09.2017	31.01.2018
MOSTAANI	Zohreh	IR	Marcel Sébastien	02.10.2017	09.04.2019
ROSSIER	Alain	CH	Fleuret François	17.07.2017	17.01.2018
SCHNEIDER	Eva	CH	Gatica-Perez Daniel	01.01.2018	30.04.2018
STANDAERT	Florian	CH	Calinon Sylvain	23.12.2017	26.01.2018
STERPU	George	RO	Motlicek Petr	05.02.2018	04.08.2018
STREHL	Thomas	CH	Gatica-Perez Daniel	02.01.2018	30.04.2018
SUOMALAINEN	Markku	FI	Calinon Sylvain	29.01.2018	15.06.2018
THURRE	Veronique	CH	Marcel Sébastien	27.11.2017	27.01.2018
VUILLEUMIIER	Noémie	CH	Gatica-Perez Daniel	01.04.2018	31.05.2018

Figure 34: List of interns in alphabetical order.

3.5 Visitors

Lastname	Firstname	Nationality	Supervisor	Start	Estimated End
CARRERA SOTO	Laura	CU	Ginsbourger David	10.11.2017	31.05.2018
HU	Rui	CN	Gatica-Perez Daniel	01.12.2017	31.05.2018
SCHÄRER	Cedric	CH	Ginsbourger David	08.03.2018	31.08.2018

Figure 35: List of visitors in alphabetical order.

4 Active and Granted Projects in 2017

An overview of the projects that have been active during the year 2017 is presented in Section 4.1. The projects are grouped by funding agencies, starting with European and International Agencies, then Swiss Agencies to finish with the projects funded by industrial partners.

Section 4.2 presents the list of projects accepted during 2017 but starting in 2018.

4.1 Projects in Progress during 2017

4.1.1 Projects Funded by European and International Agencies

1. Name: **MACADAMS** (Modifying Adhoc Centralised Advertisement with Digit Arena Multicast over Satellite)
   Funding: Eurostars Program
   Coordinator: Digit Arena
   Duration: 2016.09.01 - 2018.08.30
   Partner(s): Idiap Research Institute, Ecole Centrale de Lyon, Eurovision (European Broadcast Union)

2. Name: **SIIP** (Speaker Identification Integrated Project)
   Funding: FP7
   Coordinator: Verint System Ltd
   Duration: 2014.05.01 - 2018.04.30
   Partner(s): Idiap Research Institute, SAil Labs Technology AG, Singularlogic Anonymos etairia Pliroforiakon Systimaton & Efarmogon Pliroforikis, Green Fusion Limited (Data Fusion International), Synthema S.R.L., Ok2Go, Loquendo Spa, International Biometric Group (Uk) Limited, Cassidian SAs, Rijksuniversiteit Groningen, Inov Inesc Inovacao - Instituto de Novas Tecnologias, University of Warwick, Laboratorio Di Scienze della Cittadinanza, The International Criminal Police Organization, Police Service of Northern Ireland, Ministério Da Justiça, Lisboa

3. Name: **DEXROV** (Effective Dexterous ROV Operations in Presence of Communications Latencies)
   Funding: H2020-BG
   Coordinator: Space Applications Services
   Duration: 2015.03.01 - 2018.08.31
   Partner(s): Idiap Research Institute, Jacobs University Bremen, Comex SA., Interuniversity Center of Integrated Systems For The Marine Environment, Graal Tech S.R.L.

4. Name **TESLA** (An Adaptive Trust-based e-assesment System for Learning)
   Funding H2020-ICT
   Coordinator Fundacio Per A La Universitat Oberta de Catalunya
   Duration 2016.01.01 - 2018.12.31
   Partner(s) Idiap Research Institute, Imperial College London, The Open University, Télécom Sudparis, Open Universiteit Nederland, European Association For Quality Assurance In Higher Education Aisbl, Universite de Namur Asbl, AGència Per A La Qualitat del Sistema Universitari de Catalunya, Lplus Gmbh, Sofiiski Universitet Sveti Kliment Ohridski, Protos Sistemas de Información, S.L., Technical University of Sofia, Anadolu University, Jyvaskylan Yliopisto, European Quality Assurance Network For Informatics Education E.V., Instituto Nacional de Astrofisica, Optica Y Electronica, Wfsw, SA, Institut Mines-Telecom

5. Name **MUMMER** (MultiModal Mall Entertainment Robot)
   Funding H2020-ICT
   Coordinator University of Glasgow
   Duration 2016.03.01 - 2020.02.28
   Partner(s) Idiap Research Institute, Centre National de La Recherche Scientifique, Aldebaran Robotics, Teknologian Tutkimuskeskus Vtt, Kiinteistö Oy Ideapark Ab

6. Name **SUMMA** (Scalable Understanding of Multilingual Media)
   Funding H2020-ICT
   Coordinator University of Edinburgh
   Duration 2016.02.01 - 2019.01.31
   Partner(s) Idiap Research Institute, University College London, British Broadcasting Corporation, deutsche Welle, Priberam Informatica S.A., Leta, Qatar Computing Research Institute

7. Name **4DHEART** (4D analysis of heart development and regeneration using advanced light microscopy)
   Funding H2020-MSCA
   Coordinator Fundacion Centro Nacional de Investigaciones Cardiovasculares Carlos Iii
   Duration 2016.10.01 - 2020.09.30
   Partner(s) Idiap Research Institute, Centre National de La Recherche Scientifique, Universität Bern, Acquifer AG, Bitplane AG, Leica Microsystems Cms Gmbh, 4D-Nature ImAGing Consulting, S. L., Centre Europeen de Recherche En Biologie et Medecine

8. Name **TAPAS** (Training Network on Automatic Processing of PAthological Speech)
   Funding H2020-MSCA
   Coordinator Idiap Research Institute
   Duration 2017.11.01 - 2021.10.31
   Partner(s) University of Sheffield, Philips, Radboud University Nijmegen - Stichting Katholieke Universiteit, Ludwig-Maximilians-Universität München, Institut de Recherche En Informatique de Toulouse, Antwerpen University Hospital, Friedrich-Alexander-Universität Erlangen Nuernberg, Instituto de Engenharia de Sistemas E Computadores, Investigacao E desenvolvimento Em Lisboa, Interuniversitair Micro-Electronica Centrum Imec Vzw, Stichting Het Nederlands Kanker Instituut - Antoni Van Leeuwenhoek Ziekenhuis, Universitaet Augsburg

9. Name **MALORCA** (Machine Learning of Speech Recognition Models for Controller Assistance)
   Funding H2020-SESAR
   Coordinator Deutsches Zentrum Fuer Luft - Und Raumfahrt Ev
   Duration 2016.04.01 - 2018.03.31
   Partner(s) Idiap Research Institute, Universität des SAarlandes, Rízení Letového Provozu Ceské Republiky, Austro Control Osterreichische Gesellschaft Fur Zivilluftfahrt Mbh

10. Name **SWAN** (Secure Access Control over Wide Area Network)
    Funding Research Council of Norway
    Coordinator Hogskolen I Gjovik
    Duration 2016.01.01 - 2019.12.31
    Partner(s) Idiap Research Institute, Morpho, Bankenverband, Universitetet I Oslo, Zwipe As

11. Name **SAVI** (Spotting Audio-Visual Inconsistencies)
    Funding USA DARPA
    Coordinator Sri International
    Duration 2016.05.19 - 2020.05.18
    Partner(s) Idiap Research Institute

12. Name **BATL** (Biometric Authentification with Timeless Learner)
    Funding USA IARPA
    Coordinator University of Southern California
    Duration 2017.03.01 - 2021.02.28
    Partner(s) Idiap Research Institute

13. Name **SARAL** (Summarization and domain-Adaptive Retrieval of Information Across Languages)
    Funding USA IARPA
    Coordinator University of Southern California
    Duration 2017.10.01 - 2019.03.31
    Partner(s) Idiap Research Institute, Massachusetts Institute of Technology, Raytheon Company, Reenselaer Polytechnic Institute, University of Massachusetts Amherst, Northeastern University

4.1.2 Projects Funded by Swiss Agencies

1. Name **SOCIALFOOD** (Social Food and People: Characterizing Food Consumption Behaviour on Visual Social Media)
   Funding EPFL Collaboration
   Coordinator Idiap Research Institute
   Duration 2017.01.01 - 2017.12.31
   Partner(s) École Polytechnique Fédérale de Lausanne

2. Name **MULTI-CROWD** (Tracing Swiss Heritage Speakers' Identities in North America)
   Funding EPFL Collaboration
   Coordinator University of Lausanne
   Duration 2017.02.01 - 2018.02.28
   Partner(s) Idiap Research Institute

3. Name **CREM-IDIAP** (Pour une recherche fondamentale et appliquée au service des systèmes énergétiques territoriaux en Valais)
   Funding Etat du Valais
   Coordinator Centre de Recherches Energetiques et Municipales
   Duration 2016.11.01 - 2017.03.31
   Partner(s) Idiap Research Institute

4. Name **ADDG2SU-EXT.** (Flexible Acoustic data-driven Grapheme to Subword Unit Conversion)
   Funding Hasler Foundation
   Coordinator Idiap Research Institute
   Duration 2016.03.01 - 2017.02.28
   Partner(s)

5. Name **DCROWDLENS** (Crowdsourcing the Assessment of Deep Visual Explanations for Subjective Variables)
   Funding Hasler Foundation
   Coordinator Idiap Research Institute
   Duration 2017.05.15 - 2018.01.15
   Partner(s)

6. Name **FLOSS** (Flexible Linguistically-guided Objective Speech aSessment)
   Funding Hasler Foundation
   Coordinator Idiap Research Institute
   Duration 2017.03.01 - 2020.02.29
   Partner(s)

7. Name **MASH-2** (Massive Sets of Heuristics for Machine Learning II)
   Funding Hasler Foundation
   Coordinator Idiap Research Institute
   Duration 2013.07.01 - 2017.09.30
   Partner(s)

8. Name **MEMUDE** (Multi-view Detection with Metric-learning for Deep Network Fusion)
   Funding Hasler Foundation
   Coordinator Idiap Research Institute
   Duration 2017.06.01 - 2019.05.31
   Partner(s)

9. Name **SETVISU** (Learning and visualizing dangerous regions in multivariate parameter spaces)
   Funding Hasler Foundation
   Coordinator Idiap Research Institute
   Duration 2016.12.01 - 2017.08.31
   Partner(s)

10. Name **GPUS** (Acquisition d'un cluster de calcul dédié à l'Intelligence Artificielle)
    Funding Loterie Romande
    Coordinator Idiap Research Institute
    Duration 2017.09.01 - 2018.08.31
    Partner(s)

11. Name **VALAIS+** (Une plateforme pour mieux connaître l'espace de vie du canton)
    Funding Loterie Romande
    Coordinator Idiap Research Institute
    Duration 2015.09.01 - 2017.03.31
    Partner(s)

12. Name **ODESSA** (Online Diarization Enhanced by recent Speaker identification and Sequential learning Approaches)
    Funding SNF - ANR
    Coordinator Centre National de La Recherche Scientifique
    Duration 2016.03.01 - 2019.02.28
    Partner(s) Idiap Research Institute, Eurecom

13. Name **LIVEHEART** (The Cellular Basis of Cardiac Development Revealead by Live Imaging)
    Funding SNF - ANR
    Coordinator Institut de Génétique et de Biologie Moléculaire et Cellulaire
    Duration 2016.06.01 - 2019.05.31
    Partner(s) Idiap Research Institute, University of Bern, École Polytechnique Paris

14. Name **TACT-HAND** (Improving control of prosthetic hands using tactile sensors and realistic machine learning)
    Funding SNF - DACH
    Coordinator deutsches Zentrum Fuer Luft - Und Raumfahrt Ev
    Duration 2016.04.01 - 2019.03.31
    Partner(s) Idiap Research Institute, Universitaet Bielefeld

15. Name **MAAYA** (Multimedia Analysis and Access for Documentation and Decipherment of Maya Epigraphy)
    Funding SNF - DACH
    Coordinator Idiap Research Institute
    Duration 2013.09.01 - 2017.04.30
    Partner(s) University of Geneva, University of Bonn

16. Name **COMETS-M** (Computational Methods for Temporal Super-resolution Microscopy)
    Funding SNF - Division II
    Coordinator Idiap Research Institute
    Duration 2016.04.01 - 2019.03.30
    Partner(s)

17. Name **HFACE** (Heterogeneous Face Recognition)
    Funding SNF - Division II
    Coordinator Idiap Research Institute
    Duration 2014.07.01 - 2018.06.30
    Partner(s)

18. Name **ISUL** (Importance sampling for large-scale unsupervised learning)
    Funding SNF - Division II
    Coordinator Idiap Research Institute
    Duration 2017.03.01 - 2020.02.28
    Partner(s)

19. Name **MASS** (Multilingual Affective Speech Synthesis)
    Funding SNF - Division II
    Coordinator Idiap Research Institute
    Duration 2017.05.01 - 2020.04.30
    Partner(s)

20. Name **PHASER-QUAD** (Parsimonious Hierarchical Automatic Speech Recognition and Query Detection)
    Funding SNF - Division II
    Coordinator Idiap Research Institute
    Duration 2016.10.01 - 2018.09.30
    Partner(s)

21. Name **UNITS** (Unified Speech Processing Framework for Trustworthy Speaker Recognition)
    Funding SNF - Division II
    Coordinator Idiap Research Institute
    Duration 2015.07.01 - 2018.06.30
    Partner(s)

22. Name **I-DRESS** (Assistive Interactive robotic system for support in DRESSing)
    Funding SNF - ERA NET
    Coordinator Idiap Research Institute
    Duration 2015.12.01 - 2018.11.30
    Partner(s)

23. Name **PLATFORM-MMD** (Platform for Reproducible Acquisition, Processing, and Sharing of Dynamic, Multi-Modal Data)
    Funding SNF - R'equip
    Coordinator Idiap Research Institute
    Duration 2016.07.01 - 2017.12.31
    Partner(s)

24. Name **WILDTRACK** (Tracking in the Wild)
    Funding SNF - Sinergia
    Coordinator Idiap Research Institute
    Duration 2014.01.01 - 2017.12.31
    Partner(s) École Polytechnique Fédérale de Lausanne, Eidgenoessische Technische Hochschule Zuerich

25. Name **MOSPEEDI** (Motor Speech Disorders: characterizing phonetic speech planning and motor speech programming/execution and their impairments)
    Funding SNF - Sinergia
    Coordinator University of Geneva
    Duration 2017.10.01 - 2020.09.30
    Partner(s) Idiap Research Institute, University Hospitals of Geneva, Université Paris 3

26. Name **YOUTH@NIGHT** (A multi-disciplinary multi-method study of young people's outgoing and drinking behaviors)
    Funding SNF - Sinergia
    Coordinator Sucht Schweiz - Research Institute
    Duration 2014.01.01 - 2017.12.31
    Partner(s) Idiap Research Institute, University of Zurich

27. Name **UBIMPRESSED** (Ubiquitous First Impressions and Ubiquitous Awareness)
    Funding SNF - Sinergia
    Coordinator Idiap Research Institute
    Duration 2014.01.01 - 2017.12.31
    Partner(s) Université de Neuchâtel, Cornell University

28. Name: **MODERN** (Modeling discourse entities and relations for coherent machine translation)
    Funding: SNF - Sinergia
    Coordinator: Idiap Research Institute
    Duration: 2013.08.01 - 2017.07.31
    Partner(s): University of Geneva, Universiteit Utrecht, University of Zurich

29. Name: **SMILE** (Scalable Multimodal sign language Technology for sIgn language Learning and assessmEnt)
    Funding: SNF - Sinergia
    Coordinator: Idiap Research Institute
    Duration: 2016.03.01 - 2019.02.28
    Partner(s): University of Surrey, University of Applied Sciences of Special Needs Education

30. Name: **DUSK2DAWN** (Characterizing Youth Nightlife Spaces, Activities, and Drinks)
    Funding: SNF - Sinergia
    Coordinator: Idiap Research Institute
    Duration: 2017.07.01 - 2019.12.31
    Partner(s): University of Zurich, La Trobe University

31. Name: **ALLIES** (Autonomous Lifelong learnIng intelLigent Systems)
    Funding: SNF -ERA-NET
    Coordinator: Idiap Research Institute
    Duration: 2017.12.01 - 2020.11.30
    Partner(s): Laboratoire national de métrologie et d'essais, Université du Maine, Universitat Politecnica de Catalunya

32. Name: **DEEPCHARISMA** (Deep Learning Charisma)
    Funding: UNIL Collaboration
    Coordinator: Idiap Research Institute
    Duration: 2017.04.01 - 2017.12.31
    Partner(s):

4.1.3 Projects Funded by Industrial Partners

1. Name: **CUAV** (Feasibility Study "Countering Mini-UAVs with Mini-UAVs")
   Funding: ArmaSuisse
   Coordinator: Idiap Research Institute
   Duration: 2017.06.01 - 2017.08.31
   Partner(s):

2. Name: **CAMPRO** (Camera de profondeur)
   Funding: Fondation The Ark
   Coordinator: Idiap Research Institute
   Duration: 2017.08.01 - 2018.01.31
   Partner(s): Digit Arena

3. Name **RECOMEDIND** (RecoMed Industrialisation)
   Funding Fondation The Ark
   Coordinator Recapp
   Duration 2017.04.01 - 2017.08.31
   Partner(s) Idiap Research Institute, Clinique Romande de Réadaptation

4. Name **ELEARNING-VALAIS-3.0** (eLearning-Valais 3.0)
   Funding Fondation The Ark
   Coordinator Formation Universitaire à Distance
   Duration 2016.03.01 - 2018.01.31
   Partner(s) Idiap Research Institute, Klewel

5. Name **LIFE** (PdG-fatigue)
   Funding Fondation The Ark
   Coordinator Clinique Romande de Réadaptation
   Duration 2016.08.01 - 2017.03.31
   Partner(s) Idiap Research Institute, Life

6. Name **COBALT** (Content Based Call Filtering)
   Funding Fondation The Ark
   Coordinator Idiap Research Institute
   Duration 2017.07.01 - 2018.06.30
   Partner(s) Katia SA

7. Name **NMTBENCHMARK** (Training and Benchmarking Neural MT and ASR Systems for Swiss Languages)
   Funding Industrial Project
   Coordinator Idiap Research Institute
   Duration 2017.01.01 - 2017.12.31
   Partner(s) École Polytechnique Fédérale de LauSAnne

8. Name **SNACK** (Bites'n'Bits: Understanding Eating Routines in Context)
   Funding Industrial Project
   Coordinator École Polytechnique Fédérale de Lausanne
   Duration 2016.03.01 - 2018.05.31
   Partner(s) Idiap Research Institute

9. Name **BBC-BOB** (Adapting bespoke speaker identification models to be used in IDIAP's 'Bob.Spear' speaker recognition software)
   Funding Industrial Project
   Coordinator Idiap Research Institute
   Duration 2017.05.01 - 2017.06.30
   Partner(s)

10. Name — **SWISKO** (Swiss-Korean project to develop and integrate new wearable sensors into the existing DomoSafety ambient sensor system.)
    Funding — CTI
    Coordinator — Idiap Research Institute
    Duration — 2016.01.01 - 2018.02.28
    Partner(s) — DomoSafety S.A., University of Bern, Hes-So Vaud

11. Name — **REGENN** (Robust Eye-Gaze Estimation Deep Neural Network)
    Funding — CTI
    Coordinator — Idiap Research Institute
    Duration — 2017.09.01 - 2018.12.31
    Partner(s) — Eyeware

12. Name — **UNICITY** (3D scene understanding through machine learning to secure entrance zones)
    Funding — CTI
    Coordinator — Idiap Research Institute
    Duration — 2017.03.01 - 2018.11.30
    Partner(s) — Hes-So Fribourg, Fastcom Technology SA

13. Name — **IMIM** (Intelligent Monitoring for In-line Manufacturing)
    Funding — CTI
    Coordinator — AiSA Automation Industrielle SA
    Duration — 2016.07.01 - 2018.05.31
    Partner(s) — Idiap Research Institute

14. Name — **BIOWAVE** (A BIOmetric Watch Activated by VEins)
    Funding — CTI
    Coordinator — Centre Suisse D'Electronique et de Microtechnique
    Duration — 2016.02.01 - 2017.12.31
    Partner(s) — Idiap Research Institute, Biowatch SA

15. Name — **3DFINGERVEIN** (3D FingerVein Biometrics)
    Funding — CTI
    Coordinator — Hes-So Valais
    Duration — 2016.06.01 - 2018.03.31
    Partner(s) — Idiap Research Institute, Itservices Sonna Sàrl

16. Name — **VIEW-2** (Visibility Improvement for Events Webcasting)
    Funding — CTI
    Coordinator — Hes-So Valais
    Duration — 2016.08.01 - 2018.03.31
    Partner(s) — Idiap Research Institute, Klewel

17. Name: **FARGO** (Convenient and Secure 3D Face Recognition based on RGB-D Cameras)
    Funding: CTI
    Coordinator: Idiap Research Institute
    Duration: 2016.05.01 - 2018.02.28
    Partner(s): Keylemon SA

18. Name: **ESGEM** (Enhanced Swiss German mEdia Monitoring)
    Funding: CTI
    Coordinator: Idiap Research Institute
    Duration: 2016.06.01 - 2018.01.31
    Partner(s): Recapp, Argus der Presse AG

19. Name: **MULTIVEO** (High Accuracy Speaker-Independent Multilingual Automatic Speech Recognition System)
    Funding: CTI
    Coordinator: Idiap Research Institute
    Duration: 2013.11.01 - 2018.03.31
    Partner(s): Veovox

4.2 Projects Awarded in 2017 and Starting in 2018

1. Name: **SHAPED** (Speech Hybrid Analytics Platform for consumer and Enterprise Devices)
   Funding: CTI
   Coordinator: Idiap Research Institute
   Duration: 2018.03.01 - 2020.02.29
   Partner(s): Logitech S.A.

2. Name: **RISE** (Rich Interpersonal Skill analytics for rEcruitment)
   Funding: CTI
   Coordinator: Idiap Research Institute
   Duration: 2018.05.01 - 2020.04.30
   Partner(s): University of Lausanne

3. Name: **MEMMO** (Memory of Motion)
   Funding: H2020-RIA-ICT
   Coordinator: Centre national de la recherche scientifique
   Duration: 2018.01.01 - 2021.12.31
   Partner(s): Idiap Research Institute, University of Edinburgh, Max Planck Society for the Advancement of Sciences, University of Oxford, PAL ROBOTICS SL, AIRBUS SAS, Wandercraft, Centre de médecine physique et de réadaptation, Costain Group PLC

4. Name **MPM** (Multimodal People Monitoring)
   Funding Idiap-CSEM Program
   Coordinator Idiap Research Institute
   Duration 2018.02.01 - 2019.01.31
   Partner(s) Centre Suisse d'Electronique et de Microtechnique

5. Name **RAPPS** (Reinforced audio processing via physiological signals)
   Funding Idiap-CSEM Program
   Coordinator Idiap Research Institute
   Duration 2018.03.01 - 2019.02.28
   Partner(s) Centre Suisse d'Electronique et de Microtechnique

6. Name **DOMAT** (On-demand Knowledge for Document-level Machine Translation)
   Funding SNF - Division II
   Coordinator Idiap Research Institute
   Duration 2018.01.01 - 2018.12.31
   Partner(s) HES-SO Vaud

7. Name **ROSALIS** (Robot skills acquisition through active learning and social interaction strategies)
   Funding SNF - Division II
   Coordinator Idiap Research Institute
   Duration 2018.04.01 - 2022.03.31
   Partner(s)

8. Name **SHISSM** (Sparse and hierarchical Structures for Speech Modeling)
   Funding SNF - Division II
   Coordinator Idiap Research Institute
   Duration 2018.01.01 - 2021.12.31
   Partner(s)

5 List of Publications in 2017

5.1 Books and Book Chapters

[1] D. Gatica-Perez, O. Aran, and D. B. Jayagopi, "Analysis of small groups," in *Social Signal Processing*, Cambridge University Press. Editors J. Burgoon, N. Magnenat-Thalmann, M. Pantic, and A. Vinciarelli, 2017, pp. 349–367.

[2] D. Gatica-Perez, G. Can, R. Hu, S. Marchand-Maillet, J.-M. Odobez, C. Pallan Gayol, and E. Roman-Rangel, "Maaya: Multimedia methods to support maya epigraphic analysis," in *Arqueologia computacional: Nuevos enfoques para el analisis y la difusion del patrimonio cultural*, D. Jimenez-Badillo, Ed., INAH-RedTDPC, 2017.

[3] P. Korshunov and S. Marcel, "Presentation attack detection in voice biometrics," in *User-Centric Privacy and Security in Biometrics*, C. Vielhauer, Ed., Savoy Place, London WC2R 0BL, UK: The Institution of Engineering and Technology, 2017, ch. 7.

5.2 Articles in Scientific Journals

[1] Z. Akhtar, A. Hadid, M. Nixon, M. Tistarelli, J.-L. Dugelay, and S. Marcel, "Biometrics: In search of identity and security (q and a)," *IEEE MultiMedia*, vol. PP, 2017.

[2] A. Asaei, M. Cernak, and H. Bourlard, "Perceptual information loss due to impaired speech production," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2017.

[3] D. Azzimonti and D. Ginsbourger, "Estimating orthant probabilities of high dimensional gaussian vectors with an application to set estimation," *Journal of Computational and Graphical Statistics*, 2017.

[4] J.-I. Biel, N. Martin, D. Labbe, and D. Gatica-Perez, "Bites'n'bits: Inferring eating behavior from contextual mobile data," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (PACM IMWUT)*, vol. 1, no. 4, pp. 125–157, Dec. 2017, article 125.

[5] D. Bruno, S. Calinon, and D. G. Caldwell, "Learning autonomous behaviours for the body of a flexible surgical robot," *Autonomous Robots*, vol. 41, no. 2, pp. 333–347, Feb. 2017.

[6] A. Cerekovic, O. Aran, and D. Gatica-Perez, "Rapport with virtual agents: What do human social cues and personality explain?" *IEEE Transactions on Affective Computing*, vol. 8, no. 3, pp. 382–395, Jul. 2017.

[7] M. Cernak, A. Asaei, and A. Hyafil, "Cognitive speech coding," *IEEE Signal Processing Magazine*, 2017.

[8] M. Cernak, J. R. Orozco-Arroyave, F. Rudzicz, H. Christensen, J. C. Vasquez-Correa, and E. Nöth, "Characterisation of voice quality of parkinson's disease using differential phonological posterior features," *Computer Speech and Language*, 2017.

[9] K. G. Chan and M. Liebling, "Direct inversion algorithm for focal plane scanning optical projection tomography," *Biomedical Optics Express*, 2017.

[10] S. Dey, P. Motlicek, S. Madikeri, and M. Ferras, "Template-matching for text-dependent speaker verification," *Speech Communication*, Apr. 2017.

[11] G. Heusch, A. Anjos, and S. Marcel, "A reproducible study on remote heart rate measurement," *ArXiv*, Sep. 2017.

[12] R. Hu, J.-M. Odobez, and D. Gatica-Perez, "Extracting maya glyphs from degraded ancient documents via image segmentation," *Journal on Computing and Cultural Heritage*, vol. 10, Apr. 2017.

[13] R. Hu, C. Pallan Gayol, J.-M. Odobez, and D. Gatica-Perez, "Analyzing and visualizing ancient maya hieroglyphics using shape: From computer vision to digital humanities," *Digital Scholarship in the Humanities*, vol. 32, pp. 179–194, Dec. 2017.

[14] N. Jaquier, M. Connan, C. Castellini, and S. Calinon, "Combining electromyography and tactile myography to improve hand and wrist activity detection in prostheses," *Technologies*, vol. 5, no. 4, Oct. 2017.

[15] P. Korshunov and S. Marcel, "Impact of score fusion on voice biometrics and presentation attack detection in cross-database evaluations," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 4, pp. 695 –705, Jun. 2017.

[16] F. Labhart, F. Tarsetti, O. Bornet, D. Santani, J. Truong, S. Landolt, D. Gatica-Perez, and E. Kuntsche, "Development of the geographical proportional-to-size street-intercept sampling (gp-sis) method for recruiting urban nightlife-goers in an entire city," *International Journal of Social Research Methodology*, vol. 20, no. 6, pp. 721–736, 2017.

[17] R. Lefort, L. Fusco, O. Pertz, and F. Fleuret, "Machine learning-based tools to model and to remove the off-target effect," *Pattern Analysis and Applications*, vol. 20, no. 1, pp. 87–100, Feb. 2017, first online: 1st April 2015.

[18] N. Linde, D. Ginsbourger, J. Iriving, F. Nobile, and A. Doucet, "On uncertainty quantification in hydrogeology and hydrogeophysics," *Advances in Water Resources*, vol. 110, 166–181, Dec. 2017.

[19] E. Madonna, D. Ginsbourger, and O. Martius, "A poisson regression approach to model monthly hail occurrence in northern switzerland using large-scale environmental variables," *Atmospheric Research*, 2017.

[20] A. Mohammadi, S. Bhattacharjee, and S. Marcel, "Deeply vulnerable – a study of the robustness of face recognition to presentation attacks," *IET (The Institution of Engineering and Technology) – Biometrics*, pp. 1–13, 2017, Accepted on 29-Sept-2017.

[21] H. Muckenhirn, P. Korshunov, M. Magimai.-Doss, and S. Marcel, "Long-term spectral statistics for voice presentation attack detection," *IEEE/ACM Transactions on Audio, Speech and Language Processing*, vol. 25, no. 11, pp. 2098–2111, Nov. 2017.

[22] L. S. Nguyen, S. Ruiz-Correa, M. Schmid Mast, and D. Gatica-Perez, "Check out this place: Inferring ambiance from airbnb photos," *IEEE transactions on Multimedia*, vol. PP, no. 99, Nov. 2017.

[23] J. R. Orozco-Arroyave, J. C. Vasquez-Correa, J. F. Vargas-Bonilla, R. Arora, N. Dehak, P. S. Nidadavolu, H. Christensen, F. Rudzicz, M. Yancheva, A. Vann, N. Vogler, T. Bocklet, M. Cernak, J. Hannink, and E. Nöth, "Neurospeech: An open-source software for parkinson's speech analysis," *Digital Signal Processing*, 2017.

[24] N. Pappas and A. Popescu-Belis, "Explicit document modeling through weighted multiple-instance learning," *Journal of Artificial Intelligence Research*, vol. 58, pp. 591–626, 2017.

[25] N. Pappas, M. Redi, M. Topkara, H. Liu, B. Jou, T. Chen, and S.-F. Chang, "Multilingual visual sentiment concept clustering and analysis," *International Journal of Multimedia Information Retrieval*, 2017.

[26] E. Pignat and S. Calinon, "Learning adaptive dressing assistance from human demonstration," *Robotics and Autonomous Systems*, vol. 93, pp. 61–75, Jul. 2017.

[27] M. Razavi and M. Magimai.-Doss, "A posterior-based multi-stream formulation for g2p conversion," *IEEE Signal Processing Letters*, 2017.

[28] R. Reiter-Palmon, T. Sinha, J. Gevers, J.-M. Odobez, and G. Volpe, "Theories and models of teams and group," *Small Group Research*, vol. 48, no. 5, pp. 544–567, Oct. 2017.

[29] S. Ruiz-Correa, D. Santani, B. Ramirez Salazar, I. Ruiz-Correa, F. Rendon-Huerta, C. Olmos Carrillo, B. Sandoval Mexicano, A. Arcos Garcia, R. Hasimoto-Beltran, and D. Gatica-Perez, "Sensecityvity: Mobile crowdsourcing, urban awareness, and collective action in mexico," *IEEE Pervasive Computingg, Special Issue on Smart Cities*, vol. 16, no. 2, pp. 44–53, 2017.

[30] M. Zeestraten, I. Havoutis, J. Silverio, S. Calinon, and D. G. Caldwell, "An approach for imitation learning on riemannian manifolds," *IEEE Robotics and Automation Letters (RA-L)*, vol. 2, no. 3, pp. 1240–1247, Jun. 2017.

5.3 PhD Theses

[1] G. Can, "Visual analysis of maya glyphs via crowdsourcing and deep learning," PhD thesis, École Polytechnique Fédérale de Lausanne, Sep. 2017.

[2] O. Canévet, "Object detection with active sample harvesting," PhD thesis, École Polytechnique Fédérale de Lausanne, Feb. 2017.

[3] P.-E. Honnet, "Intonation modelling for speech synthesis and emphasis preservation," PhD thesis, École Polytechnique Fédérale de Lausanne, Jan. 2017.

[4] P. H. O. Pinheiro, "Large-scale image segmentation with convolutional networks," PhD thesis, Sciences et Techniques de l'Ingénieur (STI), Jan. 2017.

[5] M. Razavi, "On modeling the synergy between acoustic and lexical information for pronunciation lexicon development," PhD thesis, École polytechnique fédérale de Lausanne (EPFL), 2017.

5.4 Articles in Conference Proceedings

[1] S. Abbasi-Sureshjani, B. Dasht Bozorg, B. ter Haar Romeny, and F. Fleuret, "Boosted exudate segmentation in retinal images using residual nets," in *Proceedings of the MICCAI Workshop on Ophthalmic Medical Image Analysis*, 2017.

[2] S. Abbasi-Sureshjani, B. Dasht Bozorg, B. ter Haar Romeny, and F. Fleuret, "Exploratory study on direct prediction of diabetes using deep residual networks," in *Proceedings of the thematic conference on computational vision and medical image processing*, 2017.

[3] A. Anjos, L. El Shafey, and S. Marcel, "Beat: An open-science web platform," in *Thirty-fourth International Conference on Machine Learning*, https://openreview.net/group?id=ICML.cc/2017/RML, Sydney, Australia, Aug. 2017.

[4] A. Anjos, M. Günther, T. de Freitas Pereira, P. Korshunov, A. Mohammadi, and S. Marcel, "Continuously reproducing toolchains in pattern recognition and machine learning experiments," in *Thirty-fourth International Conference on Machine Learning*, https://openreview.net/group?id=ICML.cc/2017/RM Sidney, Australia, Aug. 2017.

[5] A. Asaei, M. Cernak, H. Bourlard, and D. Ram, "Sparse pronunciation codes for perceptual phonetic information assessment," in *Workshop on Signal Processing with Adaptive Sparse Structured Representations (SPARS)*, Proceeding of Abstracts for Communication, 2017.

[6] T. Bagautdinov, A. Alahi, F. Fleuret, P. Fua, and S. Savarese, "Social scene understanding: End-to-end multi-person action localization and collective activity recognition," in *Proceedings of the IEEE international conference on Computer Vision and Pattern Recognition*, 2017.

[7] P. Baqué, F. Fleuret, and P. Fua, "Deep occlusion reasoning for multi-camera multi-target detection," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017.

[8] P. Baqué, F. Fleuret, and P. Fua, "Multi-modal mean-fields via cardinality-based clamping," in *Proceedings of the IEEE international conference on Computer Vision and Pattern Recognition*, 2017.

[9] Y. Benkhedda, D. Santani, and D. Gatica-Perez, "Venues in social media: Examining ambiance perception through scene semantics," in *Proceedings of the 25th ACM International Conference on Multimedia, ACM, 2017*, 2017.

[10] D. Berio, S. Calinon, and F. F. Leymarie, "Dynamic graffiti stylisation with stochastic optimal control," in *Intl Workshop on movement and computing (MOCO)*, London, UK: ACM, Jun. 2017, pp. 1–8.

[11] D. Berio, S. Calinon, and F. F. Leymarie, "Generating calligraphic trajectories with model predictive control," in *Proc. 43rd Conf. on Graphics Interface*, Edmonton, AL, Canada, May 2017, pp. 132–139.

[12] S. Bhattacharjee and S. Marcel, "What you can't see can help you – extended-range imaging for 3d-mask presentation attack detection," in *Proceedings of the 16th International Conference on Biometrics Special Interest Group.*, Darmstadt (Germany): Gesellschaft fuer Informatik e.V. (GI), 2017.

[13] Z. Boulkenafet, J. Komulainen, Z. Akhtar, A. Benlamoudi, S. Bekhouche, F. Dornaika, A. Ouafi, A. Mohammadi, S. Bhattacharjee, and S. Marcel, "A competition on generalized software-based face presentation attack detection in mobile scenarios," in *Proceedings of the International Joint Conference on Biometrics, 2017*, Oct. 2017.

[14] G. Can, J.-M. Odobez, and D. Gatica-Perez, "Shape representations for maya codical glyphs: Knowledge-driven or deep?" In *15th International Workshop on Content-Based Multimedia Indexing*, Jun. 2017.

[15] M. Cernak, A. Komaty, A. Mohammadi, A. Anjos, and S. Marcel, "Bob speaks kaldi," in *Proc. of Interspeech*, Aug. 2017.

[16] M. Cernak, E. Nöth, F. Rudzicz, H. Christensen, J. R. Orozco-Arroyave, R. Arora, T. Bocklet, H. Chinaei, J. Hannink, P. S. Nidadavolu, J. C. Vasquez, M. Yancheva, A. Vann, and N. Vogler, "On the impact of non-modal phonation on phonological features," in *Proceedings of 2017 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2017)*, 2017.

[17] T. Chavdarova and F. Fleuret, "Deep multi-camera people detection," in *Proceedings of the IEEE International Conference on Machine Learning and Applications*, 2017.

[18] S. Dey, S. Madikeri, P. Motlicek, and M. Ferras, "Content normalization for text-dependent speaker verification," in *Proc. of Interspeech*, 2017.

[19] S. Dey, P. Motlicek, S. Madikeri, and M. Ferras, "Exploiting sequence information for text-dependent speaker verification," in *Proceedings of 2017 IEEE International Conference on Acoustics, Speech, and Signal Processing*, New Orleans, Mar. 2017, pp. 5370–5374.

[20] P. Dighe, A. Asaei, and H. Bourlard, "Exploiting eigenposteriors for semi-supervised training of dnn acoustic models with sequence discrimination," in *Proceedings of Interspeech*, 2017.

[21] P. Dighe, A. Asaei, and H. Bourlard, "Low-rank and sparse soft targets to learn better dnn acoustic models," in *Proceedings of 2017 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2017)*, 2017.

[22] A. R. Goncalves, P. Korshunov, R. P. V. Violato, F. O. Simões, and S. Marcel, "On the generalization of fused systems in voice presentation attack detection," in *16th International Conference of the Biometrics Special Interest Group*, A. Brömme, C. Busch, A. Dantcheva, C. Rathgeb, and A. Uhl, Eds., Darmstadt, Germany, Sep. 2017.

[23] I. Havoutis and S. Calinon, "Supervisory teleoperation with online learning and optimal control," in *Proc. IEEE Intl Conf. on Robotics and Automation (ICRA)*, Singapore: IEEE, May 2017, pp. 1534–1540.

[24] R. Hu, H. Pham, P. Buluschek, and D. Gatica-Perez, “Elderly people living alone: Detecting home visits with ambient and wearable sensing,” in *In Proceedings of MMHealth*, Oct. 2017.

[25] N. Jaquier and S. Calinon, “Gaussian mixture regression on symmetric positive definite matrices manifolds: Application to wrist motion estimation with semg,” in *Proc. IEEE/RSJ Intl Conf. on Intelligent Robots and Systems*, Sep. 2017.

[26] N. Jaquier, C. Castellini, and S. Calinon, “Improving hand and wrist activity detection using tactile sensors and tensor regression methods on riemannian manifolds,” in *Proc. of the Myoelectric Control Symposium*, Aug. 2017.

[27] K. Khelif, yann Mombrun, G. Backfried, F. Sahito, L. Scarpatto, P. Motlicek, D. Kelly, G. Hazzani, E. Chatzigavriil, and S. Madikeri, “Towards a breakthrough speaker identification approach for law enforcement agencies: Siip,” in *European Intelligence and Security Informatics Conference (EISIC) 2017*, Athenes, Greece: IEEE Computer Society, Sep. 2017, pp. 32–39.

[28] M. Kleinert, H. Helmke, G. Siol, h. Ehr, M. Finke, Y. Oualil, and A. Srinivasamurthy, “Machine learning of controller command prediction models from recorded radar data and controller speech utterances,” in *Proceedings of the 7th SESAR Innovation Days (SID)*, University of Belgrade, Belgrade, Serbia, Nov. 2017.

[29] N. Le and J.-M. Odobez, “A domain adaptation approach to improve speaker turn embedding using face representation,” in *ACM International Conference on Multimodal Interaction*, Glasgow, Scotland: ACM, Nov. 2017.

[30] N. Le and J.-M. Odobez, “Improving speaker turn embedding by crossmodal transfer learning from face embedding,” in *ICCV Workshop on Computer Vision for Audio-Visual Media*, Oct. 2017.

[31] N. Le, J.-M. Odobez, and et al., “Towards large scale multimedia indexing: A case study on person discovery in broadcast news,” in *15th International Workshop on Content-Based Multimedia Indexing*, Jun. 2017.

[32] R. Liepins and et al., “The summa platform prototype,” in *Proceedings of the EACL 2017 Software Demonstrations*, Valencia, Spain, Apr. 2017, pp. 116–119.

[33] N.-Q. Luong and A. Popescu-Belis, “Machine translation of spanish personal and possessive pronouns using anaphora probabilities,” in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, Valencia, Spain: Association for Computational Linguistics, 2017.

[34] S. Madikeri, M. Ferras, P. Motlicek, and S. Dey, “Intra-class covariance adaptation in plda backends for speaker verification,” in *Proceedings of International Conference on Acoustics, Speech and Signal Processing*, 2017, pp. 5365–5369.

[35] A. Maksai, X. Wang, F. Fleuret, and P. Fua, “Non-markovian globally consistent multi-object tracking,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2017.

[36] S. Marmin, J. Baccou, J. Liandrat, and D. Ginsbourger, “Non-parametric warping via local scale estimation for non-stationary gaussian process modelling,” in *Wavelets and Sparsity XVII*, ser. Proc. SPIE, vol. 10394, International Society for Optics and Photonics, 2017, p. 1 039 421.

[37] S. Marmin, J. Baccou, F. Perales, D. Ginsbourger, and J. Liandrat, “Planification adaptative d’expériences numériques par paquets en contexte non stationnaire pour une étude de fissuration mécanique,” in *23ème Congrès Français de Mécanique, 28 août - 1er septembre 2017, Lille, France (FR), AFM*, 2017.

[38] C. Mastalli, M. Focchi, I. Havoutis, A. Radulescu, S. Calinon, J. Buchli, D. G. Caldwell, and C. Semini, “Trajectory and foothold optimization using low-dimensional models for rough terrain locomotion,” in *Proc. IEEE Intl Conf. on Robotics and Automation (ICRA)*, Singapore: IEEE, May 2017, pp. 1096–1103.

[39] L. Miculicich Werlen and A. Popescu-Belis, "Using coreference links to improve spanish-to-english machine translation," in *Proceedings of the 2nd Workshop on Coreference Resolution Beyond OntoNotes (CORBON 2017)*, Valencia, Spain: Association for Computational Linguistics, Apr. 2017, pp. 30–40.

[40] L. Miculicich Werlen and A. Popescu-Belis, "Validation of an automatic metric for the accuracy of pronoun translation (apt)," in *Proceedings of the Third Workshop on Discourse in Machine Translation (DiscoMT)*, Denmark, Copenhagen: Association for Computational Linguistics (ACL), 2017.

[41] H. Muckenhirn, M. Magimai.-Doss, and S. Marcel, "End-to-end convolutional neural network-based voice presentation attack detection," in *International Joint Conference on Biometrics*, Denver, Colorado, USA, 2017.

[42] S. Muralidhar, "On job training: Automated interpersonal behavior assessment & real-time feedback," in *Proceedings of the 25th ACM International Conference on Multimedia, 2017*, 2017.

[43] S. Muralidhar and D. Gatica-Perez, "Examining linguistic content and skill impression structure for job interview analytics in hospitality," in *Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia*, 2017.

[44] S. Muralidhar, M. Schmid Mast, and D. Gatica-Perez, "How may i help you? behavior and impressions in hospitality service encounters," in *Proceddings of 19th ACM International Conference on Multimodal Interaction*, 2017.

[45] J. Newling and F. Fleuret, "A sub-quadratic exact medoid algorithm," in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, 2017.

[46] J. Newling and F. Fleuret, "K-medoids for k-means seeding," in *Proceedings of the international conference on Neural Information Processing Systems*, 2017.

[47] Y. Oualil, D. Klakow, G. Szaszak, A. Srinivasamurthy, H. Helmke, and P. Motlicek, "A context-aware speech recognition and understanding system for air traffic control domain," in *Proceedings of the IEEE Automatic Speech Recognition and Understanding Workshop*, Okinawa, Japan, Dec. 2017.

[48] N. Pappas and A. Popescu-Belis, "Multilingual hierarchical attention networks for document classification," in *8th International Joint Conference on Natural Language Processing (IJCNLP)*, Dec. 2017, pp. 1015–1025.

[49] X. Pu, L. Mascarell, and A. Popescu-Belis, "Consistent translation of repeated nouns using syntactic and semantic cues," in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, Valencia, Spain: Association for Computational Linguistics, Apr. 2017.

[50] X. Pu, N. Pappas, and A. Popescu-Belis, "Sense-aware statistical machine translation using adaptive context-dependent clustering," in *Proceedings of Second Conference on Machine Translation (WMT17)*, Sep. 2017.

[51] D. Ram, A. Asaei, and H. Bourlard, "Subspace regularized dynamic time warping for spoken query detection," in *Workshop on Signal Processing with Adaptive Sparse Structured Representations (SPARS)*, Proceeding of Abstracts, 2017.

[52] L. Rozo, N. Jaquier, S. Calinon, and D. G. Caldwell, "Learning manipulability ellipsoids for task compatibility in robot manipulation," in *Proc. IEEE/RSJ Intl Conf. on Intelligent Robots and Systems*, Sep. 2017.

[53] D. Santani, S. Ruiz-Correa, and D. Gatica-Perez, "Insiders and outsiders: Comparing urban impressions between population groups," in *International Conference on Multimedia Retrieval*, ACM, May 2017.

[54] A. Sequeira, L. Chen, J. Ferryman, P. Wild, F. Alonso-Fernandez, J. Bigün, K. B. Raja, R. Raghavendra, C. Busch, T. de Freitas Pereira, S. Marcel, S. S. Behera, M. Gour, and V. Kanhangad, "Cross-eyed 2017: Cross-spectral iris/periocular recognition competition.," in *IEEE/IAPR International Joint Conference on Biometrics*, Denver, Colorado, USA: IEEE, Oct. 2017.

[55] R. Siegfried, Y. Yu, and J.-M. Odobez, "Towards the use of social interaction conventions as prior for gaze model adaptation," in *Proceedings of 19th ACM International Conference on Multimodal Interaction*, ACM, Nov. 2017, p. 9.

[56] A. Srinivasamurthy, P. Motlicek, I. Himawan, G. Szaszak, Y. Oualil, and H. Helmke, "Semi-supervised learning with semantic knowledge extraction for improved speech recognition in air traffic control," in *Proceedings of Interspeech 2017*, Stockholm, Sweden, Aug. 2017, pp. 2406–2410.

[57] A. K. Tanwani and S. Calinon, "A generative model for intention recognition and manipulation assistance in teleoperation," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2017.

[58] T. P. Thanh and D. Gatica-Perez, "Healthy fondue dinner: Analysis and inference of food and drink consumption patterns on instagram," in *Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia*, 2017.

[59] S. Tong, P. N. Garner, and H. Bourlard, "An investigation of deep neural networks for multilingual speech recognition training and adaptation," in *Proc. of Interspeech*, Aug. 2017.

[60] S. Tulyakov, A. Ivanov, and F. Fleuret, "Weakly supervised learning of deep metrics for stereo reconstruction," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017.

[61] J. Varadarajan, R Subramanian, N. Ahuja, P. Moulin, and J.-M. Odobez, "Active online anomaly detection using dirichlet process mixture model and gaussian process classification," in *IEEE Winter Conference on Applications of Computer Vision (WACV)*, Washington, Mar. 2017.

[62] J. C. Vasquez-Correa, J. R. Orozco-Arroyave, R. Arora, E. Nöth, N. Dehak, H. Christensen, F. Rudzicz, T. Bocklet, M. Cernak, H. Chinaei, J. Hannink, P. S. Nidadavolu, M. Yancheva, A. Vann, and N. Vogler, "Multi-view representation learning via gcca for multimodal analysis of parkinson's disease," in *Proceedings of 2017 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2017)*, 2017.

[63] Y. Yu, K. A. Funes Mora, and J.-M. Odobez, "Robust and accurate 3d head pose estimation through 3dmm and online head model reconstruction," in *Proceedings of the 12th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2017)*, 2017.

[64] M. Zeestraten, I. Havoutis, S. Calinon, and D. G. Caldwell, "Learning task-space synergies using riemannian geometry," in *Proc. IEEE/RSJ Intl Conf. on Intelligent Robots and Systems (IROS)*, Vancouver, Canada: IEEE, Sep. 2017, pp. 73–78.