



---

# ANNUAL REPORT 2015

---



**Production** Martina Fellay and Léonore Miauton, Project Management Team  
**Drafting** Philippe Morel **Translation and correction** Dave Brooks, ELCS.ch  
**Photographic credits** Céline Ribordy, Sion · Sedrik Nemeth, Sion · Nicolas Sedlatchek, Sion · Idiap, Martigny  
**Graphic Design** Formaz — Andenmatten, Sion **Typographies** Cera · Franziska **Paper** PlanoJet® FSC  
**Printing** MengisDruck AG, Visp **Print run** 1,500 copies

## MAIN PARTNERS

### CITY OF MARTIGNY

### CANTON OF VALAIS

### SWISS CONFEDERATION

The State Secretariat for Education, Research and Innovation (SERI)



[www.loterie.ch](http://www.loterie.ch)



[www.swisscom.ch](http://www.swisscom.ch)



[www.groupemutuel.ch](http://www.groupemutuel.ch)



[www.epfl.ch](http://www.epfl.ch)



[www.theark.ch](http://www.theark.ch)



[www.ideark.ch](http://www.ideark.ch)



[www.snf.ch](http://www.snf.ch)



[www.kti.admin.ch](http://www.kti.admin.ch)



[www.haslerstiftung.ch](http://www.haslerstiftung.ch)



## CONTENTS

### Messages

- 04** Our dynamism: Serving innovation and competitiveness  
Olivier Dumas, President of the Foundation Council, Idiap
- 05** Valais' star is rising—including in Europe's firmament  
Prof. Hervé Bourlard, Director, Idiap

### Organization

- 07** Profile
- 10** Research themes, application domains
- 11** Organization chart
- 12** Research groups
- 14** Foundation Council 2015
- 15** Advisory Board
- 16** Employees

### Finances

- 19** Operating accounts
- 20** Sources of financing/Costs/Comments
- 21** Balance sheet

### Research

- 23** Valais\* Wallis Digital: Digitizing the collective memory of Valais
- 25** Valais+: Smartphones—a large-scale mobile data harvesting tool
- 26** Summa, Mummer, TeSLA: European success
- 28** DexROV: Developing a semi-autonomous underwater robot
- 29** Youth@Night: A study using smartphones and social networks

### Network

- 31** ICC'2015: Immersion over, it's time to take flight
- 33** MultiVEO: Multilingual speech-recognition system for fast-food outlets
- 34** Technology transfer: Helping research to bear fruit

### Faces

- 37** David Ginsbourger, senior researcher:  
Statistics as an interdisciplinary language
- 38** Elisa Bovio, program manager and financial assistant:  
The hidden face of the research project
- 39** Kenneth Funes Mora, postdoc and entrepreneur:  
Follow my gaze
- 40** 25th anniversary: 25 candles on Idiap's cake
- 42** Distinctions
- 43** Theses completed

### Scientific inserts

- II** Idiap research areas: Human and Media Computing
- IV** Scientific Report
- XXVIII** Main projects in progress
- XXXVI** Major publications/conferences



## Our dynamism: Serving innovation and competitiveness

As it approaches its 25th anniversary, to be celebrated in 2016, the Idiap Research Institute is thriving. Despite a downturn in general economic conditions, the institute continues to pursue its development harmoniously, and with dynamism and efficiency.

It is the Foundation Council's task to define the working framework and the missions entrusted to the institute's researchers in order to ensure Idiap's sustainability. Favorable conditions—autonomy, academic partnerships, and a conducive social context—must be in place if the institute is to prosper sustainably in today's tough, competitive international environment. We must offer our scientists the greatest freedom possible, without losing sight of the fact that research must, ultimately, remain useful. While our researchers independence precludes—for the large part—any purely mercantile distortions that would lead to immediate profitability being the only yardstick that counts, they still cannot escape economic reality. Only the creation of wealth in the “real world” makes funding research possible.

It is widely accepted today that the companies that manage to distinguish themselves from their competitors are those that focus on employee training and innovation, both in terms of the added value their products bring and of how they conquer markets, enabling them to offer innovative products and services. Close, both geographically and in approach, to The Ark Foundation—mandated as it is with organizing and coordinating the range of activities necessary for the formation, establishment, growth, and development of companies in the canton—Idiap fosters technology transfer and thus contributes to the creation of many start-ups. The Ark embodies the strategy developed by Valais' authorities to create a true technology park in



**"Idiap not only creates, it also helps companies, many of which are from Valais."**

**Olivier Dumas**

President of the Foundation Council, Idiap

the canton. Its particularity is its multi-site approach: one of its branches—IdeArk SA—is located in Martigny, and is run by François Foglia, Deputy Director of Idiap. IdeArk SA currently hosts 22 active companies.

But Idiap not only creates, it also helps companies, many of which are from Valais. And such partnerships are interwoven with the aim of adapting the results of our research to the specific needs of industry, particularly via projects funded by the Commission for Technology and Innovation (CTI). Many such projects are emerging and are already helping improve manufacturing processes. In this way, the competitiveness of these companies is enhanced.

The financial support Idiap receives from the confederation, the Canton of Valais, and the city of Martigny enables the institute to play an active role in both the establishment and the growth of companies. Idiap's employees can be proud of their contributions to the efforts that our country must undertake if it is to maintain its enviable position in the ranks of economically powerful nations. And I, personally, thank all my colleagues for all the hard work they put in to make this goal a reality.

## Valais' star is rising—including in Europe's firmament

In 2014, dark clouds had gathered on the horizons of Swiss research, heralding a turbulent year to come. Relations with Europe had rapidly become strained after 9 February 2014, following the Swiss people's acceptance of the initiative "Against Mass Immigration". And science was the first to pay the price, to the astonishment of all concerned. The strong Swiss franc and the Swiss National Bank's abandonment of the Swiss franc-euro floor early in 2015 enduringly compromised an economy suddenly less prone to investing in innovation. And for Idiap in particular, it was also time to prove it could reach beyond the legacy of the NCCR IM2—a long-term project (12 years) that had enabled our institution to play in the big league and had granted us a certain profile with the general public.

Despite these bad omens, 2015 proved to be another fine vintage for Idiap; in no small part thanks to the commitment of each and every one of the institute's members of staff, whom I take the opportunity to thank here. Their work has proved productive in every domain, from research projects to the creation of start-ups, not forgetting technology transfer. And Idiap certainly intends to continue along this road, fully exploiting its unique environment and keeping in mind that it is always possible to do better still.

**"2015 proved to be another fine vintage for Idiap."**

**Hervé Bourlard**

Prof., EPFL; Director, Idiap

Projects indeed: while we understandably welcome the launch of new avenues of research, we often forget that those avenues also come to an end one day. As funding is usually available only for a period of three to four years, the project renewal rate is also important. Identifying new employees is also a laborious—though exciting—task. In these respects, 2015 was a success since it saw the launch of 13 new projects. And, more gratifying still, our researchers managed the feat of being awarded funding for no fewer than three new European projects (see pp. 26-27).

And while the international recognition received by Idiap in its chosen research domains is to be applauded, one of the keys to the institute's perpetual hunt for ideas lies in the targeted expansion of our fields of expertise. This involves, notably, the recruitment of new talent. In the last twenty-four months, three new senior researchers joined, and enriched, our ranks.

Despite these successes, we remain conscious of our environment and know that we must continue to strengthen the ties we have forged with our privileged academic partners, whether they be national or international. This is why we continue to work closely with EPFL, and with its EPFL-Valais-Wallis branch, via a joint development plan. Indeed, it is only by working together that those active in the fields of research and development will be able to help Switzerland to rise to the many challenges—both current and future—facing the country. Without the confidence placed in us and the support provided by the confederation, the Canton of Valais, and the city of Martigny for nearly twenty-five years, none of this would have been possible. Another good reason, among many, to thank all three again here.







# organization

06 — 17

# An institute on a human scale, but with an international reputation

With over one hundred employees, and research domains in touch with current challenges, Idiap is committed to scientific progress in the service of humanity.

In 1991, at its foundation, the Idiap Research Institute first established its mission of promoting scientific progress in the service of the well-being of humanity. Today, over twenty years later, society's best interests remain at the heart of the institute's activities.

## A profusion of technological products

The beginning of the 21<sup>st</sup> century has witnessed the arrival of a profusion of new technological tools. On the one hand, these tools have made possible considerable improvements in efficiency and comfort. On the other, they disrupt people's habits, leaving some users impoverished, and others weary of constant system modifications. It is within this context that Idiap works, essentially to improve man-machine relationships, and to optimize human communication.

## A national and international network

In Switzerland, Idiap works mostly with the two federal institutes of technology, and with other universities of applied sciences and arts and cantonal universities at which research takes place. Idiap is currently working on several European projects, and maintains close links with its numerous partners, including in France, England, and Germany. Across the Atlantic, Idiap and the International Computer Science Institute (ICSI) in Berkley, California, share a privileged partnership.

## Structure

- Non-profit foundation
- Independent institution (but connected to EPFL through a joint development plan)
- Specializes in the management of multimedia information and man-machine, multimodal interactions
- Founded in 1991 by:
  - The city of Martigny
  - The State of Valais
  - EPFL
  - The University of Geneva
  - Swisscom

47% public funds



53% research projects



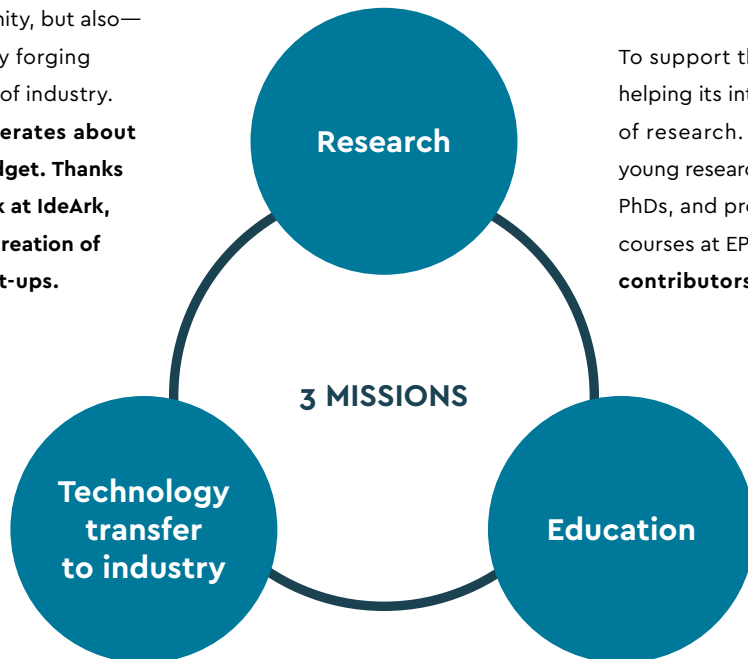
# The three missions of Idiap

The Idiap Research Institute is recognized by the Swiss confederation as part of the strategic domain of the Swiss federal institutes of technology and is supported under the Federal Act on the Promotion of Research and Innovation (RIPA) (article 15). The institute's vision is to promote quality of life through scientific progress made in relation to human and media computing. Idiap's activities are oriented in accordance with three complementary missions.

To conduct fundamental research projects at the highest level in its chosen fields of expertise, thereby assuring its position among the best, nation-, Europe-, and worldwide. On the international stage, Idiap benefits from a broad network of partners and collaborates actively with, among others, public and private research centers, and large universities. **Competitive research projects guarantee around 50 percent of the institute's financing.**

To ensure technology transfer not only by promoting the widest possible dissemination of its research results into the scientific community, but also—and most importantly—by forging close ties with the world of industry. **Technology transfer generates about 10 percent of Idiap's budget. Thanks to the incubator—The Ark at IdeArk, Idiap contributes to the creation of numerous successful start-ups.**

To support the educational process by helping its interns discover the world of research. Idiap welcomes talented young researchers working toward their PhDs, and provides them with numerous courses at EPFL. **One in two scientific contributors is a research assistant.**



# Idiap in figures (2015)

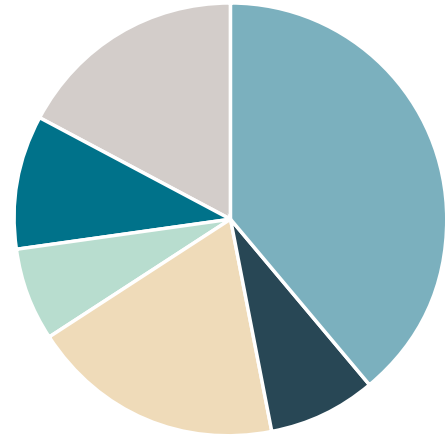
IN TOTAL, 28 NATIONALITIES ARE REPRESENTED AT IDIAP

## Human resources

- 2 professors
- 2 senior scientists (MER)
- 9 permanent senior researchers
- 28 postdocs
- 31 research assistants
- 14 engineers
- 6 system engineers
- 10 administrative staff
- 28 trainees (average/year)
- 2 visitors

**132 people in total**

**99 full-time equivalents**



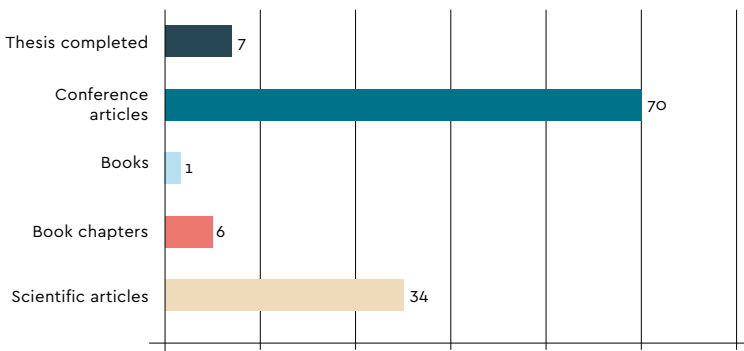
- Switzerland 39%
- India 10%
- France 8%
- Rest of Asia 17%
- Rest of Europe 19%
- Stateless, 3 robots
- America 7%

## Scientific activities

- Participation in fifty-eight research programs
- Project management in more than twenty-five consortia (Switzerland, Europe, industrial projects)
- Participation in the economic development strategy of the Canton of Valais through the program—The Ark and, in particular, the company IdeArk
- More than fifty posts in IdeArk start-ups

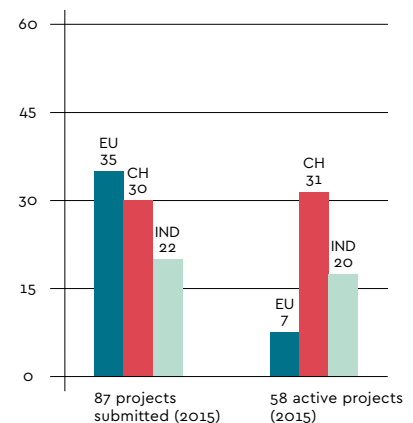
## Submission and financing of research projects

In 2015 Idiap submitted 87 projects (30 to Swiss funding agencies, 35 to European agencies, and 22 industrial projects). At the same time the institute managed 58 active projects.



### Publications

In 2015, Idiap researchers have contributed to 118 peer-reviewed publications.



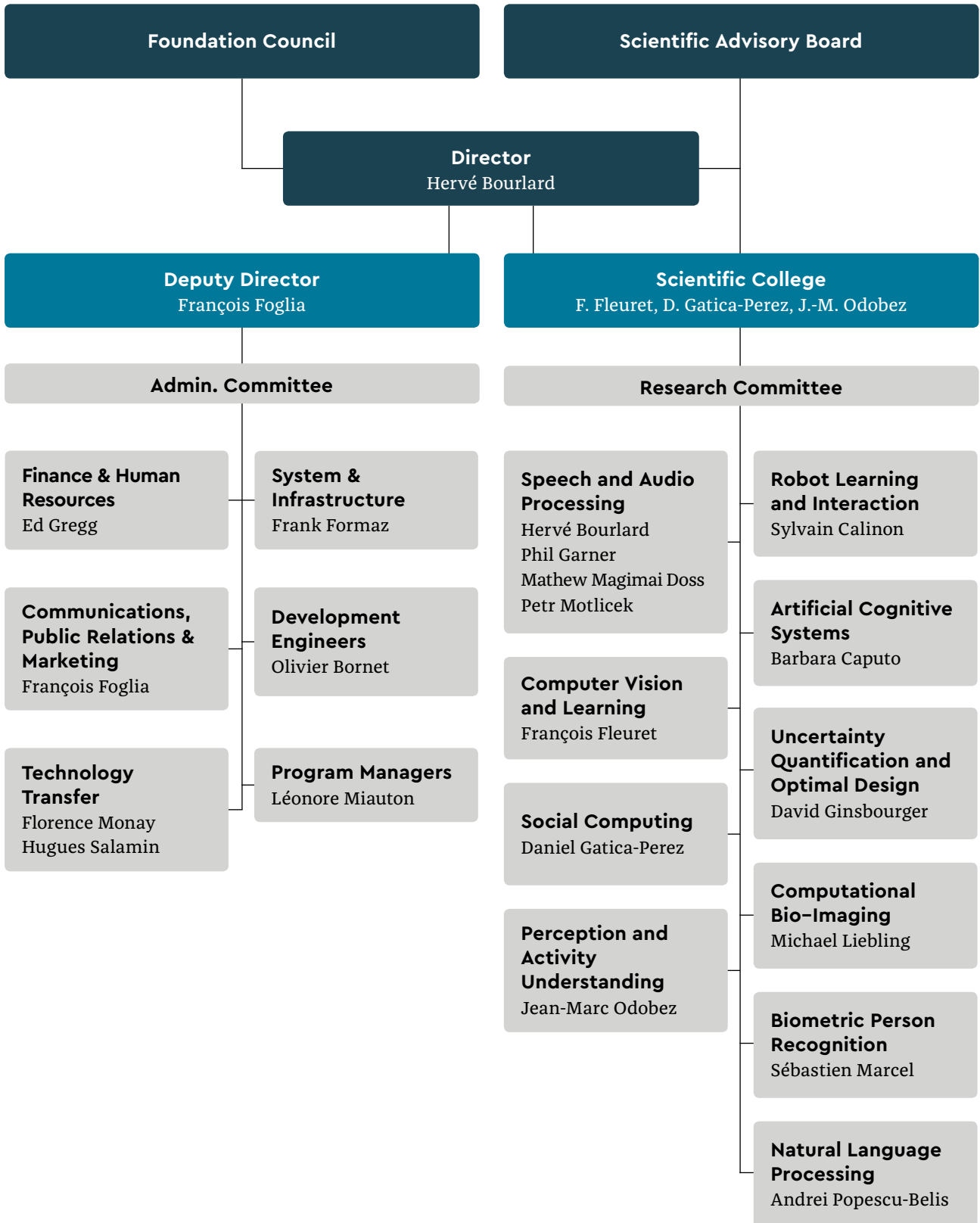
# 5 research teams

## 10 application domains

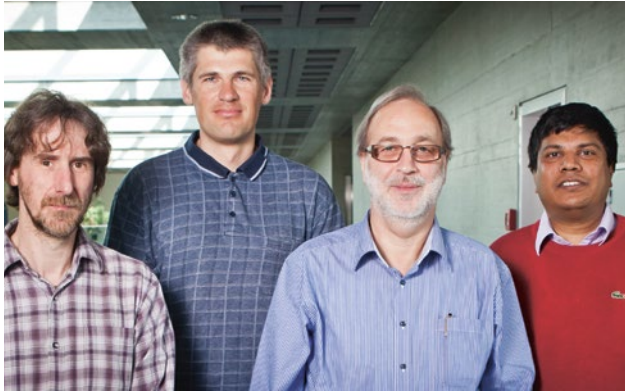




# Organization chart



# Research group



Ten groups, each led by one or more top-level scientists, divide the institute's research activities between them. The group responsible for speech- and sound processing is led by the institute's director—Hervé Bourlard—and three senior researchers. The remaining groups are all headed by senior researchers.

## Speech & Audio Processing

Prof. Hervé Bourlard (second from right),  
 Dr. Phil Garner, Dr. Petr Motlicek, Dr. Mathew Magimai Doss  
 H-indexes\*: HB—56, PG—20, PM—14, MMD—21

Speech processing has been one of the mainstays of Idiap's research portfolio for many years. Today it is still the largest group in the institute, and Idiap continues to be recognized as a leader in the field. The expertise of the group encompasses statistical automatic speech recognition (based on hidden Markov models or hybrid systems exploiting connectionist approaches), text-to-speech, 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).

## Computer Vision & Learning

Dr. François Fleuret  
 H-index: 26



The scientific objective of this group is the development of new statistical learning techniques, mainly for computer vision, with a particular interest in their computational properties. Application domains include image recognition and scene analysis, the tracking of people and biological structures, and the understanding of images in general.

## Social Computing

Prof. Daniel Gatica-Perez  
 H-index: 45



Social computing is an interdisciplinary domain that integrates theories and models from mobile and ubiquitous computing, multimedia, machine learning, and social sciences in order to sense, analyze, and interpret human and social behavior in daily life, and to create devices and systems that support interaction and communication. Current lines of research include ubiquitous sensing of face-to-face interaction, behavioral analysis of social video, crowdsourcing, and urban data-mining using smartphones and mobile social networks.

## Perception & Activity Understanding

Dr. Jean-Marc Odobez  
 H-index: 34



This group conducts research into human-human activity analysis using multimodal data. This entails the investigation of fundamental tasks such as the representation, detection, segmentation, and tracking of objects and people, the characterization of their state, and the modeling of sequential data and the interpretation of that data in the form of gestures, activities, behavior, or social relationships. These investigations take place through the design of principled algorithms that extend models from computer vision, statistical learning, or multimodal signal processing. Surveillance, traffic analysis, analysis of behavior, human-robot interfaces, and multimedia content analysis are the main application domains.



### Robot Learning & Interaction

Dr. Sylvain Calinon  
H-index: 29

The Robot Learning & Interaction group focuses on human-centric robot applications. The scientific objective is to develop probabilistic approaches for encoding movements and behaviors in robots evolving in unconstrained environments. In these applications, the models serve several purposes (recognition, prediction, online synthesis), and are shared by different learning strategies (imitation, emulation, incremental refinement, or exploration). The aim is to facilitate the transfer of skills from end users to robots, or between robots, by exploiting multimodal sensory information and by developing intuitive teaching interfaces.



### Artificial Cognitive Systems

Dr. Barbara Caputo  
H-index: 32

The Artificial Cognitive Systems group works on the development of multimodal learning algorithms to enable artificial agents to act autonomously in realistic settings, with a special emphasis on their ability to—autonomously—detect knowledge gaps and fill them with open-ended learning strategies. The focus of this work is on designing algorithms that are principled and computationally efficient and that provide robust performance in very realistic settings while at the same time providing theoretical guarantees on expected behavior. The group is funded by national and international funding agencies.



### Uncertainty Quantification and Optimal Design

Dr. David Ginsbourger  
H-index: 16

The Uncertainty Quantification and Optimal Design group focuses on quantifying and reducing uncertainties in the context of hi-fidelity models, with a main expertise in Gaussian process methods and the sequential design of computer experiments for optimization, inversion, and related problems. Application domains notably include energy and geosciences, with current collaborations ranging from safety engineering to hydrology and climate sciences.



### Computational Bio Imaging Group

Dr. Michael Liebling  
H-index: 19

This group focuses on research into computational imaging and the 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 to the observation and quantification of complex and live biological systems.



### Biometric Person Recognition

Dr. Sébastien Marcel  
H-index: 32

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



### Natural Language Processing

Dr. Andrei Popescu-Belis  
H-index: 21

The Natural Language Processing group studies how the semantic and pragmatic analysis of texts can improve the execution of two important tasks—machine translation and information retrieval. In particular, the group studies how the recommendation of information from networked multimedia repositories can be improved by analyzing a network's structure and the information content of its nodes.

\*The h-index attempts to quantify the productivity and impact of a given researcher based on the quotation pattern and frequency of that researcher's publications. The higher the index number, the more the citation level is important.



## Foundation council

The Foundation council is responsible for the economic and financial management of Idiap. It defines the institute's structure, appoints its director, and—more generally—ensure Idiap's development and defends its interests.



**Mr. Olivier Dumas, President**  
Independent manager  
and business consultant



**Mr. Patrick Furrer**  
Vice-Chancellor Research and  
Innovation, HES-SO



**Mr. Jean-Daniel Antille, Vice-President**  
Manager of the Regional Office  
for the Economic Development of  
French-Speaking Valais



**Mr. Jean-René Germanier**  
Oenologist, Cave Jean-René Germanier  
Balavaud



**Prof. Karl Aberer, Vice-President**  
Vice-President for Information Systems,  
Ecole Polytechnique Fédérale  
de Lausanne (EPFL)



**Mr. Jordi Montserrat**  
Regional Manager Venturelab



**Mr. Marc-André Berclaz**  
Operational Director of the EPFL  
Valais-Wallis Cluster



**Prof. Christian Pellegrini**  
Senior Member  
of the Foundation Council of Idiap  
Honorary Professor in the Faculty of  
Sciences at the University of Geneva



**Mr. Stefan Bumann**  
Head of  
"Head of the Service  
des Hautes Ecoles (SHE)"



**Mr. Dominique Perruchoud**  
President of the Board of Cimark SA



**Mr. Marc-Henri Favre**  
President of the city of Martigny



**Mr. David Watrin**  
Head of the product unit  
Security & Intelligence at Swisscom

# Advisory board

The Advisory Board is comprised of members of the scientific community selected by Idiap's management for their exceptional skills and avant-garde vision. Although their role is strictly advisory, their advice is frequently sought and proves to be invaluable when making decisions regarding research, training, and technology transfer.



**Dr. Jordan Cohen**  
Independent Consultant,  
SPELAMODE, Half Moon Bay, CA, USA



**Prof. Klaus-Robert Müller**  
Professor of Computer Science, TU Berlin  
Director, Bernstein Focus on Neurotechnology, Berlin, Germany



**Prof. Anil K. Jain**  
Distinguished Professor, Department  
of Computer Science & Engineering  
Michigan State University, USA



**Dr. David Nahamoo**  
Senior Manager, Human Language  
Technologies, IBM Research  
Yorktown Heights, NY, USA



**Dr. John Makhoul**  
Chief Scientist, Speech and Signal  
Processing, BBN Technologies,  
Cambridge, MA, USA



**Prof. Gerhard Sagerer**  
Rector, Bielefeld University  
Germany



**Prof. Kenji Mase**  
Professor, Graduate School  
of Information Science  
Nagoya University, Japan



**Prof. Bernt Schiele**  
Max-Planck-Director, MPI Informatics  
Professor at Saarland University,  
Saarbrücken, Germany



**Prof. Nelson Morgan**  
Deputy Director (and former Director)  
of the International Computer Science  
Institute (ICSI)  
Berkeley, CA, USA



**Prof. Bayya Yegnanarayana**  
Professor and Microsoft Chair,  
International Institute of Information  
Technology (IIIT)  
Hyderabad, India

# Employees

## SCIENTISTS

First name, last name, position, country of origin, residence, year of arrival

Oya Aran Karakus, scientific collaborator, Turkey, 2009  
 Afsaneh Asaei, Postdoc, Iran, 2008  
 Dario Azzimonti, research assistant, Italy, 2015  
 Joan Isaac Biel, postdoc, Spain, 2008  
 Hervé Bourlard, director, Belgium/Switzerland, Saxon, 1996  
 Sylvain Calinon, researcher, Switzerland, Martigny, 2014  
 Gulcan Can, research assistant, Turkey, 2013  
 Olivier Canévet, research assistant, France, 2012  
 Barbara Caputo, senior researcher, Italy, 2005  
 Tatjana Chavdarova, research assistant, Macedonia, 2014  
 Ivana Chingovska, research assistant, Macedonia, 2011  
 Tiago De Freitas, research assistant, Brazil, 2014  
 Subhadeep Dey, research assistant, India, 2014  
 Pranay Dighe, research assistant, India, 2013  
 Trinh-Minh-Tri Do, scientific collaborator, Vietnam, 2009  
 Elie El Khoury, postdoc, Lebanon, 2011  
 Laurent El Shafey, research assistant, France, 2010  
 Marc Ferras Font, scientific collaborator, Spain, 2011  
 François Fleuret, senior researcher, France, 2007  
 Kenneth Funes Mora, postdoc, Costa Rica, 2011  
 Philip Garner, senior researcher, England, 2007  
 Daniel Gatica-Perez, senior researcher, Mexico, 2002  
 Sucheta Ghosh, postdoc, India, 2014  
 David Ginsbourger, senior researcher, France, 2015  
 Manuel Günther, postdoc, Germany, 2012  
 Ioannis Havoutis, postdoc, Greece, 2015  
 Alexandre Heili, postdoc, France, 2010  
 Guillaume Heusch, scientific collaborator, Switzerland, Lausanne, 2015  
 Ivan Himawan, postdoc, Australia, 2014  
 Pierre-Edouard Honnet, research assistant, France, 2012  
 Rui Hu, postdoc, China, 2013  
 David Imseng, postdoc, Switzerland, St. German, 2009  
 Cijo Jose, research assistant, India, 2013  
 Pavel Korshunov, postdoc, Estonia, 2015  
 Tipaluck Krityakierne, postdoc, Thailand, 2015  
 Ilja Kuzborskij, research assistant, Lithuania, 2012  
 Alexandros Lazaridis, scientific collaborator, Greece, 2012  
 Nam Le, research assistant, Vietnam, 2015  
 Rémi Lebret, research assistant, France, 2012  
 Joël Legrand, research assistant, France, 2012  
 Michael Liebling, senior researcher, Switzerland, St. Sulpice, 2014  
 Ngoc-Quang Luong, postdoc, Vietnam, 2014  
 Srikanth Madikeri, postdoc, India, 2013  
 Mathew Magimai Doss, researcher, India, 2007  
 Parvaz Mahdabi, postdoc, Iran, 2014  
 Sébastien Marcel, senior researcher, France/Switzerland, Martigny, 2000  
 Petr Motliceck, researcher, Czech Republic, 2005  
 Hannah Muckenhirn, research assistant, France, 2015  
 Skanda Muralidhar, research assistant, India, 2014  
 James Newling, research assistant, England, 2013  
 Laurent Nguyen, postdoc, Switzerland, Lausanne, 2011  
 Jean-Marc Odobez, senior researcher, France/Switzerland, Clarens, 2001  
 Pedro Oliveira Pinheiro, research assistant, Brazil, 2012  
 Dimitri Palaz, research assistant, Switzerland, Martigny, 2011  
 Nikolaos Pappas, research assistant, Greece, 2012  
 Novi Patricia, research assistant, Indonesia, 2012  
 Trung Phan, research assistant, Vietnam, 2015  
 Andrei Popescu-Belis, senior researcher, France/Romania/Switzerland, Lausanne, 2007  
 Blaise Potard, postdoc, France, 2013  
 Xiao Pu, research assistant, China, 2014  
 André Rabello Dos Anjos, scientific collaborator, Brazil, 2010  
 Dhananjay Ram, research assistant, India, 2014  
 Ramya Rasipuram, research assistant, India, 2010  
 Marzieh Razawi, research assistant, Iran, 2013  
 Lakshmi Saheer, scientific collaborator, India, 2008  
 Darshan Santini, research assistant, India, 2012  
 Ajay Tanwani, research assistant, Pakistan, 2015  
 Pedro Tome, postdoc, Spain, 2014  
 Raphaël Ullman, research assistant, Switzerland, Lausanne, 2012  
 Matthias Vanoni, research assistant, France, 2013  
 Di Wu, postdoc, China, 2015  
 Yu Yu, research assistant, China, 2015

## • Development engineers

Philip Abbet, senior development engineer, Switzerland, Conthey, 2006  
 Olivier Bornet, head of development engineers, Switzerland, Pont de la Morge, 2004  
 Milos Cernak, senior development engineer, Slovakia, 2011



Frédéric Dubouchet, development engineer, Switzerland, Lens, 2014  
 Charles Dubout, development engineer, France, 2015  
 Samuel Gaist, development engineer, Switzerland, Ardon, 2013  
 Salim Kayal, development engineer, Switzerland, Vevey, 2011  
 Vasil Khalidov, senior development engineer, Russia, 2010  
 Andreas Kobler, development engineer, Switzerland, 2015  
 Christine Marcel, development engineer, France/Switzerland,  
 Martigny, 2007  
 Florent Monay, technology transfer officer, Switzerland,  
 Choëx, 2008  
 Alexandre Nanchen, senior development engineer, Switzerland,  
 Martigny, 2008  
 Hugues Salamin, technology transfer officer, Switzerland,  
 Dorénaz, 2014  
 Flavio Tarsetti, senior development engineer, Switzerland,  
 Martigny, 2008

#### ADMINISTRATIVE STAFF

Elisa Bovio, junior program manager and financial assistant,  
 Switzerland, Martigny, 2015  
 Sergio Calabretta, industrial relations, Italy, 2014  
 Antoine Dorsaz, financial assistant, Switzerland, Fully, 2012  
 Christophe Ecoeur, program manager, Switzerland, Collombey, 2010  
 Martina Fellay, program manager, Austria, 2012  
 François Foglia, deputy director, Switzerland, Saxon, 2006  
 Edward-Lee Gregg, financial manager, United States of America, 2004  
 Léonore Miauton, responsible program managers, Switzerland,  
 Chexbres, 2012  
 Sylvie Millius, administrative assistant, Switzerland, Vétroz, 1996  
 Nadine Rousseau, administrative assistant, Belgium/Switzerland,  
 Saxon, 1998

#### • System engineers

Bastien Crettol, system administrator, Switzerland, Sion, 2005  
 Norbert Crettol, system administrator, Switzerland, Martigny, 2002  
 Cédric Dufour, system administrator, Switzerland, Aigle, 2007  
 Frank Formaz, system manager, Switzerland, Fully, 1998  
 Louis-Marie Plumel, system administrator, France, 2011  
 Vincent Spano, webmaster, Switzerland, Martigny-Combe, 2004

#### INTERNS

First name, last name, country of origin, home institution

Idiap interns generally spend between six and ten months at the research institute. Some are students at the Ecole Polytechnique Fédérale de Lausanne (EPFL) and spend time at Idiap as part of their degree studies. Others come on student exchange programs set up with European projects in which Idiap participates.

Christian Abbet, Switzerland, EPFL, Lausanne  
 Afroze Baqapuri, Pakistan, EPFL, Lausanne  
 Daniel Berio, Italy, University of London (GBR)  
 Sushil Bhattacharjee, Switzerland, Talaris Ltd (GBR)  
 Fabian Brix, Germany, EPFL, Lausanne  
 Kevin Chan, United States of America, University of California,  
 Santa Barbara (USA)  
 Gilberto Chávez-Martínez, Mexico, CIMAT, Guanajuato (MEX)  
 Yiqiang Chen, China, INSA Lyon (FRA)  
 Artur Costa Pazo, Spain, Universidade de Vido (ESP)  
 Ailbhe Finnerty, Ireland, University of Trento (ITA)  
 Branislav Gerazov, Macedonia, University of Skopje (MKD)  
 Nicolas Gninenko, Switzerland, EPFL, Lausanne  
 David Guennec, France, Université Rennes 1 (FRA)  
 Maryam Habibi, Iran, research assistant Idiap  
 Serife Kucur Ergunay, Turkey, EPFL, Lausanne  
 Jeevanthi Uthpala Liyanapathirana, Sri Lanka,  
 Copenhagen Business School (DNK)  
 Gil Luyet, Switzerland, University of Fribourg  
 Lesly Miculicich, Peru, University of Fribourg  
 Terry Niederhauser, Switzerland, HES-SO  
 Emmanuel Pignat, Switzerland, EPFL, Lausanne  
 Dairazalia Sanchez-Cortes, Mexico, University Ensenada of Baja  
 California (MEX)  
 Ashtosh Sapru, India, research assistant Idiap  
 Milan Secujski, Serbia, University of Novi Sad (SRB)  
 Alexandre Sierro, Switzerland, HES-SO  
 Dominique Thao, Switzerland, EPFL, Lausanne  
 Julien Theux, Switzerland, Kollegium Brig  
 Bogdan Vlasenko, Ukraine, Otto-von-Guericke Universität,  
 Magdebourg (GER)  
 Sree Harsha Yella, India, research assistant Idiap

#### VISITORS

First name, last name, home institution

Visitors are researchers or industry representatives who only spend a few days or weeks at the institute, some to strengthen inter-institutional relationships and others to get an insight into the work carried out at Idiap.

Nilceu Marana, Brazil, University of Sao Paulo (BRA)  
 Shogo Okada, Japan, Tokyo Institute of Technology (JPN)





# finances

18 — 21



# Profit and loss statement (CHF)

INCOME	2014	2015	%
City of Martigny	700 000	<b>700 000</b>	6.8%
Canton of Valais	1 720 000	<b>1 720 000</b>	16.7%
Swiss Confederation	2 550 000	<b>2 467 300</b>	23.9%
Loterie Romande	50 000	<b>100 000</b>	1%
NCCR IM2 projects	476 648	-	-
Swiss National Science Foundation projects	1 427 889	<b>1 715 249</b>	16.5%
Hasler Foundation projects	371 045	<b>606 689</b>	5.9%
European Commission projects	1 210 890	<b>1 395 595</b>	13.5%
The Ark projects	173 797	<b>225 070</b>	2.2%
CTI projects	607 217	<b>388 170</b>	3.8%
EPFL contribution	72 000	<b>72 000</b>	0.7%
Industrial financing	366 736	<b>538 818</b>	5.2%
Other income / extraordinary income	685 926	<b>396 274</b>	3.8%
<b>TOTAL INCOME</b>	<b>10 412 148</b>	<b>10 325 165</b>	<b>100.0%</b>
<b>EXPENSES</b>			
Personnel expenses	7 879 097	<b>7 884 132</b>	76.4%
Education and travel	355 766	<b>371 774</b>	3.6%
Third-party expenses	302 540	<b>167 114</b>	1.6%
Computer equipment and maintenance	246 284	<b>185 680</b>	1.8%
Administrative costs	233 001	<b>148 148</b>	1.5%
Promotion and communication	50 079	<b>45 468</b>	0.5%
Rent	818 888	<b>791 284</b>	7.7%
Depreciation	310 853	<b>263 690</b>	2.5%
Exceptional expenses	-	-	-
Provisions	200 000	<b>455 000</b>	4.4%
<b>TOTAL EXPENSES</b>	<b>10 396 508</b>	<b>10 312 290</b>	<b>100.0%</b>
<b>OPERATING PROFIT/LOSS</b>	<b>15 640</b>	<b>12 875</b>	<b>0.12%</b>

# Management comments, 2015

Idiap enjoyed an excellent financial period 2015, despite the effect of the abandonment of the exchange-rate floor that had—for over three years—pegged the Swiss franc to the euro. The annual accounts exceed, once again, the symbolic 10 million franc mark.

The proportion of public subsidies continues to respect the strategic goal set by the institute’s direction—to be below 50 percent. In 2015, the figure is approximately 47 percent.

For a number of years now Idiap has focused on developing its relations with industry on cantonal, national, and international levels. And clearly these efforts are bearing fruit. Over 10 percent of the budget is now dedicated to projects closely related to the economy (financing TheArk, CTI, industrial).

Staff costs (salaries plus social charges) make up the bulk of expenses at a little under 80 percent.

Idiap closes its 2015 accounts with a surplus of CHF 12,875.

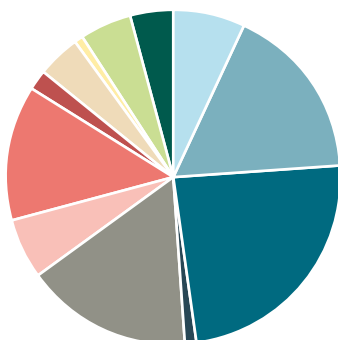
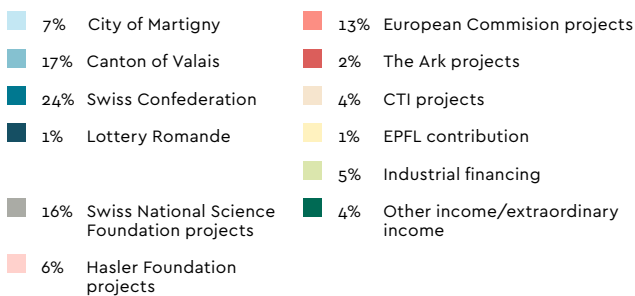
## Federal, cantonal, and municipal subsidies

(In thousands of Swiss francs)

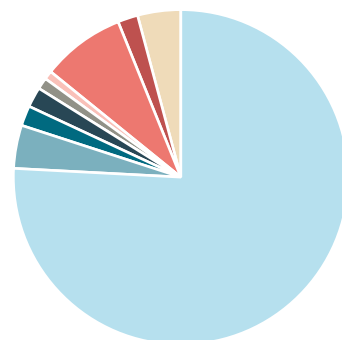
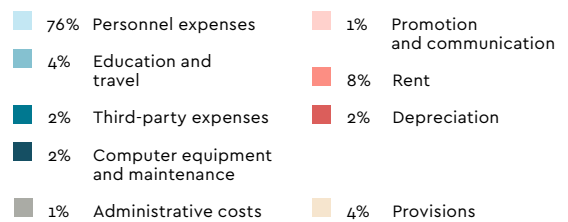
YEAR	2013	2014	2015	2016*
Confederation	2 400	2 550	2 467	2 424
Canton	1 320	1 720	1 720	1 720
Municipality	700	700	700	700

\*Budget

## Distribution of sources of financing



## Distribution of costs





# Balance sheet (CHF)

ASSETS	31.12.2014	31.12.2015
Cash	3 172 241.55	<b>3 467 573.13</b>
Accounts receivable	315 159.11	<b>319 121.25</b>
Accrued income and other	502 413.88	<b>622 013.91</b>
<b>TOTAL CURRENT ASSETS</b>	<b>3 989 814.54</b>	<b>4 408 708.29</b>
Equipment	515 818.64	<b>434 031.64</b>
Financial assets	10 000.00	<b>10 000.00</b>
<b>TOTAL NON-CURRENT ASSETS</b>	<b>525 818.64</b>	<b>444 031.64</b>
<b>TOTAL ASSETS</b>	<b>4 515 633.18</b>	<b>4 852 739.93</b>
<b>LIABILITIES</b>	<b>31.12.2014</b>	<b>31.12.2015</b>
Accounts payable	300 607.00	<b>349 413.85</b>
Accrued expenses	2 058 214.07	<b>1 878 638.93</b>
Provisions	950 000.00	<b>1 305 000.00</b>
<b>TOTAL FOREIGN FUNDS</b>	<b>3 308 821.27</b>	<b>3 533 052.78</b>
Share capital	40 000.00	<b>40 000.00</b>
Special reserve	1 000 000.00	<b>1 100 000.00</b>
Retained earnings	151 171.95	<b>166 812.11</b>
Net income	15 640.16	<b>12 875.04</b>
<b>TOTAL OWN FUNDS</b>	<b>1 206 812.11</b>	<b>1 319 687.15</b>
<b>TOTAL LIABILITIES</b>	<b>4 515 633.18</b>	<b>4 852 739.93</b>



$$p(s|M) | p(x|M)$$

$$h(n): \text{graph of a signal between } -50 \text{ and } 50$$

$$f(n) = g_0 + \sum_{i=1}^K \left( \int_{\tau} g_i(\tau) x_i(n-\tau) d\tau \right)$$

$$f(n) = \sum_{i=-50}^{50} h(i) y(n-i), \quad n: 0 \text{ to } K$$

$$\sum_{n=0}^K f(n) y(n-i)$$

$$\sum_{i=-50}^{50} h(i) y(n-i)$$
  
$$\sum_{j=n-50}^{50} h(j) x(i) + \dots$$

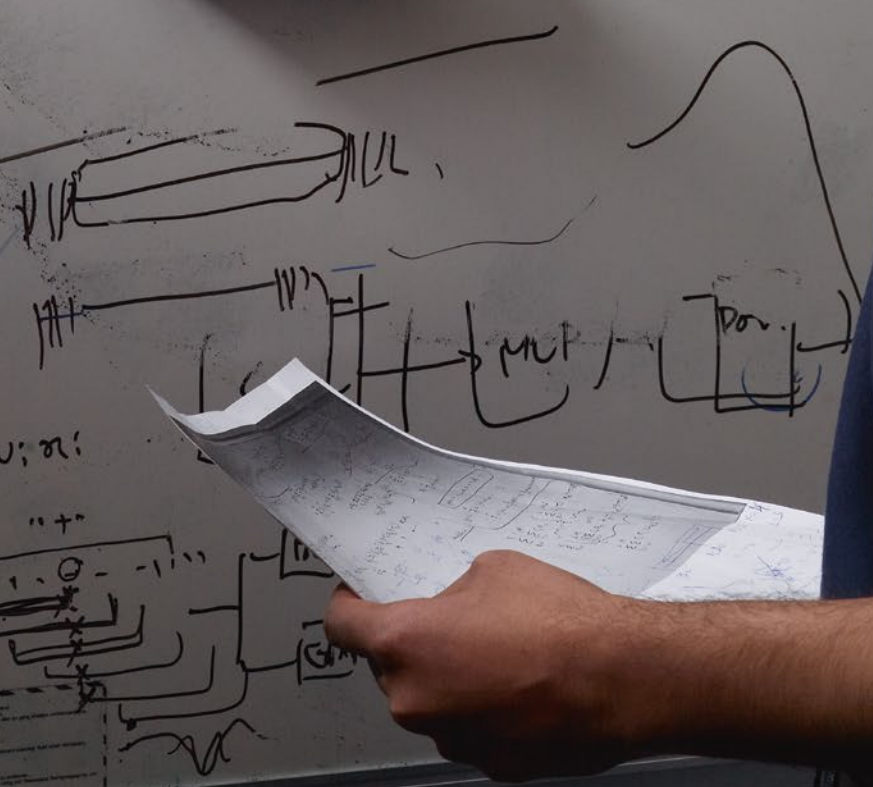
$$x_0 h_{-50} + x_0 h_{-49} \dots x_0 h_0 + x_1 h_1 \dots$$

$$x_1 h_{-50} + \dots + x_0 h_1 + x_1$$

$$\left[ \sum_{-50}^0 h_i + \sum_{-4}^0 \dots \right]$$

# research

22 — 29



# Digitizing the collective memory of Valais

In 2015, the Canton of Valais celebrated the 200th anniversary of its entry into the Swiss confederation. The canton launched a call for ideas—How would it commemorate this bicentenary? A call to which Idiap and its partners responded, proposing a project entitled Valais\*Wallis Digital. The goal? To draw on the power of new technology to digitize the collective memory of the “Old Country”.

“Collective memory” includes, of course, archived documents from public institutions. But it is also made up of the memories of associations and private individuals. Taken in isolation, such memories have—above all—a certain personal worth. But once collected, contextualized, and linked together, they assume their full and true value. Often these private archives are hidden away in the basements or attics of individual citizens. Yet they represent a gold mine to any efforts to illustrate the social and cultural history of a city, canton, or region.

## Collecting and sharing memories

To mark the bicentenary of Valais’ entry into the Swiss confederation, Idiap launched—in collaboration with its partners the Canton of Valais’ multimedia library and cantonal archives, the cooperative Migros Valais, and the EPCA Academy of Contemporary Arts—the project Valais\*Wallis Digital. The project aimed to offer the citizens of Valais a platform via which they could deposit and assemble—in a digital form—their photos, home movie footage, audio files, or written documents of historical value. And that platform obviously would not only be used to collect these archives, but also to share them, making them accessible to one and all, at [www.valais-wallis-digital.ch](http://www.valais-wallis-digital.ch).



**“A project like this goes beyond what we at Idiap do on a daily basis.”**

**François Foglia**

Deputy Director, Idiap

And contributing to safeguarding Valais’ heritage also means contributing to the research into the digital sciences carried out at Idiap. That research involves processing and indexing the large multimedia databases generated by such mass collaborative or participative endeavors. So such a project has an impact that goes far beyond Valais\*Wallis Digital alone, and will help the institute to develop new IT solutions with which to address the major societal challenges of the future, including those in the fields of energy management, healthcare, and the environment.

## Cards to collect

Another aspect of the project was the creation—in Migros “mania” form—of a set of cards illustrating 200 events that marked the history of the canton over the past two centuries. The set covered topics as diverse as agriculture and gastronomy, art, culture and spirituality, the generations, history and tradition, immigration and emigration, innovation and technology,





24

music and media, sports, and tourism. Creating this set of collectable cards was a lengthy undertaking. For François Foglia, Deputy Director of the institute and head of the Valais\*Wallis Digital project, “a project like this goes beyond what we at Idiap do on a daily basis, whether in terms of the choice of illustrations, the size of the cards, the appearance of the wild cards, or the distribution logistics.”

As the moment comes to take stock, Foglia expresses a somewhat subdued sense of satisfaction: “In terms of the platform, the technology itself worked well. But fewer items than we had hoped for were uploaded. We should, perhaps, have focused on specific topics. Our recent call for archive material related to *guggens* and other brass bands (*guggens* are troops of costumed, masked musicians—principally brass and percussion players—that “false play” easily recognizable pieces, creating a strongly rhythmical, cacophonous sound that may be unique to the Swiss, and Swabian-Alemannic carnival season) has, for example, been more successful: the circle of interested parties is smaller, and its members are more inclined to search their archives and share them with other enthusiasts. But, in a scientific sense, this finding has also enabled us to launch new research efforts, including the project Valais+ (see article, p. 25), one component of which addresses the issue of how to encourage the public to participate in the collection of thematic digital data.”



# Valais+: Smartphones— a large-scale mobile data harvesting tool

With its Valais+ project—funded by the Loterie Romande—Idiap aims to provide the Swiss population with an Internet platform that can harvest mobile data on a large scale. Based on the concept of “challenges”, the platform will be designed to meet individuals’ and communities’ data-collection needs. “Challenges” can be set up, customized, and made available on the mobile application Civique.

Smartphones enable their owners to gather images, videos, and texts with great ease and to “geotag” them. But while an individual can collect such data using commonly available applications, sharing and structuring that information is another question entirely.

## An innovative social project

Aided by grants from the Loterie Romande, Professor Daniel Gatica-Perez (head of Idiap’s Social Computing group) decided to embark on an innovative social project, with an eye to placing Idiap’s expertise at the service of the populations of Switzerland in general, and of Valais in particular. To do this, he is leading—alongside Dr. Joan-Isaac Biel and in collaboration with a development group—efforts to create a versatile online platform and a mobile application aimed at helping public authorities, clubs, and associations to comfortably gather whatever information they require.

“Imagine that a municipality wishes to solve a planning problem such as the redevelopment of a park,” Gatica-Perez explains. “It could set up a ‘challenge’ and invite local residents to photograph the elements of the current park that they consider unsatisfactory—a

dilapidated playground, for example. This is the sort of participatory approach that local authorities could put to good use. These geo-tagged data would be made available on a dedicated web page, along with various ways of visualizing them.”

## How to attract potential users

And the creators of a challenge will be its promoters. They share the information and their reasons for creating their challenge so that as many people as possible get involved. A dedicated web page for each challenge and a mechanism to synchronize challenges with the mobile application make sharing and participation simple.

## Cultural heritage as a pilot scheme

Daniel Gatica-Perez aims to begin with a small number of pilot projects. These include the creation of a challenge for the project City-Zen, for which the project’s leader, Professor Maria Sokhn of HES-SO Sierre, was approached. For Sokhn, the information available to tourists with regards to Valais’ cultural heritage is very broad, but can also be quite scattered. With City-Zen, she plans to offer a spatiotemporal search platform that will highlight the canton’s patrimony and any information relating to it. The creation of a Valais+ challenge tied to Valais’ cultural heritage will allow the canton’s citizens to make their own contributions to expanding this knowledge base.

## European success

While the success rate for submissions to the European Community’s research and innovation framework program Horizon 2020 (H2020) is about 14 percent, in 2015 Idiap researchers achieved the feat of securing funding for three of the eleven projects they co-proposed under the section dedicated to information and communication technologies (ICTs). Ranging from entertainment robotics, through biometrics, to automated analysis of multimedia data, these three projects have a combined budget of EUR 22 million, slightly more than CHF 3.8 million of which will come directly to Idiap. This brilliant success contrasts starkly with what had been a rather subdued year, in European terms, 2014.

figures, make links between them, and identify what is truly new. For the project, Idiap researchers led by Professor Hervé Bourlard are working on transcribing and indexing audio content in nine languages. “Here at Idiap, we know how to do this for English, French, and German. But with SUMMA, we’re also dealing with Mandarin, Persian, and even Latvian,” Bourlard explains. The content also has to be linked to news threads on social networks like Facebook or Twitter. For Bourlard, another challenge is the volume of information to be processed: “The project’s media partners—the BBC, Deutsche Welle, Latvian News Agency, and Qatar TV—have asked us to handle 1.5 terabytes (TB) of data, the equivalent of five-hundred hours of audiovisual content, per day.” Coordinated by the University of Edinburgh, the European H2020 project SUMMA will employ 4-5 people at Idiap full-time for three years. [www.summa-project.eu/](http://www.summa-project.eu/)

26



Scalable Understanding  
of Multimodal Media  
(SUMMA)



MultiModal Mall  
Entertainment Robot  
(MuMMER)

At each moment of the day and night, a huge flow of information is arriving in newsrooms across the globe. It is used as the raw material from which new news content is produced—assembling, combining, completing, or commenting on what already exists in order to provide context. In this way, broadcasting footage of an event may lead to articles being written and published on news platforms, which—in turn—may generate tweets that can be taken and reused on that platform, providing a new approach to the subject when it is covered on that evening’s news, perhaps complemented by archive footage. To ensure the traceability of such a multifaceted information thread, the employees of large media organizations spend their days manually indexing the very multimedia content they analyze. An even more complex task given that this must be done in several languages.

The SUMMA project aims to automate this task, making it possible to more rapidly and more effectively manage the emergence of news and the importance of news-related

Developing robots capable of interacting socially with the clients of a large shopping mall is the principal aim of the MuMMER project, led by a consortium of seven institutions and coordinated by the University of Glasgow. The interactions planned are designed firstly to entertain the public—telling stories; coordinating social games such as quizzes—but will also lead to more commercial forms of interaction, such as conducting surveys or proposing specific offers to consumers based on their needs. So, rather than confining itself to waiting for someone to address it, the robot may also adopt more proactive means and approach clients. Such behavior will require the robot to master a social perception of potential clients that is sophisticated enough to enable it to rapidly detect whether such an interaction is desired or not. To complicate matters still further, the interactions will be with groups and will take place in an unconfined environment (the testing phase will take place in a Finnish shopping mall with no fewer than 3,000 stores).

From the Idiap standpoint, the project is in the hands of Dr. Jean-Marc Odobez (in collaboration with Dr. Petr Motlicek for the audio aspect). “In the consortium, we are responsible for the robot’s perception of the people it meets, including localization, speech, and the extraction of non-verbal cues (social signals). In terms of our *savoir faire*, MuMMER is a relatively complex project. Both the robot and the client may be distracted by what is happening around them. When the robot begins to approach an individual, others can pass between the two, forcing the robot to deviate from its goal if its identification system isn’t robust enough. Similarly, a client could—say—rotate their head from left to right and back simply because their attention has been captured by an unrelated (to the robot) audio announcement; but the robot could interpret this as a ‘no’. When working with groups, the robot will need to maintain continuous contact with all of a group’s members and a coherent view of all their activities, despite its limited field of vision, the actions and activities it must maintain—actions that affect its vision and hearing—in order to reproduce social behavior, and the complexity of the interactions it engages in. In general, the robot’s lexicon of perception will have to be much more robust, more comprehensive, and more sophisticated.”

Finally, the project will be an opportunity for Jean-Marc Odobez to meet up with a former colleague from Idiap, Alessandro Vinciarelli—now a researcher at the University of Glasgow and responsible, at MuMMER, for the synthesis of those social signals that members of the public perceive positively.

[www.mummer-project.eu/](http://www.mummer-project.eu/)



**Adaptive, trust-based e-assessment online at different learning stages (TeSLA)**

ICTs offer an opportunity to overcome the physical and financial limitations imposed by traditional teaching models. But changing this paradigm, in which students and educators interact in person, requires new identification standards, especially with regards to sitting for examinations. How can we ensure—in a simple, reliable manner—that it truly is the person registered for an examination that sits for the test and demonstrates that they have acquired the requisite knowledge?

This is the central issue addressed by the European project TeSLA, coordinated by the Fundació per a la Universitat Oberta de Catalunya. Thanks to the biometrics expertise acquired by Idiap researchers—notably via FP7 and Swiss National Science Foundation (SNSF) projects, and industrial collaborations, including for example the work they have carried out with Google—the members of the TeSLA consortium invited Idiap, in the person of Dr. Sébastien Marcel, to join the project. At Idiap, TeSLA will employ two full-time researchers for two years and an engineer for six months in order to develop the necessary methods of voice biometrics and co-develop those for facial biometrics. The Idiap researchers will also manage attack detection.

For Marcel, scientific lead at Idiap, “This project is an opportunity to apply those techniques that we have developed in recent years. But the problems that we will inevitably encounter will also no doubt allow us to continue our progress from a fundamental research standpoint.”

[www.tesla-project.eu/](http://www.tesla-project.eu/)

# Developing a semi-autonomous underwater robot

The ocean floor is as mysterious as it is inhospitable. Prevailing conditions make difficult the nevertheless significant number of industrial and scientific activities (biology, geology, and archeology) that take place there. To perform these tasks, divers often prove essential. But in light of the dangers present and of the complexity of the tasks these divers perform, the use of (semi-) autonomous robots opens up new and interesting opportunities. And developing such a robot is exactly what the European DexROV project is aiming to do.

The seabed makes up about two-thirds of the Earth's surface. Yet much of it remains less well known to us still than the surfaces of distant planets. This, of course, is understandable given the dangerous, complex, and aggressive environment created by the ocean floor's extremely low light levels, and pressure, temperature, and salinity conditions.

In a strange paradox, this inhospitable place is packed with infrastructure dedicated to telecommunications, the exploitation of raw materials, and energy production. The installation, maintenance, and dismantling of this infrastructure is an extremely onerous task. Robots are often used for these jobs, but these robots are usually limited to observation missions, and—in the end—divers often have to manually perform the most complex tasks. Yet taking divers down to great depths is a complicated task in itself due to the very long decompression stops required—some lasting up to ten days. And, given the dangers of the trade—with its duties more befitting a mercenary than a manual worker—it is not always easy even to find divers to do the job.

**It is a case of teaching basic sequences that the robot will adapt and assemble itself.**

## Remote control—and its limits

As for remote-controlled robots, their use requires the in situ presence of a team of specialists dedicated to that one specific mission. An alternative would be to control the robots remotely. But this option comes up against a limitation of a physical nature: satellite transmission of the information necessary to control a robot's movements imposes a delay of about one second between sending a command and receiving the return signal, which means that any complex movements have to be sequenced, extending their duration still further.

With this in mind, Dr. Sylvain Calinon's Robot Interaction & Learning group, working within the framework of the European DexROV project, are developing a semi-autonomous robot that will, in fact, be capable of operating autonomously during this one second of latency necessary for the transmission of commands. Having learned to execute various types of movements, the robot will be able to recognize them and either

terminate or prolong them during signal breaks. If a valve needs to be opened, for example, once the rotation of the valve's handwheel has begun, the robot should be able to continue the action until the valve is completely open.

## Gestures and LEGO bricks

A learning period is, however, necessary and this can take place in two ways: the robot can process information via a camera that shows it the movement that needs to be performed, or it can literally be taken "by the hand" and guided—as an ergotherapist would do—in how to perform the gesture in question. "In contrast to an industrial robot, programmed to repetitively perform perfectly identical tasks," explains Sylvain Calinon, the project's scientific lead at Idiap, "this is a case of teaching basic sequences that the robot will adapt and assemble itself in order to create a complete sequence that will never be the same twice."

As Idiap researchers do not have an underwater robot, they are working on an industrial robot—named Baxter—with two articulated arms controlled by force commands. In 2018, at the conclusion of a project begun in March 2015 and coordinated by the Belgian company Space Applications Services, those collaborating on DexROV hope to have developed a working demonstrator that they will have tested down to a depth of 1,300 meters.



# Youth@Night: Using smartphones and social networks to study nightlife and alcohol consumption

Use mobile technologies and social networks and the way they function to study the habits of young people with regard to alcohol and nightlife. This is the aim of the research project Youth@Night, and Professor Daniel Gatica-Perez.

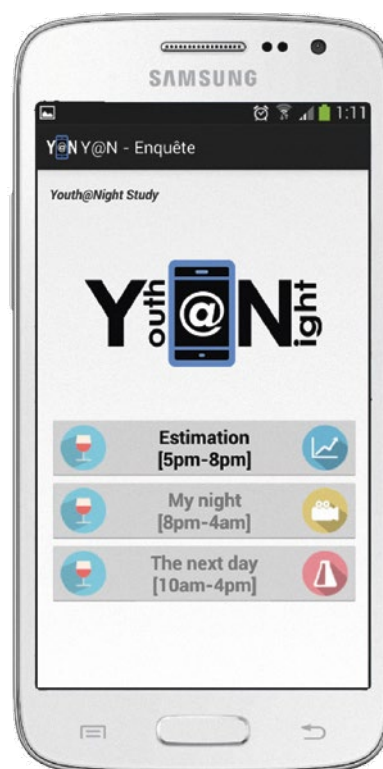
To be effective, a prevention strategy must understand its target with precision. But data on the drinking habits of young adults in Switzerland are limited—they are difficult to obtain and usually come from questionnaires completed retrospectively.

## Understanding the dynamics of the night

Launched by Addiction Switzerland (Addiction Suisse), Idiap's Social Computing research group, and the Department of Geography of the University of Zurich, Youth@Night seeks to better understand the use of public space, to analyze the influence of context on nightlife, and to explore the role urban structure plays in alcohol consumption patterns, at both the individual and the collective level.

## A technology in tune with how we live today

To better understand the problematic and to gather data non-intrusively from study participants, Idiap developed—in 2014—a mobile phone application. “Our idea is to use a technology that’s in tune with what young people between 16 and 25 years of age actually do,” explains Daniel Gatica-Perez, the project’s scientific lead at Idiap. “Using their smartphones and social media networks, young people are accustomed to posting information about their activities and where those activities are taking place.” Throughout the evening, the application—Y@N—requests information on alco-



hol intake and the venues frequented. Participants also have to take pictures of their drinks and capture short videos that indicate the consumption context. “These small glimpses of Swiss nightlife will also form the basis of a series of qualitative interviews that will be conducted later.”

Data from the study participants—a hundred or so in Zurich and another hundred or so in Lausanne—are aggregated and compared to data from the commercially available application “Foursquare”, with which users inform each other with regards to their positions, to ascertain whether common patterns emerge. Both data sets are anonymized—an important point.

“Thanks to Foursquare, we’ve also been able to map the nightlife hot spots of the two cities. This has allowed us to better locate our target audience and to recruit participants.”





**network**

**30 — 35**



# Immersion over, it's time to take flight

The International Create Challenge (ICC), launched by Idiap in 2012, helps young entrepreneurs develop their ideas into commercial prototypes—the conclusion of three weeks of immersion at the institute. In 2015, this unique program was once again a resounding success, honoring three projects likely to improve our daily lives.

From 25 August to 15 September 2015, Idiap welcomed seven teams of aspiring entrepreneurs—scientists, designers, and business hopefuls—hailing from Switzerland, France, Spain, Romania, and even from as far afield as Costa Rica. At the conclusion of ICC'2015, the jury—composed of leading figures from the worlds of research, venture capital, and innovation—chose to honor three projects that combine innovation and high scientific quality: Eyeware (see article, p. 39), Dialexy, and App4Autism.

## EYEWARE\*

### TEAM

- Dr. Kenneth Funes Mora—postdoc at Idiap
- Carlos Becker—PhD candidate in the field of Computer Vision, EPFL
- Serban Mogos—PhD candidate in the field of Technological Change and Entrepreneurship, Carnegie Mellon University (USA)

### GOAL

Develop solutions that give computers and robots the ability to interpret the head and eye movements of a user as commands (move the cursor, click, etc.). Applications using patented eye-tracking technologies are being developed for the fields of health, research, and robotics.

### STATUS

Winner of Arkathon Hacking Health Valais. Recipient of a "Helping Hand" (*Coup de pouce*) grant from the Dr. René Liechti Foundation. Integrated into the program The Ark. Founded the company Eyeware Tech SA, in Martigny.

## DIALEXY

### TEAM

- Mónica Camino Moreno—entrepreneur and certified translator
- Dr. Blaise Potard—postdoc at Idiap and at CereProc Ltd. (Edinburgh, Scotland)
- Albert Solà—senior web developer at Lyles Sutherland (Edinburgh, Scotland)

### GOAL

Development of machine learning algorithms with the goal of improving and automating the translation process, guaranteeing a high-quality end result.

### STATUS

Founded the company Dialexy, in Edinburgh (GBR).

## APP4AUTISM

### TEAM

- Cristina Costescu—research assistant, Department of Clinical Psychology and Psychotherapy, Babeş-Bolyai University, Cluj-Napoca (Romania)
- Josef Kalleder—Sales Manager for Romania and the Republic of Moldova, Apator SA, Poland
- Mihai Calinescu—Coordinator of the Video and e-Learning Production Department, and video editor, SC Frontline Solutions SRL, Cluj-Napoca (Romania)

### GOAL

Development of a modern therapeutic and educational tool to facilitate communication between autistic children and between such children and their families and educators.

### STATUS

A start-up is currently being established in Romania.



A number of projects honored during previous editions of the ICC continue to pursue their respective goals—sometimes enjoying significant success along the way.

#### **HORUS**

The project to create a visual assistant for the visually impaired led to the creation of Horus Technology, in Italy. Having been awarded several prestigious prizes (EIT Digital Idea Challenge, Unicredit Startlab, and IBM SmartCamp) the company has attracted investment of USD 900,000 from the American holding company 5Lion.

#### **ANEMOMIND**

This platform for analyzing and optimizing the performance of sailing vessels led to the creation of a company of the same name, located in Ecublens (VD). In addition to various sporting successes, Anemomind was nominated for the METS Design Award at the Marine Equipment Trade Show in Amsterdam, 2015, and has received seed funding CHF 100,000 from Switzerland's Foundation for Technological Innovation.

#### **BIOWATCH\***

This biometric wristwatch, which functions thanks to the wrist's network of veins, won the UBS employees' prize at the worldwide innovation competition the UBS Future of Finance Challenge, 2015. Thanks to funding from the Commission for Technology and Innovation (CTI), the start-up is pursuing its vision and is currently working both with CSEM—the Neuchâtel-based research and technology organization specializing in microtechnologies—and with Idiap.

#### **REMEETING\***

Record and archive meetings and search their content? An idea that earned ReMeeting—now recap IT AG\*—prestigious awards including the grand prize of UC Berkeley's start-up competition, LAUNCH. The young company has been supported by the Hasler Foundation in its efforts to transform its technical innovation into a marketable product.

#### **LUMA7**

The text analysis technology of this young Valais company—used to search the content of text files and understand the relationships between texts' constituent elements—is currently employed by Swiss media companies and by Valais' parliament.

#### **CLOUDGUIDE**

This company created an application that allows smartphones to be used as audio guides when visiting museums. Cloudguide provides museums with the tools necessary for creating their own content and for making it available to visitors via the start-up's mobile app. The idea has already attracted interest from a number of institutions worldwide. The company has also been honored with the Innovation Award of Spanish bank Caja de Ingenieros.

\*Idiap spin-offs



# Multilingual speech-recognition system for fast-food outlets

As part of a project funded by the Commission for Technology and Innovation (CTI), Idiap and the Swiss company Veovox worked together to develop a multilingual speech-recognition system particularly suited to the noisy conditions typical of the average fast-food restaurant.

When the client of a drive-in fast-food outlet places an order, an operator transmits the list of items to the kitchens, where many employees will be involved in preparing the order before it is handed over to the consumer. Employees use their hands to prepare the order, and have to break off and interrupt what they are doing to validate each item as it is added, using a touch interface. Errors are now relatively common, since this validation system is poorly adapted to the prevailing conditions.

## Reducing the number of errors

To overcome this problem, Veovox developed—in collaboration with Idiap—an order management system based on speech recognition. As they fill each bag, employees name, out loud, each item they add to the order. The system checks that the item corresponds to the order as placed, and if it does, that item is deleted from the list. Veovox's solution allows the number of order picking errors in the fast-food sector to be significantly reduced, an improvement that cannot help but boost customer satisfaction. The true centerpiece of the solution—the voice-recognition system—was developed at Idiap, with the support of Dr. Mathew Magimai Doss.

## A little piece of Idiap in your burger

This innovative solution was developed between 2013 and 2015 in the framework of the project MultiVEO, and was supported by the CTI. The project is the logical extension of two previous successful collaborations between the partners concerned. From the outset, tests measured the performance of the voice recognition approach developed by Idiap and Veovox against that of other, commercially available solutions. In order to ensure they were developing the most competitive system, the researchers involved defined scenarios those comprised only a limited list of vocal commands. More important still, the approach applied by Idiap enables the system to adapt quickly to new languages without the need to purchase additional data.

**So, the next time you order a burger, maybe there'll be a little piece of Idiap "inside".**

# Technology transfer: Helping research to bear fruit

In 2015, Dr. Florent Monay and Dr. Hugues Salamin took over the management of Idiap's Technology Transfer Office (TTO). Veteran researchers turned developers, their aim is to offer a second life to their colleagues' findings by reaching out—armed with those findings—to companies that are more than keen to innovate.

In the field of basic research, success is measured in numbers of publications and in the citations such publications generate. This logic of “publish or perish” means that—in career terms—a researcher does not necessarily have much to gain by devoting his or her time to developing a marketable product.

Financed mainly with public money and thus by the taxpayer—including, of course, tax-paying companies—Idiap bet on technology transfer almost from the outset, setting up an office dedicated to the task, the TTO, and maintaining a team of developers, now twelve in number, to ensure that office was a success. The aim? Offer a return on investment to the industries of Switzerland in general, and Valais in particular.

**“It's our researchers who take the first step, our job is then to grow that contact.”**

**Florent Monay and Hugues Salamin**

## Beyond research

Since early 2015, Florent Monay and Hugues Salamin have run Idiap's Technology Transfer Office. In parallel, each also works as a developer at the institute. Issuing from the world of academia and armed each with their respective doctorate, their profiles are ideally situated at the point where research, development, and applications meet. “Our goal is to bring researchers and entrepreneurs together. To do so, it's essential to understand both worlds, and their respective needs and capabilities,” Monay explains.

On their first day at the TTO, Salamin and Monay began updating the institute's technology portfolio. They drew up an inventory of the technologies developed by Idiap's researchers and of their respective degrees of

completion. Salamin takes up the story: “The portfolio now contains around 40 technologies. But it's rare that even one of them corresponds exactly to the specific needs of any given company. So we use them primarily to showcase what we can do, allowing us to show potential partners concrete examples of what Idiap can offer.”

## Linking research and industry together

On the basis of this showcasing, a company can decide to set up a project and can ask for support from the Commission for Technology and Innovation (CTI) or the Valais innovation foundation—The Ark. “In this initial phase,” Monay explains, “our job is to find, within Idiap, the researcher the best qualified to take the scientific lead on the project.” This done, the two help the company to prepare the funding application.

## TECHNOLOGY TRANSFER

If that application is successful, they then assist the company in submitting a patent request. “This is carried out on the request of the researcher. Prior to submission, the request passes before Idiap’s patent committee, composed of the institute’s management, a member of the TTO, and a researcher who has extensive patent experience,” explains Monay. “The filing of a patent is a long and costly process. We do all we can to ensure the intrinsic value of any given invention before launching ourselves on this path.”

### Motivating researchers

In terms of Idiap’s internal operations, the updating of the institute’s technology portfolio was also carried out with the aim of motivating researchers to become more active in the field of technology transfer. “The portfolio, and technology transfer, also enable researchers to identify new problems to which their research can be

applied, and thus to benefit from additional resources with which to improve their algorithms,” Salamin explains.

At the end of their first year, the two managers of the TTO draw positive conclusions. “In 2015, Idiap researchers have applied for and received funding for three Ark projects and four CTI projects, and filed three patent applications in the fields of biometrics and speech processing.”

Hugues Salamin and Florent Monay are delighted with what companies now know about Idiap. “At seminars, trade fairs, and conferences, it’s our researchers who take the first step and make the first contact. Our job is then to grow that contact and win the loyalty of companies that have already placed their trust in the institute.”







**faces**  
**36 — 43**



# Statistics as an interdisciplinary language

Specialist in applied mathematics and statistics, David Ginsbourger cultivates his taste for the interdisciplinary approach—a path that has led to collaborations in fields as diverse as engineering, hydrogeology, and climate science.

When an Idiap researcher talks about “machine learning”, one expects to hear a lot about visual recognition techniques and robotics. David Ginsbourger, though, associates the field with climate science. But then, what one takes away from a meeting with David, besides his interest in statistics, is his taste for interdisciplinary interaction.

For David, “similar approaches appear in parallel in different disciplines, with uses and terminology that can vary significantly from one to the next. The use of mathematical language and statistical concepts often helps clarify things. Machine learning researchers come from a range of backgrounds, each with their own codes, ways of thinking, and aims, but—with regard to many of the fundamental questions—they share a certain common ground.”

## In silico experiences

After a dual degree at the École des Mines de Saint-Etienne and the Technische Universität Berlin, David Ginsbourger remained at Saint-Etienne and began doctoral studies in applied mathematics, working on the digital experience planning problematic. “Simulation is used in a whole range of areas to supplement, or even replace, real experiences. But digital experience itself has a cost—particularly in computation time. This is why experience planning strategies are used.” On a trip to the US, David’s path crosses that of hydrogeologist Philippe Renard. Later, he joins Renard’s team at the University of Neuchâtel, as an associate researcher.

**“Similar approaches appear in parallel in different disciplines.”**

**David Ginsbourger**  
Senior researcher



## Idiap gains a foothold in German-speaking Switzerland

And David’s work with that team is in full swing when he is offered a lecturing post at the University of Bern’s Institute of Mathematical Statistics and Actuarial Science. “I found myself at an institute with a tradition both of top-level theoretical work and of academic freedom—this allowed me to deepen my knowledge of statistics while simultaneously cultivating a number of interdisciplinary collaborations,” he explains.

Upon qualification, while searching for a professorship or a position as a permanent researcher, an opportunity to join Idiap presents itself. “We managed to put together a post split across both sites—Idiap and Bern”.

Since September 2015, David has worked on developing an Idiap research group on the topics of optimization and uncertainty quantification, while retaining a part-time post at the university. The appointment has come at just the right moment in David Ginsbourger’s career: “It will allow me to engage in a new cycle of original research projects, in the highly energizing context of Idiap”. True to that taste for the

interdisciplinary approach, he is looking forward to working with the institute’s other permanent researchers.

# The hidden face of the research project



38

**In the world of research, as elsewhere, money is indispensable. At Idiap, Elisa Bovio supports researchers with their funding applications and keeps the accounts of a range of research projects.**

Having previously worked for an insurance company, Elisa Bovio—a young graduate of Sierre’s School of Business Administration—joined Idiap’s ranks in June 2015. She has a dual role: Financial Assistant and Junior Program Manager.

## Helping scientists

As a financial assistant, Elisa is responsible for salaries and expense reports. She also manages the cost accounting process for research projects, ensuring that they respect their various budgetary envelopes. This last task enables her to better connect with her second set of responsibilities—project administration.

As a Junior Program Manager, Elisa’s principal tasks are to help researchers with project submissions, and to

manage the many contracts established between Idiap and a whole range of partners. More specifically, she supports Dr. Jean-Marc Odobez with regard to the European project MuMMER (see p. 26) and handles those Idiap projects funded by the US government agency DARPA and the subsidies received from the Hasler Foundation and the European H2020 program. All in all, a very diverse mandate, with in-house and wider-world connections—a particularly motivating mix.

## A cosmopolitan environment

Strange to think then that Elisa first came across Idiap purely by chance during her studies, thanks to a visit to the institute organized for the University of Applied Sciences that she was attending. “At the time,” she says, “I couldn’t imagine myself working in the research sector—it seemed very remote.” Finding herself—in professional terms—somewhat cramped in the world of insurance, this yoga lover with a passion for travel decided to change direction and join Idiap. “I found an environment and people that were fascinating. I especially appreciate the multilingual aspect of the institute and the diversity of backgrounds, whether personal or professional. I have a certain wanderlust—I’ve already spent almost a year in North America and have traveled several times in Southeast Asia. Which is why I appreciate all the more the cosmopolitan atmosphere here.”

Elisa also enjoys the variety of tasks assigned to her, and the freedom that goes with that variety. “Having two roles is, in this regard, a real plus. My schedule varies significantly throughout the year, which helps me to some extent to avoid the negative side of routine, and allows me to better balance the needs of the researchers and the institute. This year I was also able to get involved in the organiza-

tion of a seminar on ISO standards in biometrics, and in events related to the bicentennial of Valais joining the Swiss Confederation. Next year, I’ll be handling the International Create Challenge (see pp.31-32)—yet another opportunity to submerge myself in the world of new technologies. And, to be honest, that’s something I’m enjoying more and more.”

**“I found an environment and people that were fascinating.”**

**Elisa Bovio**

Financial Assistant and  
Junior Program Manager

## Follow my gaze



**Fresh from ICC'2015, which he won brilliantly, Kenneth Funes Mora decided to exploit the results of his doctorate by creating his own company—Eyeware Tech. At the same time, this specialist in 3-D eye tracking is pursuing his research career as a postdoc at Idiap.**

It was a long road and an Erasmus Mundus scholarship that combined to lead this Costa Rican researcher to Martigny. A road that began with a degree from the Instituto Tecnológico of his homeland, and took in Heriot-Watt University, Edinburgh; Dijon; Girona; and INRIA, Grenoble, where Kenneth Funes Mora completed his Master's thesis.

He joined Idiap and Dr. Jean-Marc Odobez's group in January 2011 and began a PhD on 3-D eye tracking. This work has a number of outlets, in terms both of research and of applications. The goal is to determine, at each moment, where the eyes and attention of a person are focused, without recourse to any particularly sophisticated interfaces. And while researchers in the humanities will be able to apply such a feature to studying social interactions, the technology also has applications in the fields of robotics, cognitive science, and marketing.

### **Innovation—an adventure**

His thesis complete—he was awarded his academic title in October 2015—Kenneth embarked on the innovation adventure, enrolling in the Arkathon: Switzerland's first “hackathon” dedicated to healthcare, organized by The Ark. Along with his friend and colleague Carlos Becker—himself pursuing PhD studies at EPFL in the field of Computer Vision—Kenneth set himself the task of finding an application for his research results. “Even if the human face is very expressive,” Kenneth explains, “our recognition algorithms are quite capable of deciphering it. By using predetermined expressions as a code, we'll be able to propose to quadriplegics a system that allows a computer to be operated quite easily by associating movements of the head and face with commands. Such hands-free systems already exist, but they're expensive and often require specific hardware (reflective dots applied to the forehead or to headgear) that only a third party can put in place. Our system—which could be described as a virtual mouse—only requires a simple 3-D camera, a basic peripheral that will soon replace the webcam found in most computers.”

Following the success they encountered at the Arkathon, the duo decided to enroll for ICC'2015 (see article, pp.31-32). Kenneth's partner then put him in touch with Serban Mogos, an acquaintance with a doctorate in Entrepreneurship from Carnegie Mellon University. Their skills combined, the project of the “virtual mouse” advanced and took shape—to the extent that the trio won the competition. “There were long and intense days of work, but we learned a lot there. It was very exciting—a really great experience,” Kenneth recalls.

After receiving a “Helping Hand” (*Coup de pouce*) grant from the Dr. René Liechti Foundation, the trio make that great leap forward. Along with Jean-Marc Odobez, they launch their own business—Eyeware Tech SA—with the support of Idiap and the business incubator IdeArk.

Kenneth Funes Mora's thoughts turn to the future: “We aim to put our virtual mouse system on the market, at the same time improving it by incorporating the 3-D eye-tracking technology that I worked on for my PhD—technology that I'll continue to work on as a postdoc at Idiap, again in Jean-Marc Odobez's group. I'm looking forward to continuing my work with Jean-Marc. After all, he supported me during my PhD and was there at the birth of Eyeware Tech.”

# 25 candles on Idiap's cake

The year 2015 marks 25 years of activity at Idiap. From its foundation to its silver jubilee, we look back on 25 events that make up the history of an institute that, this year, attained its majority.

## Idiap timeline

**1991**

With the support of the Dalle Molle Foundation for Quality of Life, Idiap first sees the light of day, taking up residence in Villa Tissières, Martigny.

**1996**

A new director—Professor Hervé Bourlard—comes on board.



**2001**

Coordinated by Idiap, the National Centre for Competence in Research (NCCR) Interactive Multimodal Information Management (IM2) is launched.

**2002**

Idiap creates a stir—analyzing the Osama bin Laden recordings.

**2002**

Idiap expands, building the Dalle Molle Pavilion in Martigny.

**2002**

The 12th IEEE (Institute of Electrical and Electronic Engineers) conference on the subject of “Neural Networks for Signal Processing” becomes the first international conference organized by Idiap in Martigny.

**2003**

First project with the United States government.

**2004**

The funding of a new round of large-scale European projects (AMI/AMIDA) has a significant international impact.

**2005**

IdeArk SA—the start-up incubator whose mission is to develop Idiap technologies—is created.

**2007**

Launch of Klewel, Idiap's first spin-off.



**2007**

Idiap moves into new premises at the Centre du Parc, Martigny.



**2008**

Signing of a strategic alliance with the ETH domain.



**2009**

Opening of a showroom at the institute.

**2010**

Idiap unveils the first version of its technology portfolio.

**2011**

Idiap signs a collaboration agreement with the prestigious Indian Institute of Technology Guwahati (IITG) and the International Institute of Information Technology, Hyderabad (IIIT).

**2011**

First EPFL academic posts awarded to three of Idiap's permanent researchers.

**2012**

For the first time, Idiap's operating accounts exceed the symbolic 10 million franc mark.

**2012**

The project MediaParl begins, with the support of the Loterie Romande and Valais' cantonal parliament.

**2012**

Creation of the Association Interactive Multimodal Information Management (AIM2)—the continuation of the NCCR IM2.

**2012**

The first International Create Challenge—ICC'2012—is launched.

**2013**

Idiap's first patents validated.



**2014**

The Swiss Center for Biometrics Research and Testing—the first European center for biometrics—opens its doors; hosted by Idiap.

**2014**

At the instigation of the institute's director, seven leading figures from the world of science carry out an in-depth audit of Idiap's activities.

**2015**

Idiap participates in the festivities organized to mark the bicentennial of Valais joining the Swiss Confederation; its contribution—the project Valais\*Wallis Digital.

**2016**

To celebrate its twenty-fifth anniversary, Idiap organizes an "Idiap Innovation Day" to present its activities to companies in the region.

# Awards

## Internal

Each year Idiap awards prizes to two of its research assistants. The first award is for research, the second for papers published. Candidates for the Research Prize are evaluated, by an internal commission, against five criteria: papers published, teamwork, commitment to a project, communication skills, and autonomy. For the Publication Prize, the institute's leadership draws up an initial list of candidates from papers of which an Idiap research assistant is the primary author. The members of the Advisory Board then evaluate, independently and anonymously, the papers selected.

In 2015 the Research Prize was awarded to **Ivana Chingovska** for the novelty of her research contributions to biometrics anti-spoofing applied to face recognition, for the number and quality of her publications—especially in 2015—and for her participation in the Idiap and EDEE communities. The Research Publication Prize went to **Nikolaos Pappas** for his outstanding scientific article entitled *Adaptive sentiment-aware one-class collaborative filtering*.



Ivana Chingovska

## External

This year, Idiap wished to acknowledge its researchers' brilliant participation at international conferences. The quality of their research was rewarded with several distinctions.

N. LE, D. WU, S. MEIGNIER AND J.-M.ODOBEZ.  
 MEDIAEVAL 2015, WURZEN, SEPTEMBER 2015,  
 WINNER OF THE PERSON DISCOVERY SUBTASK.  
**EUMSSI team at the MediaEval Person Discovery challenge.**

M. RAZAVI, R. RASIPURAM, AND M. MAGIMAI DOSS.  
 LANGUAGE AND TECHNOLOGY CONFERENCE (LTC)  
 2015, POZNAN, POLAND, BEST STUDENT PAPER AWARD.  
**Pronunciation Lexicon Development for Under-Resourced Languages Using Automatically Derived Subword Units: A Case Study on Scottish Gaelic.**

C. MCCOOL, R. WALLACE, M. MCLAREN,  
 L. EL SHAFEY, AND S. MARCEL.  
 IET 2015 BIOMETRICS PREMIUM BEST PAPER AWARD.  
**Session Variability Modelling for Face Authentication.**

N. CHACKO, K. CHAN, ET M. LIEBLING.  
 IEEE INTERNATIONAL SYMPOSIUM ON BIOMEDICAL  
 IMAGING, BROOKLYN, NY, APRIL 2015, BEST STUDENT  
 PAPER AWARD.  
**Intensity-Based Point-Spread-Function-Aware Registration for Multi-View Applications in Optical Microscopy.**



Nikolaos Pappas

## Theses completed

Seven students completed their theses in 2015: Kenneth Alberto Funes Mora, Ashtosh Sapru, Laurent Son Nguyen, Mohammad J. Taghizadeh, Maryam Habibi, Sree Harsha Yella, and Ivana Chingovska.

3D GAZE ESTIMATION FROM REMOTE RGB-D SENSORS

**Kenneth Alberto Funes Mora**

May 29, 2015

Thesis director: Dr. Jean-Marc Odobez

Members of the thesis committee:

Prof. Kamiar Aminian, D<sup>r</sup> Ronan Boulic,  
Prof. Dan Witzner Hansen,  
Prof. Louis-Philippe Morency

AUTOMATIC SOCIAL ROLE RECOGNITION AND  
ITS APPLICATION IN STRUCTURING MULTIPARTY  
INTERACTIONS

**Ashtosh Sapru**

April 28, 2015

Thesis director: Prof. Hervé Bourlard

Members of the thesis committee:

Prof. David Atienza Alonso,  
Prof. Fabio Pianesi, D<sup>r</sup> Alessandro Vinciarelli,  
Dr. Jean-Marc Vesin

COMPUTATIONAL ANALYSIS OF BEHAVIOR  
IN EMPLOYMENT INTERVIEWS AND VIDEO RESUMES

**Laurent Son Nguyen**

February 18, 2015

Thesis director: Prof. Daniel Gatica-Perez

Members of the thesis committee:

Prof. Jean-Philippe Thiran,  
Prof. Marianne Schmid Mast, Prof. Mohamed  
Chetouani, Prof. Sabine Sússtrunk

ENABLING SPEECH APPLICATIONS USING AD HOC  
MICROPHONE ARRAYS

**Mohammad Javad Taghizade**

March 27, 2015

Thesis directors: Prof. Hervé Bourlard,

Dr. Philip N. Garner

Members of the thesis committee:

Prof. Jean-Philippe Thiran,  
Prof. Emanuel Habets, Prof. Dietrich Klakow,  
Dr. Hervé Lissek

MODELING USERS' INFORMATION NEEDS  
IN A DOCUMENT RECOMMENDER FOR MEETINGS

**Maryam Habibi**

November 30, 2015

Thesis theses: Prof. Hervé Bourlard,

Dr. Andrei Popescu-Belis

Members of the thesis committee:

Prof. Jean-Philippe Thiran, D<sup>r</sup> Jim Glass,  
Prof. Stéphane Marchand-Maillet,  
Dr. Martin Rajman

SPEAKER DIARIZATION OF SPONTANEOUS MEETING  
ROOM CONVERSATIONS

**Sree Harsha Yella**

January 19, 2015

Thesis director: Prof. Hervé Bourlard

Members of the thesis committee:

Prof. Anja Skrivervik,  
Dr. Andreas Stolcke, D<sup>r</sup> Xavier Anguera,  
Dr. Jean-Marc Vesin

TRUSTWORTHY BIOMETRIC VERIFICATION  
UNDER SPOOFING ATTACKS:  
APPLICATION TO THE FACE MODE

**Ivana Chingovska**

November 23, 2015

Thesis directors: Prof. Hervé Bourlard,

Dr. Sébastien Marcel

Members of the thesis committee:

Prof. Sabine Sússtrunk, Prof. Julian Fierrez-Aguilar,  
Prof. Raymond Veldhuis, Prof. Jean-Philippe Thiran



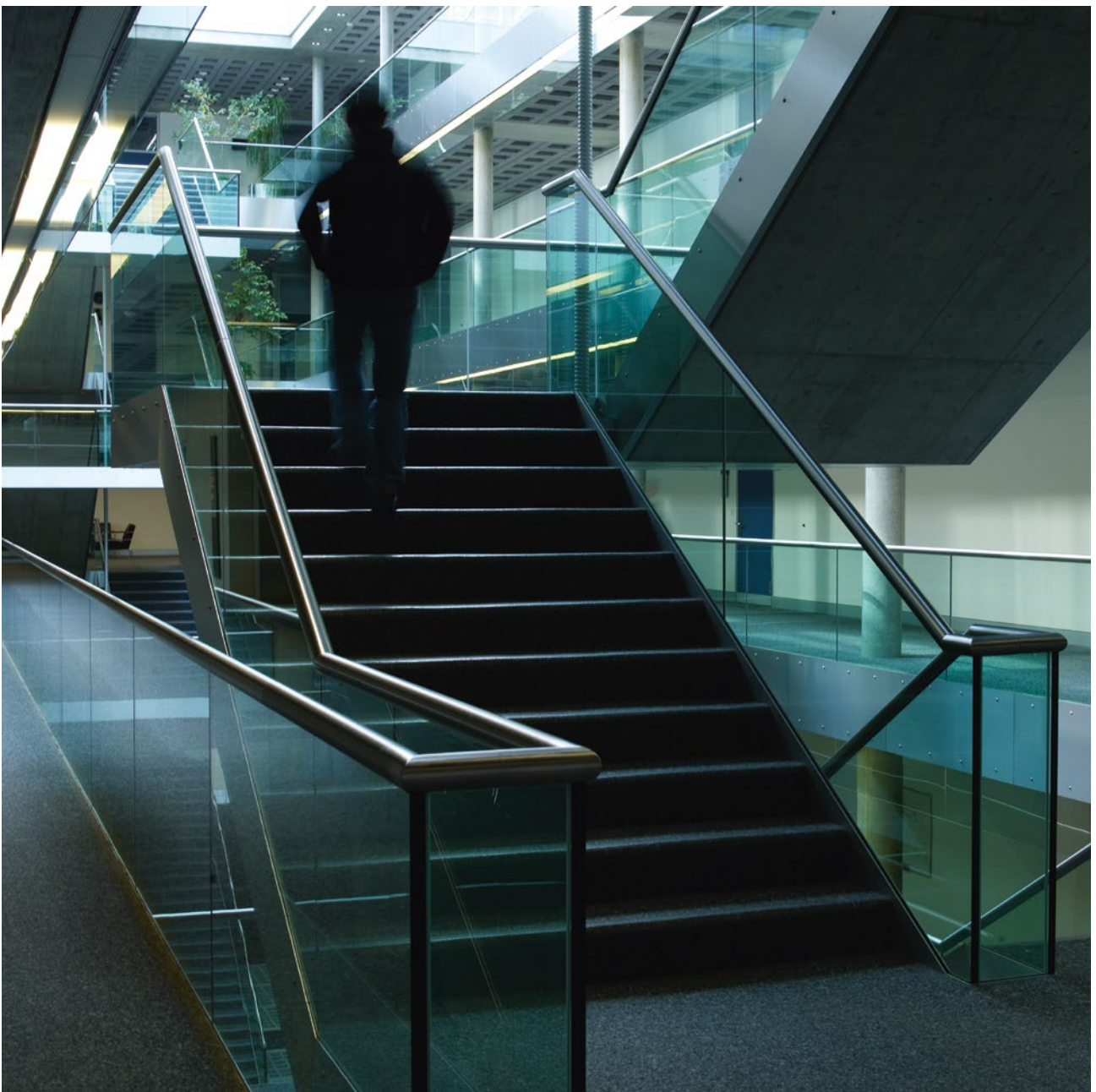
Centre du Parc, Rue Marconi 19,  
case postale 592, CH-1920 Martigny

T +41 27 721 77 11

F +41 27 721 77 12

M [info@idiap.ch](mailto:info@idiap.ch)

[www.idiap.ch](http://www.idiap.ch)





# **scientific inserts**

**I — XLIV**

# Research areas

To face its continuous growth and diversification in different, but complementary, research directions, while building upon its general institutional theme of "Human and Media Computing" (as defining ourselves on our website [www.idiap.ch](http://www.idiap.ch)),

Idiap's expertise mainly spans five research areas, which are summarized in Figure 1 below. In that table, each research area is also described in few keywords. As described later, those generic research areas are covered by 10 research group.

Idiap Research Areas (horizontal)	Keywords
<b>Perceptual and cognitive systems</b>	Speech and audio processing, computer vision, handwriting recognition, document processing, computational cognitive sciences, robotics, natural language processing, machine translation
<b>Human and social behavior</b>	Web and mobile social media, social interaction sensing, social signal processing, verbal and nonverbal communication analysis, mobile phone sensing, computational social science
<b>Information and presentation interfaces</b>	Multimedia information systems, user interfaces, contextualization, personalization, system evaluation, mobile HCI using Big Data, data driven services
<b>Biometrics</b>	Speaker recognition, face recognition, multimodal biometric fusion, mobile biometry, spoofing and anti-Spoofing
<b>Machine learning</b>	Statistical and neural network based ML, computational efficiency, online learning, multi-sensor processing, very large datasets

**Figure 1**

Overview of Idiap research areas.

## PERCEPTUAL AND COGNITIVE SYSTEMS

Speech processing; Natural language understanding and translation; Document and text processing; Vision and scene analysis; Multimodal processing; Cognitive sciences and Robotics.

Idiap combines its multi-disciplinary expertise to advance the understanding of human perceptual and cognitive systems, engaging in research on multiple aspects of human-computer interaction with computational artefacts such as natural language understanding and translation, document and text processing, vision and scene analysis (with a particular emphasis on human sensing), multimodal interaction, computational cognitive systems, robotics, and methods for automatically training such systems.

## HUMAN AND SOCIAL BEHAVIOUR

Social media; Mobile media; Social interaction analysis; Social signal processing.

This area spans methods for analyzing human and social behaviour from a variety of information sources with the goals of understanding social phenomena and developing human-centered applications. Idiap has investigated new approaches for behavioral analysis in face-to-face communication, online interaction in social media sites like YouTube, and smartphone-mediated interaction.

## INFORMATION AND PRESENTATION INTERFACES

Multimedia information systems, User interfaces; System evaluation.

Information processing by computers must be accompanied by human-computer interfaces that present information and receive input in an efficient and usable way, possibly acquiring information from users in a non-disruptive way. Current research directions at Idiap focus on multimedia information systems, search and recommendation, and interactive information retrieval, and several new projects are currently being initiated in those directions.

## BIOMETRIC PERSON RECOGNITION

Face recognition (detection-localization-identification-verification); Speaker identification and verification; Multimodal biometric person recognition; Counter-measures to spoofing attacks.

Biometric person recognition (Biometrics) refers to the process of automatically recognizing a person using distinguishing behavioural patterns (gait, signature, keyboard typing, lip movement, hand-grip) or physiological traits (face, voice, iris, fingerprint, hand geometry, EEG, ECG, ear shape, body odour, body salinity, vascular). Idiap is currently a recognized leader in that field, fully exploiting our multi-disciplinary expertise in image processing, computer vision, pattern recognition and machine learning. This expertise has recently been confirmed by the official launch of a 'Swiss Center for Biometric Security Research and Testing' (<http://www.biometrics-center.ch/>) hosted at Idiap.

## MACHINE LEARNING

Statistical and neural network based machine learning; Computational efficiency, targeting real-time applications; Very large datasets; Online learning.

Research in machine learning aims at developing computer programs able to learn from examples. Instead of relying on a careful tuning of parameters by human experts, machine-learning techniques use statistical methods to directly estimate the optimal setting, which can hence have a complexity beyond what is achievable by human experts. Today, Idiap is also recognized as a key leader in that field with new trends towards "collaborative machine learning", deep neural network architectures, and large-scale distributed learning algorithms.

## NEW RESEARCH THEMES INITIATED IN 2015

In addition to these "horizontal" research themes, new activities recently started to be developed towards bio-medical applications (through the new Computational Biomedicine Group) and Environmental Modeling (through the new group in "Uncertainty Quantification and Optimal Design").

# 1. Speech and Audio Processing

## OVERVIEW

**Heads:** Hervé Bourlard (MS and PhD, Polytechnic University, Mons, Belgium, 1982 and 1992), Philip N. Garner (MEng, University of Southampton, UK, 1991; PhD, University of East Anglia, UK, 2011), Mathew Magimai-Doss (MS by Research, Indian Institute of Technology Madras, India, 1999; PhD, Ecole Polytechnique Fédérale de Lausanne, Switzerland, 2005), 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. Today it is still the largest group within the institute, and Idiap continues to be recognised as a leading proponent in the field. The expertise of the group encompasses statistical automatic speech recognition (based on hidden Markov models, or hybrid systems exploiting connectionist approaches), text-to-speech, speaker recognition 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).

The Speech and Audio Processing group in 2015 was composed of 1 head of group, 3 Principal Investigators, 9 postdocs, 1 scientific collaborator, 10 PhD students and 4 interns. In addition to that, the group also worked closely with two Senior Development Engineers from the development team at Idiap on industrial research projects.

## KEY SCIENTIFIC OUTPUTS

We have always used both neural network and Markov model based approaches, leading to the KL (Kullback Leibler)

HMM approach to fusion of neural network and HMM based speech recognition, especially for multilingual speech modelling. Use of techniques from HMM based speech recognition in HMM based speech synthesis resulted in a unified approach to speech recognition and synthesis. The group was well placed to take advantage of recent advances in "deep" neural networks (DNNs) technology, and the associated Kaldi automatic speech recognition (ASR) toolkit. Template-based speech processing has been always in attention to utilize the power of big data with less statistical assumptions. In this context, Idiap is recognized to exploit neural network posterior features as speech templates, and structured sparsity of posterior features have led to hierarchical frameworks for ASR, query-by-example spoken term detection (QbE-STD), speech coding and linguistic parsing applications. Several key contributions in microphone arrays, including ad-hoc microphone arrays, are well recognized. Sparse model based approaches have led to novel algorithms for source separation. The agglomerative clustering approach to speaker diarization is well regarded. Finally, the work in this group also resulted in several spin-offs, including dev-audio, Koemei and, more recently (end of 2014), recapp.

Much of the group's recognition has come from contributions to the speech and audio community; these include the AMI corpus<sup>1</sup> and the Juicer ASR decoder<sup>2</sup>. More recently, the MediaParl<sup>3</sup> and Walliserdeutsch<sup>4</sup> databases have been released; other contributions such as ISS and SSP are on the GitHub site<sup>5</sup>. This highlights that, although the group has traditionally worked with English speech, many recent projects are multi-lingual in nature. This has in turn become a focus of the group; one on which we are well placed to capitalise given our geographical location. The group is also very active in contributing to the Kaldi open-source ASR toolkit<sup>6</sup>.

## AUTOMATIC SPEECH RECOGNITION

In recent years, at Idiap the 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 parametrization and acoustic modeling**

We are still investigating new features (e.g., posterior-

<sup>1</sup> [www.idiap.ch/mmm/corpora/ami](http://www.idiap.ch/mmm/corpora/ami)

<sup>2</sup> [juicer.amiproject.org/juicer/](http://juicer.amiproject.org/juicer/)

<sup>3</sup> [www.idiap.ch/dataset/mediaparl](http://www.idiap.ch/dataset/mediaparl)

<sup>4</sup> [www.idiap.ch/dataset/walliserdeutsch](http://www.idiap.ch/dataset/walliserdeutsch)

<sup>5</sup> [github.com/idiap](https://github.com/idiap)

<sup>6</sup> [kaldi.sourceforge.net](http://kaldi.sourceforge.net)

<sup>7</sup> [www.idiap.ch/scientific-research/projects/dbox](http://www.idiap.ch/scientific-research/projects/dbox)

<sup>8</sup> [www.idiap.ch/scientific-research/projects/a-muse](http://www.idiap.ch/scientific-research/projects/a-muse)

<sup>8</sup> [www.idiap.ch/scientific-research/projects/a-muse](http://www.idiap.ch/scientific-research/projects/a-muse)

<sup>9</sup> [www.idiap.ch/scientific-research/projects/phaser](http://www.idiap.ch/scientific-research/projects/phaser)

<sup>9</sup> [www.idiap.ch/scientific-research/projects/phaser](http://www.idiap.ch/scientific-research/projects/phaser)

<sup>9</sup> [www.idiap.ch/scientific-research/projects/phaser](http://www.idiap.ch/scientific-research/projects/phaser)



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). Over the last 12 months, our research has largely focussed on the use of DNNs for far-field speech recognition; novel approaches to adapt DNNs across different domains and languages; sparse modeling and low-rank representation to structure DNN testing posteriors according to the training domain for unsupervised acoustic model adaptation; development of articulatory feature (phonological feature) based ASR system; development of a unified framework for recognition and detection; (funded through the EC Eurostars project D-Box<sup>7</sup> SNSF projects like A-MUSE<sup>8</sup>, PHASER<sup>9</sup>, FlexASR<sup>10</sup>, and IM2 NCCR sub-project Walliserdeutsch, HASLER Foundation project AddG2SU<sup>11</sup>, CTI project ScoreL2<sup>12</sup>, and Samsung Electronics Co., Ltd. (South Korea) sponsored project AMASSE<sup>13</sup>).

• **Cross-lingual and multi-lingual speech recognition**

Within the scope of the D-Box project, 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 bootstrapped using well-resourced data and adapted to the target language (or environment). In the context of an industry sponsored project (Samsung, Inc. South Korea sponsored project DAUM<sup>14</sup>), we are also investigating Markov model based acoustic model adaptation techniques (currently referred to as "subspace Gaussian modeling") for different domains (e.g., accent or dialect).

The recogniser is often required to perform in real-time and to support closed- and open-dictionary (i.e., detection of key-words) decoding graphs. Multi-lingual or domain adaptation is often combined with speaker-adaptation techniques to further improve recognition accuracies.

• **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 (illustrated in Figure 4), 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 towards the first commercial product that performs Swiss German (dialect) speech recognition. The subject of dialects remains pertinent as an important local focus, with plans underway to expand to other regions and countries such as Germany.

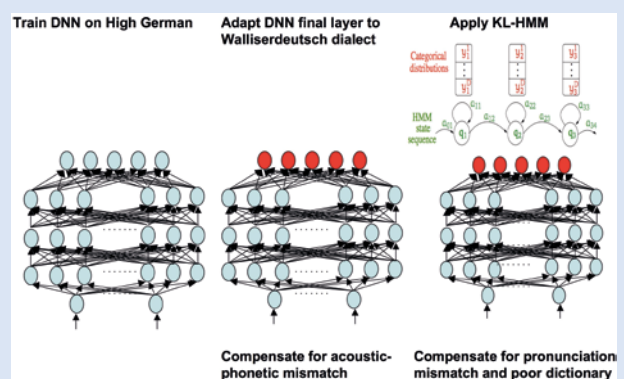


Figure 4

Illustration of development of Walliserdeutsch ASR system, merging Deep Neural Network (DNN) technology and a new form of Hidden Markov Model (KL-HMM).

• **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 project A-MUSE, SNSF project PHASER, and CTI project MultiVEO<sup>15</sup>. In particular, our current research is focussing on linking template-based speech recognition and statistical sequence model based speech recognition. Towards this end there are two on-going efforts. The first focuses on development of a "sparse" HMM framework by establishing a link between statistical speech recognition formalism and the recent theories of compressive sensing and sparse modeling. The framework integrates the advantages of template-based modelling within the HMM sequence matching scheme. The second effort focusses

<sup>10</sup> <https://www.idiap.ch/scientific-research/projects/flexasr>

<sup>12</sup> <https://www.idiap.ch/scientific-research/projects/scorel2>

<sup>14</sup> <http://www.idiap.ch/scientific-research/projects/daum>

<sup>11</sup> <https://www.idiap.ch/scientific-research/projects/addg2su>

<sup>13</sup> <http://www.idiap.ch/scientific-research/projects/amasse>

<sup>15</sup> <https://www.idiap.ch/scientific-research/projects/multiveo>

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. Over the past 12 months, this vision was realized through sparse modeling of posterior features. It has been shown that dictionary learning for sparse recovery of posteriors enables enhanced acoustic modeling for statistical ASR through hierarchical reconstruction of posterior templates (funded by SNSF project PHASER).

#### • Grapheme-based ASR

Over the past five 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). In the past 12 months, our research activities in that direction have primarily focussed on exploiting the approach for lexical resource development through (a) acoustic data-driven grapheme-to-phoneme conversion and (b) derivation of "phone-like" subword units automatically from the acoustic data.

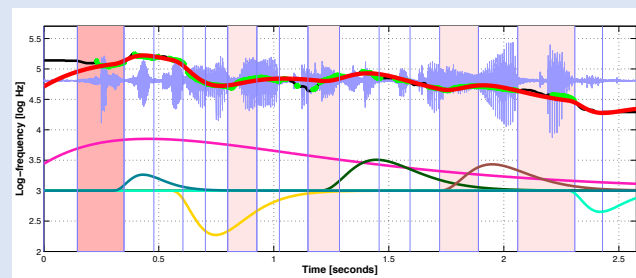
These research directions are particularly addressing practical challenges such as non-native or accented speech recognition, rapid development of ASR systems for new or under-resourced languages and/or domains. In addition, building on top of the core ASR research, we are also actively carrying out research on development of novel approaches for (a) utterance verification, (b) keyword spotting (spoken term detection), (c) objective assessment of coded speech and synthetic speech, and (d) automatic assessment of second language learners (funded through SNSF project A-MUSE, Idiap internship project, EC Eurostars project D-Box, armasuisse<sup>16</sup>, CTI project ScoreL2<sup>17</sup>). Keyword spotting is also addressed in the Applied Machine Learning group.

## SPEECH SYNTHESIS

#### • Text-to-speech synthesis (TTS)

Although newer than ASR, TTS is now an established venture for the speech group at Idiap. TTS is central to two ongoing projects: SIWIS<sup>18</sup> (Spoken Interaction With Interpretation in Switzerland), D-BOX, is DBOX still active? and SP2<sup>19</sup> (SCOPES project on speech prosody).

The TTS work at Idiap is largely aligned with the current trend in the field towards statistical synthesis, which uses the same technology as ASR. SIWIS aims to do speech



**Figure 5**

Modelling prosody using distinct atomic muscular twitches.

to speech translation in Swiss languages, including the concept of accents. A key theme of SIWIS is adaptation of speech prosody, i.e., spoken emotion. This will ensure that, e.g., questions and stressed phrases remain as such after translation. To this end, we are developing novel models of speech prosody. The goal of SIWIS are mirrored in the SP2 project with a focus on eastern-European languages. Figure 5 illustrates one of our recent approaches to decompose the pitch contour (superimposed on top of the speech signal) into "atomic" patterns (displayed below the speech signal and pitch contour), which can be related to the muscular twitches at the glottal source. SIWIS and SP2 highlight the requirement for a collaborative approach: Speech to speech translation requires a large infrastructure that is difficult to maintain at a single site. To achieve this, SIWIS brings together partners at ETHZ and the Universities of Geneva and Edinburgh; SP2 adds partners from Budapest, Skopje and Novi Sad.

R&D work in TTS is also one of active areas for D-BOX project. Besides integrating the engine into an open-gaming platform through the implemented API supporting real-time multi-

<sup>16</sup> [www.ar.admin.ch/internet/armasuisse/en/home.html](http://www.ar.admin.ch/internet/armasuisse/en/home.html)

<sup>17</sup> [www.idiap.ch/scientific-research/projects/scorel2](http://www.idiap.ch/scientific-research/projects/scorel2)

<sup>18</sup> [www.idiap.ch/project/siwis/](http://www.idiap.ch/project/siwis/)

<sup>19</sup> [www.idiap.ch/scientific-research/projects/sp2](http://www.idiap.ch/scientific-research/projects/sp2)

<sup>20</sup> [www.siip.eu](http://www.siip.eu)

player response, research in fast adaptation of TTS engine towards target speaker or different language is carried out.

- **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. In 2015, work on speech coding continued under the SP2 project, mainly on aspects concerning prosody.

## SPEAKER RECOGNITION AND SPEECH ANALYTICS

Starting from 2014, within the scope of the SIIP (EC FP7 project)<sup>20</sup>, there has been a rejuvenation of core speaker recognition research in the Speech and Audio Processing group. In particular, as illustrated in Figure 6, Idiap is involved in developing a cutting edge suspect identification solution for identifying speakers captured in lawfully intercepted calls, including multiple types of social-media information streams. In addition to core speaker recognition, the technology will capitalise on fusion of other speech analytics algorithms such as gender, age, language, or accent identification, keyword and taxonomy spotting and voice cloning detection. Speaker recognition is also addressed in the Biometric Person Recognition group, as discussed in Section 9.



**Figure 6**

Illustration of suspect identification system exploiting lawfully intercepted calls, including multiple types of social-media information streams.

## AUDIO PROCESSING

Besides ASR, TTS and speaker recognition, Idiap has been active in other areas of generic audio signals (pre-)processing and metadata extraction, including:

- **Query-by-example spoken term detection**

Idiap continues studies on the subject of keyword spotting with a focus on searching large audio archives using audio queries. The task is referred to as query-by-example spoken term detection (QbE-STD). The leading edge systems rely on template matching using posterior features as templates. Posterior features are sparse, and sparsity indicates their low-dimensional attributes. In 2015, our research focused on exploiting sparse modeling for characterizing the low-dimensional subspaces of query and background posteriors. QbE-STD is thus cast as a subspace detection problem (funded through SNSF project A-MUSE).

- **Sparse structured representations**

Idiap continues theoretical studies of sparse structured representations of audio signals, more specifically on model-based sparse component analysis, with different practical applications, such as source separation and localisation based on a small number of microphones, room acoustic modelling, and new beamforming techniques to improve recognition of speech recorded from farfield microphones. Exploiting structured sparsity pertained to speech perception and multipath propagation enables auditory and acoustic informed source separation and localization for far field ASR (supported by SNSF project PHASER).

- **Microphone arrays**

Idiap also works on audio recorded from microphone arrays, especially those organized in "ad-hoc" geometries (funded originally through the NCCR IM2, and more recently PHASER). Recently, we mainly focused on the automatic calibration of arbitrary microphone array geometries, fusing Euclidean constraints into generic mathematical concepts. This in turn enables the localization required for high quality data acquisition. In this context, we also consider environmental ambiguities such as reverberation and overlapping speech, thus requiring multi-source localization using ad-hoc microphones in reverberant rooms. Recently, we investigated the properties of time difference of arrival matrices specifying relative positioning of a source with respect to the pairs of microphones. Their low-rank property was exploited in an algebraic formulation to enable denoising and robust source localization using ad hoc microphones.

<sup>21</sup> <http://www.idiap.ch/scientific-research/projects/rodi>

<sup>23</sup> <http://www.inevent-project.eu/>

<sup>22</sup> <https://www.idiap.ch/scientific-research/projects/dimha>

<sup>24</sup> <http://www.idiap.ch/scientific-research/projects/sesame>

- **Speaker diarization**

In speaker diarization—the task of annotating temporal regions of audio recordings with labels indicating “who spoke when”—Idiap focuses on processing spontaneous overlapping speech through the use of multiple sources of information investigated in the context of various projects, including the SNSF project RODI<sup>21</sup>, SNSF project DIMHA<sup>22</sup> and EU project inEvent<sup>23</sup>. Over the last 12 months, our research has focussed on (a) linking and uniquely identifying speakers across large speech databases by making use of multiple modalities, (b) extraction of novel features in the frameworks of i-vectors and ANNs, and (c) scaling the KL-HMM approach, originally developed for ASR, to speaker diarization.

- **Audio indexing**

In the context of the Hasler Foundation project SESAME<sup>24</sup>, Idiap aims at advancing the state-of-the-art in speech processing and spoken language understanding for automatically structuring data from spoken cultural heritage archives. Technical challenges include speaker segmentation and clustering, role recognition as well as topic/story segmentation and tracking over very heterogeneous. In 2014, our research focused on unsupervised topic segmentation using social role information.

- **Very low bit-rate speech coding**

Over the last three years, Idiap has become active in the domain of very low bit-rate speech coding through a collaboration with armasuisse, in the context of the RECOD project. We have proposed a novel architecture for low bit rate coding based on neural networks. This coder differs from previous attempts by replacing HMMs by deep neural networks, and by using the phonological speech representation instead of the phonetic one. In addition, the coder further lowers the operating latency. In 2015, structured sparsity of phonological representations was discovered that led to construction of a small size codebook of sparsity patterns for very low-bitrate speech coding (supported by SNSF project PHASER); a patent application is submitted.

- **Objective speech intelligibility assessment**

In the context of armasuisse project OMSI-2015\_ ARMASUISSE<sup>25</sup> and CTI project ScoreL2, Idiap has been

actively developing novel methods to objectively assess intelligibility of speech signal. Specifically, leveraging from our posterior feature based ASR research, we have developed a novel approach where the intelligibility is objectively assessed by estimating a phonetic difference score between reference speech signal and test speech signal. A distinctive advantage of the proposed approach is that the reference speech signal can be replaced by a text transcription of the test speech signal and word level intelligibility assessment can be effectively carried out. Furthermore, the proposed approach can be applied to diverse speech assessment tasks, namely, (a) assessment of speech communication systems based on both traditional and more recent, very low bit-rate coding schemes in the presence of background noise; (b) synthetic speech intelligibility assessment; and (c) assessment of accentedness of non-native speech.

## ROADMAP FOR CONVERSATIONAL INTERACTION TECHNOLOGIES

Over the past two years, Idiap was involved in an EC funded large scale roadmapping exercise in the area of “Conversational Interaction Technologies” – ROCKIT (Roadmap for Conversational Interaction Technologies; <http://rockit-project.eu/> – which resulted in a detailed strategic (and interactive) roadmap for research and innovation in the area of multilingual conversational interaction technologies (CIT), compiled from more than 1,000 inputs from the research and innovation stakeholders.

The ROCKIT consultation brought together all of the major European stakeholders in multilingual and cross-sectorial CIT across two years of discussion about how to develop a strong innovation community with a common outlook. Our consultation resulted in a detailed understanding of what Europe could achieve and what the future demand for multilingual products, services and research will be.

A key aspect of ROCKIT was also to organise a European research and innovation community in the area of conversational interaction technologies, integrating a wide-range of commercial organisations with application and use links to the area. ROCKIT will be structured around a set of sector-based clusters including mobile applications, healthcare, education, games, broadcast media, robotics, law enforcement, and security.

<sup>24</sup> [www.idiap.ch/scientific-research/projects/sesame](http://www.idiap.ch/scientific-research/projects/sesame)

<sup>25</sup> [www.idiap.ch/scientific-research/projects/omsi-2015](http://www.idiap.ch/scientific-research/projects/omsi-2015)



## 2. Computer Vision and Learning

### OVERVIEW

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

### RESEARCH OVERVIEW

The main objective of the Computer Vision and Learning group is the development of novel machine-learning and statistical methods, with an emphasis on their algorithmic efficiency. The application domain we focus on encompasses any processing of images and video streams, for instance for object detection and scene analysis, tracking of people and biological structures, and image recognition in general.

Our current research follows two main axes. The first is the study of novel generic learning methods to cope with feature spaces of large dimension, or large training sets. The second is the development of new algorithms for scene understanding, which involves practical problems such as object detection in natural scenes, or pedestrian tracking with a multi-camera setup.

The Computer Vision and Learning group in 2015 was composed of one head of group, four PhD students and one intern, with strong interactions with the software development team.

### KEY SCIENTIFIC OUTPUTS

Our work has resulted in several novel algorithms for learning in high dimension, based on joint modeling of feature response, and a new "reservoir" strategy to leverage memory in an on-line setting. We have also developed a novel learning-based approach to pose estimation from very low quality image, and a hard-sample harvesting procedure with sub-linear computing time. In a long-lasting collaboration with EPFL, we have developed a multi-camera tracking algorithm able to combine the detection of pedestrians and vehicles in a principled manner.

### MACHINE LEARNING

- **Joint feature selection**

To classify signals using very large number of measurements, or features, it is often desirable to apply a pre-processing step

during learning which selects a "good" sub-set of them. Many criteria have been developed for that task over the years, but while it is easy to measure how good a single feature is, these techniques often fail at taking into account the joint quality of a set of features.

We have developed new methods which relies on a joint models of feature responses, under which we explicitly measure the information content of the subset of features we are selecting. Our main contribution is a "Gaussian compromise", to approximate the entropy of a mixture of Gaussian densities, for which no analytical or tractable expression exist, and that we use to approximate the mutual information.

- **Efficient clustering**

The k-means algorithm is one of the cornerstone 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 insures that it is not necessary.

In the context of the MASH2 project, 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 two. We also have developed a novel exact algorithm which is the fastest in low dimension.

In the context of the MASH2 project, 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 two. We also have developed a novel exact algorithm which is the fastest in low dimension.

In parallel to this work on exact methods, we have investigated the "batch" approaches, and proposed a new algorithm combining the triangular inequality and batch 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 are 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.

Since the performance of detectors can be extremely good, finding a sufficient number of hard samples is computationally-intensive, and in all available methods, has a cost proportional to the number of images to visit.

Our approach consists in using the prior structure provided for any large collection of images, that is its recursive decomposition into homogeneous sub-families, to concentrate computation in a principled manner. We organize the set of images into a recursive tree consistent with the similarity between images, and adapt the Monte-Carlo Tree Search to recursively apply a bandit approach to the exploration-exploitation dilemma.

In practice, our approach alternate between looking into set of images and images it has poor information about, and images from which it has observed a high success rate in finding hard samples.

• **High-dimension similarity measures**

The WILDTRACK project is 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, that is a method able to automatically figure out that a person leaving the field of view of a camera re-enters later the field of view of that camera or another one.

x

Such a problem boils down to metric learning, 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 person. In the application at hand, the challenges is to deal with changes in illumination, sensor calibration, and geometry between cameras.

The technique we have developed learns a linear projection with good metric properties, and casts this problem as a rank optimization with a regularizer that enforces the linear projection to keep a large rank. Our method allows very large scale learning and beats all existing state-of-the-art methods on standard benchmark data-sets.



**Figure 7**

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 (left). The Digit-Arena project aims at replacing ads on perimeter boards automatically in broadcast video (right).

• **Playground learning with pose-indexed features**

Related to our work on the stationary features, we are working in collaboration with the Universitat Politècnica de Catalunya, on the development of a new technique to leverage high-quality HD training images, to build models able to cope with extremely degraded test images of very low resolutions. The key idea is to use the high-quality signal to build strong priors, namely 3-D models, and to design pose-indexed features automatically from these models. Ultimately, we analytically define measurements in the images, parametrized by the pose of the object whose presence we want to check, and train a predictor from these measurements.

• **Multi-camera tracking**

We have a long-lasting collaboration with the CVLab at EPFL around the design of a robust multicamera tracking system, which now takes place in the context of the SNSF WILDTRACK project, in collaboration with ETHZ. Our most recent results consist of a new model and optimization scheme to handle properly the appearance / disappearance of "containees" (e.g. pedestrians) given the presence/absence of "containers" (e.g. cars).

Our system is composed of the Probabilistic Occupancy Map, and a flow-based convex multitarget tracker. The former estimates in individual time frames marginal probabilities of presence of people by minimizing the Kullback-Leibler divergence between the "true" posterior under a joint generative model, and a product law. The latter connects these detections properly by optimizing a flow in a spatio-temporal graph. Both components are extremely efficient computationally and process in real-time multiple camera streams to produce state-of-the-art results. Open-source implementations of both components are available under the GPL3 license.

• **Advertisement substitution in sport events**

In a collaboration with E-S. Concept. SA, we are developing a novel pipeline of algorithms to perform advertisement substitution on-the fly in broadcast video streams. The objective of this collaboration is to deliver a solution able to automatically and accurately localize perimeter advertisement during a sport event, and replace it with another one.

We have already implemented a demonstration pipeline which demonstrates the feasibility of such a system, in particular in term of stability of the board positions in the image plan.

# 3. Social Computing

## OVERVIEW

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

### GROUP OVERVIEW

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 2015 was composed of one group head, one scientific collaborator, one SNSF Ambizione Research Fellow, three postdoctoral researchers, four PhD students, one visiting faculty, and one visiting student. Research lines investigated in 2015 included: social media and phone data analytics for cities, mobile crowdsourcing, ubiquitous face-to-face interaction analysis, and analysis of Maya hieroglyphic media collections.

### KEY SCIENTIFIC OUTPUTS

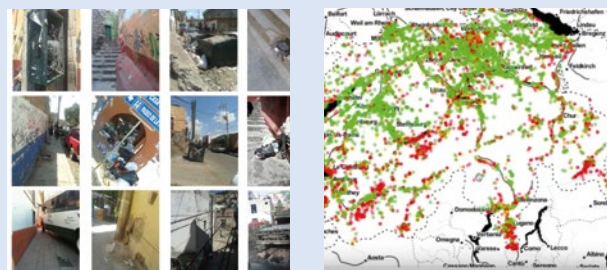
Published studies on (1) social media analytics and mobile and online crowdsourcing to document and understand urban perception in cities; (2) multimodal analysis of socioaffective attributes of social video users (YouTube) integrating audio-visual perception and video crowdsourcing; (3) multimodal inference of social constructs in employment interviews; and (4) Maya hieroglyphic retrieval integrating shape and language models. 11 EPFL PhD students graduated since 2002.

### SOCIAL MEDIA ANALYTICS AND MOBILE CROWDSOURCING TO UNDERSTAND PHENOMENA IN CITIES

Our work in this domain spans several research lines. First, we designed and implemented a mobile crowdsourcing methodology to engage youth in cities to collect multimedia items (geo-localized images and videos) that document urban

concerns like waste management, danger, or accessibility. Field studies have been implemented in multiple cities in Mexico (Figure 8). Furthermore, we are studying how urban perception can be crowdsourced and modeled to understand how sensitive citizens are with respect to such issues, and how different urban dimensions relate to each other. This work was done in the context of the SenseCityVity project<sup>26</sup> (Mobile Sensing, Urban Awareness, and Collective Action) funded by EPFL Center for Cooperation and Development (CODEV), and carried out with a multidisciplinary team from IPICYT, Mexico. This work will be expanded through a new project grant from Mexico's National Council of Science and Technology. This work has received national press coverage in Mexico<sup>27 28</sup>.

In other work, through the SNSF Youth@Night project<sup>29</sup> (A Multidisciplinary Study of Young People's Going Out and Drinking Behaviors, in collaboration with Addiction Switzerland and the University of Zurich), and the MCSC project<sup>30</sup> (Mi Casa es su Casa, supported by the EPFL-UNIL Program on Cross-Disciplinary Research on Science and Society), we are investigating the use of mobile and online crowdsourcing and social media analytics (Twitter, Foursquare, Airbnb) to characterize urban phenomena at scale in cities both in Switzerland and other world regions<sup>31</sup>. This includes the characterization of urban areas according to patterns of human activity or the state of road conditions, and the perception of social ambiance in both popular urban places and personal spaces (Figure 8). Our work in this domain has been recently covered in the media<sup>32 33</sup>.



**Figure 8**

Left: SenseCityVity: Young people in Mexico document and reflect upon urban concerns. Right: Mapping social media activity in Switzerland from Twitter and Foursquare.

<sup>26</sup> <http://www.idiap.ch/project/sensecityvity>

<sup>27</sup> <http://www.conacytprensa.mx/index.php/centros-conacyt/2143-sensecityvity-estudio-de-la-geografia>

<sup>28</sup> <http://conacytprensa.mx/index.php/>

[centros-conacyt/5198-crece-proyecto-sensecityvity-en-mexico-nota](http://centros-conacyt/5198-crece-proyecto-sensecityvity-en-mexico-nota)

<sup>29</sup> <https://www.youth-night.ch/>

<sup>30</sup> <http://www.idiap.ch/project/mcsc>

<sup>31</sup> <https://www.youtube.com/>

[watch?v=UZT4bfVnL9Q](https://www.youtube.com/watch?v=UZT4bfVnL9Q)

<sup>32</sup> <http://canal9.ch/comportements-humains-lidiap-sonde-les-smartphones/>

<sup>33</sup> <http://actu.epfl.ch/news/an-app-to-steer-clear-of-kenya-s-road-hazards/>

**SOCIAL VIDEO ANALYTICS**

This research line develops methods to characterize users of social video in sites like YouTube through the use of (1) video crowdsourcing techniques through platforms like Amazon Mechanical Turk to collect interpersonal judgments, which scales up to large amounts of data and demographically diverse viewers; and (2) automatic extraction of behavioral features including prosody, body motion, and facial expressions, which enables high-level social analysis. Our research, studying video resumes and video blogs, has shown that certain impressions of personality, mood, and hirability can be inferred to some degree from nonverbal and verbal cues. This research was supported until 2014 by the NISHA<sup>34</sup> project (NTT-Idiap Social Behavior Analysis Initiative) funded by NTT Communication Science Laboratories, Japan, and in 2015 has been done in the context of the SNSF UBImpressed project<sup>35</sup> (Ubiquitous Frist Impressions and Ubiquitous Awareness, Sinergia Interdisciplinary Program).

**FACE-TO-FACE INTERACTION ANALYTICS**

We are developing computational models to analyze dyadic and small-group interactions in workplace settings using a variety of sensors (cameras, Kinect devices, microphone arrays, smartphones, and wearables), and automatically infer key social variables for psychologists and recruiters (Figure 9). This is collaborative work with the University of Lausanne, Cornell University, and Vatel Switzerland. In the dyadic case, in the context of organizational behavior situations, we have investigated connections between hirability and personality. In the small-group case, we have studied links between behavioral cues and variables like emergent leadership and personality traits. In all cases, nonverbal cues are automatically measured from audio and video streams.

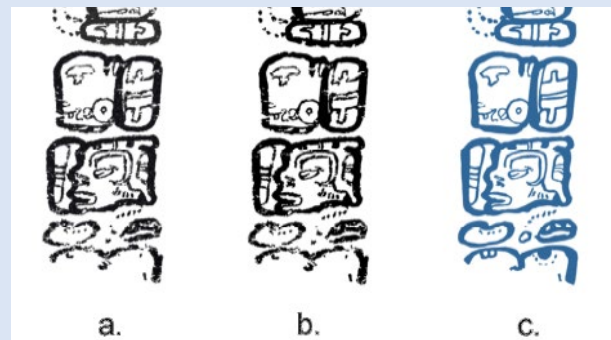
This research has been conducted in the framework of two SNSF projects, namely UBImpressed and SOBE, an Ambizione Research Fellowship. Our work in this domain has been covered by business media<sup>36</sup> and the European Commission as a H2020 Project Story<sup>37</sup>.

**VISUAL ANALYSIS OF MAYA HIEROGLYPHS**

We are developing methods for automatic visual analysis of hieroglyphs for the three most celebrated ancient Maya codices, which are located in museums in Europe – Dresden, Paris, Madrid (Figure 10).



**Figure 9**  
Sensors available in the social sensing lab. (1) Kinect V2, (2) Microcone, (3) Q-Sensor, (4) Android smartphone, (5) Google Glass.



**Figure 10**  
Preprocessing steps to generate Maya glyph data for visual analysis: (a) color images, (b) binarized images, (c) reconstructed glyphs by expert epigraphists.

Our methods are designed to support specific expert needs and include visual retrieval and classification of glyphs. This is a collaboration with Maya archaeologists from the University of Bonn and computer scientists from the University of Geneva, and is supported by the SNSF-DFG project MAAYA<sup>38</sup> project (Multimedia Analysis and Access for Documentation and Decipherment of Maya Epigraphy). This work has received recent press coverage<sup>39 40</sup>.

xii

<sup>38</sup> [www.idiap.ch/project/maaya/](http://www.idiap.ch/project/maaya/)

<sup>40</sup> [www.nzz.ch/wissenschaft/archaeologie/](http://www.nzz.ch/wissenschaft/archaeologie/)

<sup>39</sup> [www.letemps.ch/sciences/2015/11/02/un-google-translate-ecriture-maya](http://www.letemps.ch/sciences/2015/11/02/un-google-translate-ecriture-maya)

[google-translate-fuer-maya-hieroglyphen-1.18639959](http://www.nzz.ch/wissenschaft/archaeologie/google-translate-fuer-maya-hieroglyphen-1.18639959)



# 4. Perception and Activity Understanding

## OVERVIEW

**Head: Dr. Jean-Marc Odobez (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 (mainly vision and audio). This entails the investigation of fundamental tasks like the representation, detection, segmentation and tracking of people, the characterization of their state, and the modeling of sequential data and their interpretation in forms of gestures, activities, behavior or social relationships. These tasks are addressed through the design of principled algorithms extending models from computer vision, statistical learning, or multimodal signal processing. In particular, the group has focused on the recent years on the exploitation of probabilistic graphical models as one of its main modeling tools. Another objective of the group is the investigation, whenever possible, of unsupervised adaptation methods to leverage either the availability of potentially large amounts of test data from which exploitable statistics can be extracted, or of prior models that can help the soft tagging of test data, or of both. Surveillance, traffic analysis, analysis of behavior, human-robot interfaces, and multimedia content analysis are the main application domains.

In 2015, the group was composed of 2 post-doctoral members, 3 PhD students, and 1 master student.

### KEY SCIENTIFIC OUTPUTS

The group is well known for its work on video sequence analysis, probabilistic tracking and non-verbal behavior extraction. In 2006, it developed a unique video stabilization algorithm

that is currently applied automatically on all super-8 and videos digitized by the Cinetis company. In 2007, the Klewel company was created using its slide OCR technology. More recently, the work on 3D face and gaze tracking is leading to the creation of the Eyeware start-up. During the period 2011–2015, it published 20 journal papers and around 50 conference papers, and received 1 best paper and 1 best student paper awards. The group has pioneered work on unsupervised surveillance scene activity, visual attention modeling, and its recent work on gaze estimation from cheap RGB-Depth sensors is under patenting. With their LIUM partner in the EUMSSI project, the team ranked first out of 8 teams at the MediaEval Person discovery challenge in 2015.

### MULTIPLE OBJECT TRACKING

The recent work in this research line has resulted in an enhanced tracking-by-detection system with several important characteristics that make it successful in challenging conditions, and which are illustrated in Fig. 11. Given the individual object detections (humans, faces) at each instant, the problem is set as a graph partitioning problem (within a Conditional Random Field CRF framework) such that detections corresponding to the same object are associated together based on fine grained multicue measurements (position, image-based motion, multiple patch colors, set of SIFT descriptors). Important novel features of our model are the following. Links between detections can be a few seconds apart, rather than only a few frames apart. All links are used in defining the cost function, not only those between consecutive nodes from the same track. This allows assessing a solution using both short and long-term links. Importantly, pairwise cue measurement costs (defined as similarity/dissimilarity likelihood ratio) are sensitive to the time interval between two detections, integrate multi-cue association reliability factors, and all likelihood distribution parameters are learned in an unsupervised fashion, thus accounting for the data, scene, and cue specificities while avoiding tedious annotations. Results equal or better than the state-of-the-art were obtained on several benchmark data of the field, regarding both human and face tracking. The work received the best paper award at the annual IEEE Performance Evaluation of Tracking and Surveillance (PETS) workshop. The software has been licensed for research purposes to a large leading IT company active in camera manufacturing.

XIII

<sup>26</sup> <http://www.idiap.ch/project/sensecityvity>

<sup>27</sup> <http://www.conacytprensa.mx/index.php/centros-conacyt/2143-sensecityvity-estudio-de-la-geografia>

<sup>28</sup> <http://conacytprensa.mx/index.php/>

[centros-conacyt/5198-crece-proyecto-sensecityvity-en-mexico-nota](http://centros-conacyt/5198-crece-proyecto-sensecityvity-en-mexico-nota)

<sup>29</sup> <https://www.youth-night.ch/>

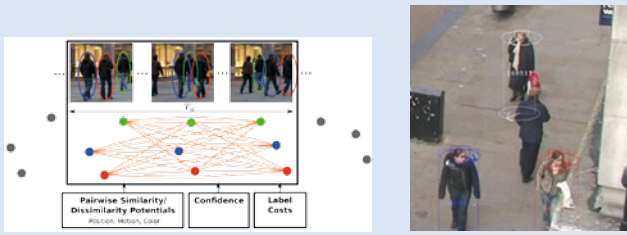
<sup>30</sup> <http://www.idiap.ch/project/mcsc>

<sup>31</sup> <https://www.youtube.com/>

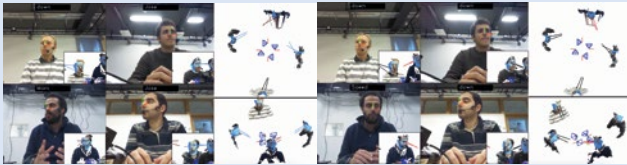
[watch?v=UZT4bfVnL9Q](http://watch?v=UZT4bfVnL9Q)

<sup>32</sup> <http://cana19.ch/comportements-humains-lidiap-sonde-les-smartphones/>

<sup>33</sup> <http://actu.epfl.ch/news/an-app-to-steer-clear-of-kenya-s-road-hazards/>



**Figure 11**  
 Left: the multi-person tracking framework. Human detections at each frame are associated based on long-term multi-cue and time-interval sensitive discriminative costs. Right: estimated head and body orientation of people.



**Figure 12**  
 Heads are tracked, eye localized and aligned before inferring gaze and attention towards people or other semantic labels. The method analyses all information in 3D (cf different renderings), allowing an easy inference of attention cues.

**BEHAVIOR CUE EXTRACTION IN SURVEILLANCE SCENES**

We have worked in recent years on body and head pose estimation (see Fig. 11, Right) which are important cues for activity analysis. Our work addressed the problem through the joint adaptation to new scene of head and body pose classifiers. The method accounted for the different coupling (between head and body pose due to anatomical constraints, between motion and body pose conditioned on a walking status). It also handled the discrepancy problems resulting from the mismatch between training and test data, using a synchronized manifold alignment approach which learn two feature projection mappings enforcing train and test features in the projected space to form a single smooth pose manifold. The method thus leverages at several modeling places the availability of weak pose labels (e.g. from people motion) in the test data to learn better pose classifiers.

**GAZE AND ATTENTION MODELING**

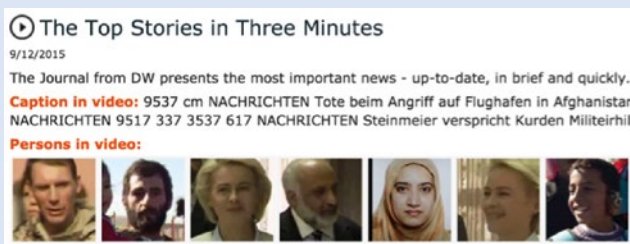
Our long-standing work on gaze and attention been continued in the context of the SNSF funded project G3E, which more specifically addresses gaze extraction from cheap RGB-D (depth) cameras like Kinect. Building on our the

head-pose free gaze estimation framework, we have proposed several extensions to enhance the performance at different levels (gaze estimation, software system, usability) and gain robustness against low eye image resolution and imprecise pose estimation, e.g. through synchronized delaunay triangulation eye alignment. The figure 12 illustrates the result of applying our tools for gaze coding in multiparty situations.

**NON-VERBAL BEHAVIOR EXTRACTION AND HUMAN INTERACTION ANALYSIS**

In addition to gaze, our work on non-verbal behavior cue extraction and human interaction modelling has continued in the context of the UBIMPRESSED project. In particular, we have worked on the 23 / 69 2 RESEARCH GROUPS 2015 Idiap Annual Report – March 2016 Figure 13: Person identification within the EUMSSI project. Left illustrating the found face clusters of a given Deutsche Welle video document (as well as automatically extracted OCR). Right: fast browsing of documents based on face occurrences. recognition of head gestures like nods which by conveying agreement, emphasis, or doubts, constitute a potential strong cue for the inference of first impressions in natural interactions. To handle the recognition of subtle nods and be immune to the camera view point, we have proposed a novel gesture detection method that leverages on our 3-D face tracker to compute (i) view-invariant relative head rotation from which rich dynamics oscillatory features based on Fourier transform are derived, (ii) rotation axis features allowing to distinguish oscillations coming from the trunk or from head oscillations around the neck. The method has been extended to the multi-class head gesture case.

In collaboration with the 'Speech and Hearing group' of KTH, we have worked on the analysis of group interaction and more specifically studied the classification of silent participants in three categories (attentive, side participants, bystander). Turn-taking, gaze tracking and head nodding cues have been automatically extracted using our tools and used to better characterize the engagement of participants. The study showed that many of the audio-visual cues (mutual gaze, backchannels) distinguishing listeners in dyadic interactions were also holding in multi-party interactions.



**Figure 13**

Person identification within the EUMSSI project. Left: illustrating the found face clusters of a given Deutsche Welle video document (as well as automatically extracted OCR). Right: fast browsing of documents based on face occurrences.

### MULTIMODAL FACE AND PERSON DIARIZATION AND NAMING

Identifying people appearing and speaking in multimedia data allows to monitor who said what and when and is thus important for the design of search and fast browsing tools of personal data collection or broadcast programs, as illustrated in Fig. 13. This is one of the challenge that we are addressing in the EU funded EUMSSI project, whose global aim is to provide journalists with text and audio visual search tools for event understanding. This year, we have improved the robustness and efficiency of the overall processing pipeline, with an emphasis on the face diarization part (face detection, tracking and clustering). In addition, we have evaluated the benefit of state-of-the-art face representations (i-vectors, neural-network), and shown that in the diarization context, matching representations based on SURF descriptors are still necessary to handle the large variety of pose and expression found in the media documents.

### VISUAL ANALYSIS OF MAYA HIEROGLYPHS

In collaboration with the Social computing group, we are working on the automatic visual analysis of hieroglyphs from the ancient Maya codices, see Fig. 10.

## 5. Robot Learning & Interaction

### OVERVIEW

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

### GROUP OVERVIEW

The Robot Learning & Interaction group, created in May 2014, focuses on human-centric robot applications. The group targets the problem of transferring skills to robots with intuitive human-robot teaching interaction. The goal is to endow robots with learning systems that can acquire rich motor skills adaptively without requiring the users to program the robots through a computer language. The objective is to develop probabilistic models of movements and behaviors for robots evolving in unconstrained environments and sharing the same workspace as humans. In these applications, the models serve several purposes (recognition, prediction and synthesis), and are shared by different learning strategies (imitation, emulation, exploration and self-refinement). The Robot Learning & Interaction group in 2015 was composed of one postdoctoral fellow, one PhD student, one visiting PhD student and two MSc students.

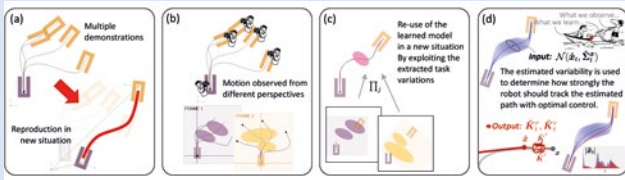
### TASK-ADAPTIVE MOVEMENT LEARNING

The group investigates the problem of analyzing and generating movements by relying on demonstrations collected in diverse situations, such as recordings with different positions of objects in a manipulation task.

### KEY SCIENTIFIC OUTPUTS

We developed an approach modeling the observed motion in multiple coordinate systems (frames of reference). The different projections of the collected data reveal regularities that are exploited to generalize and adapt the task to new situations in a probabilistic manner. Notably, the retrieved variability and correlation information in the different coordinate systems is used to regulate the stiffness of the robot in order to generate safe and natural movements that concur with the task objectives. Fig. 14 illustrates the proposed approach. It enables new haptic communication capability that has great potential in human-robot collaboration.

Aspiring to reproducible research, the group released open source codes of the proposed approach and published an accompanying tutorial article. The source codes are available in two versions (either 25 / 69 2 RESEARCH GROUPS 2015 Idiap Annual Report – March 2016 C++ or Matlab/GNU Octave)<sup>41</sup>.

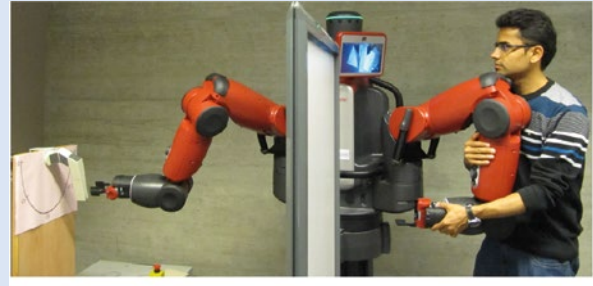


**Figure 14**

Task-parameterized probabilistic model to synthesize movements in new situations. (a) Observation of the task in different situations and generalization to new contexts. Multiple demonstrations help at extracting the structure of the task. (b) Probabilistic encoding of continuous movements in multiple coordinate systems. (c) Exploitation of variability and correlation information to adapt the motion to new situations. With cross-situational observations of the same task, the robot can generalize the skill to new situations. (d) Computation of the underlying optimal control strategy to reproduce the observed behavior.

### LEARNING FROM DEMONSTRATION FOR SEMI-AUTONOMOUS TELEOPERATION

Robots are getting closer to humans in diverse manners, sharing the users' space (humanoids at home and in manufactures, robots as co-workers), being in direct contact with the users (exoskeletons, prosthetics, rehabilitation) or inside their bodies (surgical robots). Such human-centric robot applications require a tight integration of learning and control. This connexion can be facilitated by a shared representation of the tasks and objectives to achieve in a probabilistic form. In human-robot collaboration, such representation can take various forms. Movements must be enriched with perception, force and impedance information to anticipate the users' behaviors and generate safe and natural gestures.



**Figure 15**

Teleoperation with the Baxter robot, in which one of the two arms is used as a haptic input device, and the other arm is used to imitate the movement in a semi-autonomous manner.

### KEY SCIENTIFIC OUTPUTS

The group proposed an approach to extend learning from demonstration techniques to assistive tasks and semi-autonomous teleoperation. The approach is exploited within the DexROV European Project, in which a bimanual underwater robot is distantly controlled by a user wearing an exoskeleton with force feedback. The transmission delays are handled by treating teleoperation as a joint classification, prediction and synthesis problem, all relying on a shared model encoding a set of motion, coordination and impedance primitives that can be assembled in sequence and in parallel. On the teleoperator side, the model is used to recognize existing movement/behavior primitives that were used in previous missions and predict their most likely outcomes. On the manipulator side, the same model is used to generate autonomously a movement/behavior until a new signal from the teleoperator is available. Fig. 15 shows the semi-autonomous teleoperation setup developed at Idiap to study such challenge.

<sup>41</sup> [www.idiap.ch/software/pbdlib/](http://www.idiap.ch/software/pbdlib/)



# 6. Artificial Cognitive Systems

## OVERVIEW

**Head: Prof. Barbara Caputo**  
(PhD, Royal Institute of Technology, Sweden, 2005)

### GROUP OVERVIEW

The Artificial Cognitive Systems group works on the development of multi-modal learning algorithms to enable artificial agents to act autonomously in realistic settings, with a special emphasis on the ability to autonomously detect knowledge gaps and fill them autonomously with open-ended learning strategies. The focus of this work is on designing algorithms that are principled, computationally efficient, and that provide robust performance in very realistic settings while at the same time providing theoretical guarantees on expected behavior. The group is funded by national and international funding agencies.

The Artificial Cognitive Systems group in 2015 was composed of 1 head of group and 2 PhD students.

### KEY SCIENTIFIC OUTPUTS

The group is internationally recognized for its work on life long learning, adaptive control of dexterous prosthetic hands and benchmarking activities. In 2010, we developed the first max-margin based transfer learning algorithm for visual detection. The algorithm is the current state of the art and is considered a reference in the field. Since 2011 we have been creating the first publicly available database of surface ElectroMyoGraphy (sEMG) recordings of hand postures from intact and amputated subjects. The database is the largest existing for number of postures and number of subjects recorded, and it has become the reference benchmark in the field. From 2007 to 2013 we organized the Robot Vision Task challenge under the ImageCLEF umbrella, which has seen the participation of more than 100 teams over the years. As a further recognition

of the value of our benchmarking activities, during 2013 and 2014 Prof. Caputo was the main organizer of ImageCLEF. In 2014, Prof. Caputo received an ERC Starting Grant.

### SEMANTIC SPATIAL UNDERSTANDING OF INDOOR PLACES

The work on semantic spatial understanding of indoor places, started in 2005, has been continued in the context of several projects, including the VISION@HOME project, funded by the Swiss National Science Foundation. It proposes the innovative Situated Vision paradigm, developing 3D visual perception capabilities from the view of the robot, its task and the environment it operates in.

The Situated Vision approach is inspired by recent work in cognitive science, neuroscience and interdisciplinary work in EU projects: it fuses qualitative and quantitative cues to extract and group 3D shape elements and relate them to affordance categories. Task-oriented and 3D attention mechanisms let the robot execute primitive actions to exploit the affordances perceived. Perception integrates quantitative and qualitative shape information from multiple 2D and 3D measurements. The analysis of the shapes is used to find instances of semantic 3D concepts, such as providing support to objects, enclosing space, etc. that can be used to those spatial concepts to find semantic entities, such as table surfaces, cupboards, closets, drawers and to learn which perceived affordances belong to which object category.

Over the last year, we casted the semantic spatial modeling problem into that of learning from multiple cues. Our contributions have been a principled online Multi Kernel Learning algorithm able to combine optimally multiple features while providing theoretical guarantees on the expected performance, and a general learning to learn algorithm, based on the cue integration paradigm, able to leverage over prior

---

<sup>42</sup> <http://www.idiap.ch/project/ninapro/>

knowledge in an extremely flexible and adaptive manner, over a wide range of application scenarios. This last algorithm, the first of its kind in the literature, achieves state of the art results on benchmark databases widely used in the literature and is currently implemented in a running robot system.

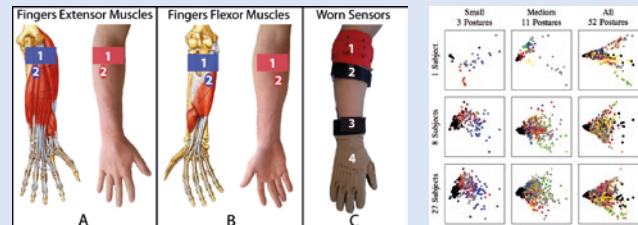
### ADVANCED HAND PROSTHESES

The work on advanced hand prostheses has been initiated in 2011 with the SNSF funded NINAPRO<sup>42</sup> project. Its goal is to develop a family of algorithms able to significantly augment the dexterity, and reduce the training time, for sEMG controlled prosthesis. Indeed, the state of the art in hand prosthetics does not offer more than 2–3 degrees of freedom and a very coarse control of the force, as there is no haptic feedback. Patients interface with the prosthesis via surface electromyography (sEMG), recorded using surface electrodes. Learning how to control the device through many input sEMG channels is a long and difficult process for most patients, that therefore settles for limited and very simplified movements (open/close).

This contrasts with recent advances in mechatronics, thanks to which mechanical hands gifted with many degrees-of-freedom and force control are being built. There is a need for prosthetic hands able to naturally reproduce a wide amount of movements and forces, while at the same time requiring a lower effort in learning how to control hand postures. This goes beyond mechatronic dexterity: the real challenge is how to provide patients with a cheap, easy and natural way of controlling the prosthesis. By testing its findings on a very large collection of data, NINAPRO has the ambition to pave the way for a new generation of prosthetic hands. The work is organized along four themes: data acquisition and analysis (Theme 1), augmented dexterity: posture classification (theme 2), augmented dexterity: natural control (theme 3) and adaptive learning (Theme 4). Prof. B. Caputo is the coordinator of the project, and the principal investigator in the Themes 2 and 4 (Figure 16).

The objective of theme 2 (augmented dexterity) is to push the current state of the art in prosthetic hand posture classification from handling a maximum of 12 postures up to 40–50. Over the last year, we have demonstrated the importance of using multi modal features to increase the robustness of posture

classification and therefore of actual control of the prosthesis, by merging information acquired from sensors applied on the stump of the user with those recorded by an accelerometer mounted on the prosthesis. This work has resulted in a publication on the international journal IEEE Transaction on Neural Systems and Rehabilitation Engineering.



**Figure 16**

One of the most important results of the NinaPro project has been the creation of the very first open access database of sEMG recordings of hand and arm movements from intact subjects and amputees. As of today, the database contains recording of more than 50 postures from over 100 intact subjects and 15 amputees. On the left, we show the placement of sensors with respect to the muscles activations; on the right, a probabilistic analysis of a first batch of data, illustrating as the signals acquired from the different subjects become more and more difficult to interpret as the number of subjects and hand postures grows.

The goal of theme 4 (adaptive learning) is to develop learning algorithms to better interpret the sEMG signals acquired from the patients, with the ultimate goal of boosting the learning process necessary for the patient to effectively use the prosthesis. Our plan is to build pre-trained models of various data postures, on the data acquired in theme 1, and adapt these general models to the needs of individual users as new data will become available using adaptive online learning methods. Over the last year, we did build over our results obtained in the large-margin classification framework, and we explored how to maintain the stability and effectiveness of our method when in presence of a very large number of prior models. This would be indeed the case when leveraging over priors derived from the actual experience of intact subjects, from which it is relatively easy to acquire large amounts of data and hence build strong prior models. We developed a new version of our previous algorithm able to scale well over a very large number of priors in terms of computational efficiency, while preserving a very high accuracy in terms of hand posture recognition.

<sup>42</sup> [www.idiap.ch/project/ninapro/](http://www.idiap.ch/project/ninapro/)

# 7. Uncertainty Quantification and Optimal Design

## OVERVIEW

**Head: PD. Dr. David Ginsbourger (Ph.D. Mines Saint-Etienne 2009, Habilitation Univ. Bern 2014)**

### GROUP OVERVIEW

The Uncertainty Quantification and Optimal Design group focuses on quantifying and reducing uncertainties in the context of natural and artificial complex system modelling. 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 high-fidelity experiments and/or numerical simulations depending on a number of inputs including both controlled design parameters and uncontrolled environmental variables.

Uncertainty Quantification comprehends a set of statistical approaches with the aim to model, elicit and characterize uncertainties arising in the study of complex systems, be it relying on data, numerical simulation and/or on expert knowledge. Optimal Design essentially deals with the choice of controlled design parameters relying on global optimization methods. In particular, elaborated model evaluation strategies are needed for making optimal decisions, both for deterministic numerical simulations and for design under uncertainty.

More specifically, the main focus of the UQOD group is on stochastic models and adaptive design of experiments for optimization, inversion, and related problems. Typical stochastic models studied within UQOD encompass spatial random fields and times series, notably with Gaussian Processes but also with an ongoing interest for non-stationary extreme value modelling.

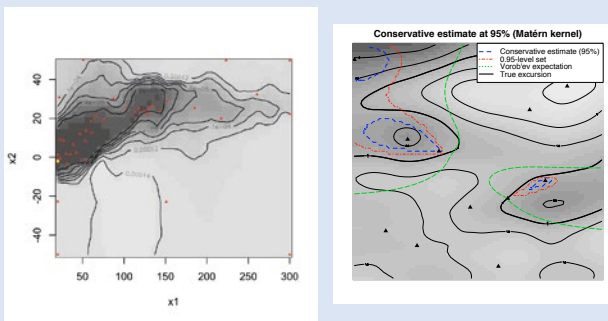
UQOD started at Idiap in September 2015, with members coming from and keeping strong academic ties to the Institute of Mathematical Statistics and Actuarial Science (IMSV) of the University of Bern. In its initial form it was composed of one head of group, one IMSV postdoc and one IMSV PhD student.

## KEY SCIENTIFIC OUTPUTS

Current contributions include efficient algorithms for Bayesian set estimation, notably for estimating and quantifying uncertainties on overcritical parameter regions and other implicitly defined sets of interest with Gaussian random field models. Other recent results deal with the interplay between the choice of the covariance kernel and degeneracy/invariances properties of Gaussian random field paths, with implications in high-dimensional Gaussian Process modelling and in function prediction under structural constraints. Ongoing work also encompasses novel algorithms for non-stationary modelling of extremes with application in climate sciences, as well as a collaboration with hydrogeologists towards simulation-based optimal placement of decontamination pumping wells.

## DETERMINISTIC AND STOCHASTIC FUNCTION MODELLING WITH KERNEL METHODS

Kernel methods, and in particular Gaussian Process (GP) / Gaussian random field models, have become quite popular in the last decades for approximating costly-to-evaluate functions relying on evaluations at a finite set of points. Often, kernels used in such methods are taken among off-the-shelf families, typically isotropic or anisotropic stationary kernels, e.g. from the Matérn family. On the other hand, in a number of problems some expert knowledge is available about the function to be approximated, e.g. being zero-mean, possessing given symmetries and invariances, being harmonic, etc. It turns out that these properties can be cast as degeneracies under linear operators, and so kernels can be designed that incorporate such structural properties, leading both to a predictor honouring the expert knowledge but also to a posterior distribution of elements honouring it as well. Previously and recently obtained results along this line are currently being resumed, with the aim to explore ways of putting more expressiveness in GP models. In particular, links to global sensitivity analysis are considered, paving the way to original high-dimensional GP-based function emulation approaches.



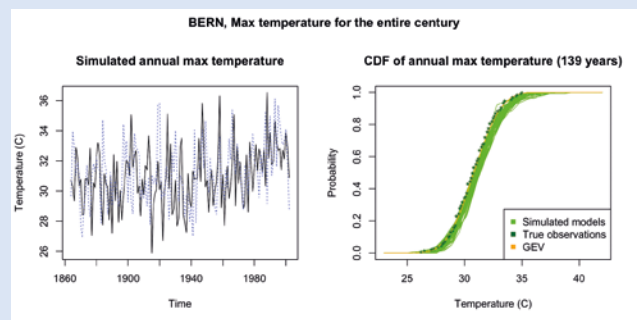
**Figure 20**  
 Surrogate-based minimization of the highest concentration (over time) escaping the right outlet of a domain as a function of coordinates of a remediation pumping well (left). Right: conservative excursion set estimation under a Gaussian random field prior.

Also, as illustrated in Fig. 20 (left panel), the group's know-how in surrogate-based strategies for global optimization is put to good use with a simulation-based application in well placement for decontamination set up together with hydro-geologists from the Universities of Lausanne and Neuchâtel. Here a response quantifying how bad some contaminant escapes from a given area is minimized as a function of the location of a candidate pumping well. The yellow point stands for the optimal placement (obtained from a very time-consuming extensive grid search) while the red points are points explored by a state-of-the-art global optimization algorithm relying on a GP model. Comparisons with other algorithms and extensions to more input variables (e.g., placing several wells) are currently under study. Also, extensions of GP-based optimization algorithms to the case of larger number of points than what basic GP models can accommodate is part of current research topics of the group.

**EFFICIENT ALGORITHMS FOR PROBABILITY OF EXCURSION AND CONSERVATIVE SET ESTIMATION**

In a number of practical problems from natural sciences or the industry that involve costly-to-evaluate functions, the goal is not exactly to optimize the function but more precisely to locate regions of the input space leading to a response above or below a given threshold, that correspond for instance to a dangerous or abnormal behaviour of the considered system. Gaussian random field models have been increasingly used to approximate such excursion sets, and also to quantify and reduce uncertainties on them by notably using stochastic simulation techniques and sequential evaluation strategies. Recently, approaches have been proposed for conservative estimation of such sets; conservative here means that under the considered model the actual excursion set is with a given probability

completely inside (or outside, depending on the way the problem is formulated) the estimate returned by those methods. Hence deriving such estimates necessitates to calculate the probability that the maximum of a high-dimensional vector (say, in a discretized case) remains below or above a given threshold. While algorithms have been proposed to calculate such probabilities, we have been working on a special Monte Carlo technique to reduce the overall computational cost of this kind of calculation, leading in particular to speed ups in the computation of conservative excursion set estimates. The right panel of Fig. 20 presents the 95% conservative excursion set estimate obtained for a Gaussian random field simulation test case, and a comparison with the actual excursion set and other candidate estimates which level of confidence is substantially lower than 95%.



**Figure 21**  
 Left: Two realizations of annual max temperature at Bern for the horizon 1864-2002 under a proposed stochastic model. Right: Corresponding simulated (longitudinal) empirical CDFs together with the empirical CDF of observed yearly maxima and a Generalized Extreme Value distribution fitted to them.

xx



## EXTREME VALUE ANALYSIS FOR NON-STATIONARY TIME SERIES WITH APPLICATION IN CLIMATE SCIENCES

Extreme value analysis is a key statistical approach to a number of phenomena from nature, economic and industry activities. In order to assess risks and make design or policy decisions accordingly, it is of importance to understand and forecast how extreme some quantities of interest may be or become, be they for instance climatological or economical variables. The theory of extreme value distributions has been studied for numerous decades, and statistical tools are available to fit so-called "Generalized Extreme Value" (GEV) distributions to observed maxima and quantify how severe and with what probability extreme events may occur. However, the most spread usage of GEV assumes a sample of independent identically distributed maxima in order for the underlying mathematical theory to be valid. Some attempts have been done to fit GEV distributions in non-stationary conditions, but a number of practical and research questions remain open. This poses serious challenges to climate scientists and beyond, as phenomena turn out to exhibit dependences and also to change over time. How then to speak about extremes, e.g. about return levels, based on non-stationary sequences of observations?

In approaches currently being developed within the UQOD group and in collaboration with the Oeschger Center for Climate Change Research we notably consider times series data at several scales, departing from the plain block-wise maxima approach where distributions are fitted to a sample of yearly maxima while usually neglecting what is happening at a finer temporal frequency. Our investigated algorithms for characterizing and forecasting extremes of times series essentially rely on state-of-the-art time series decompositions, on Monte Carlo simulations, and on statistical learning prediction algorithms.

# 8. Computational Bioimaging

## OVERVIEW

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

## 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 2015, the Computational Bioimaging Group was composed of the head of group and one trainee.

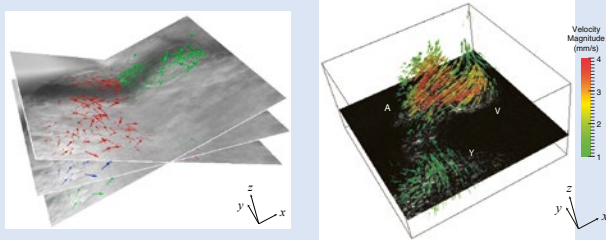
## KEY SCIENTIFIC OUTPUTS

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.

## QUANTIFYING 3D BLOOD-FLOW AT THE MICROSCOPIC LEVEL

Conventional fluid flow estimation methods for in vivo optical microscopy are limited to two-dimensions and are only able to estimate the components of flow parallel to the imaging plane. This limits the study of flow in more intricate biological structures, such as the embryonic zebrafish heart, where flow is three-dimensional. To measure three-dimensional blood flow, we have proposed, together with UCSB student and Idiap intern Kevin Chan, an algorithm to reconstruct a 3D, divergence-free flow map from multiple 2D flow estimates computed from image stacks captured from different views. This allows us to estimate the out-of-plane velocity component that is normally lost with single-view imaging. This paper describes our 3D flow reconstruction algorithm, evaluates its

performance on a simulated velocity field, and demonstrates its application to in vivo cardiac imaging within a live zebrafish larva.



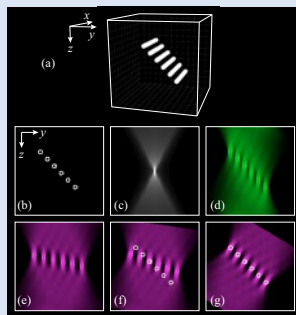
**Figure 22**

Left: We acquired focal stacks of heart in a zebrafish larva from three views:  $-18^\circ$ ,  $0^\circ$ ,  $18^\circ$  relative to the  $xy$ -plane and rotated about the  $y$ -axis. At each view, 2D optical flow was used to estimate velocity vectors in each plane. For visibility, a single slice is shown for each view. Right: We combine 2D flow estimates from three different views to recover a divergence-free 3D velocity map of blood flow through the heart (A: atrium, V: ventricle, Y: yolk). Adapted from Chan et al., ISBI 2015.

### IMAGE PROCESSING METHODS FOR MULTI-VIEW MICROSCOPY

Optical microscopy offers the unique possibility of imaging living samples in 3D under conditions similar to their native state. However, most techniques suffer from anisotropic 3D resolution (worse in the axial than the lateral dimensions). This makes quantification of images more difficult. Fusion of multiple datasets acquired from different view points has the potential to overcome this limitation. Together with UCSB graduate student Nikhil Chacko we developed an algorithm to spatially register two volumetric datasets related via a rigid-body transform and degraded by an anisotropic point-spread-function (PSF), as a first step towards solving the fusion task. Our automatic registration approach, a multi-scale pixel-based registration method accounts for the anisotropic image formation process by including a re-blurring operation before minimizing the mean squared intensity difference between a reference and a test volume.

XXII



**Figure 23**

(a) 3D perspective of the object used in simulations. (b-g) Maximum intensity projection (MIP) in the  $yz$  plane of  $f$ ,  $h$  (after gammacorrection of  $= 0:3$ ),  $fR$ ,  $fT$ , the registered results using the traditional, and our proposed approach, respectively. For comparison, (b) has been overlaid on the registered results shown in (f) and (g). Scale bar is 25m: Adapted from Chacko et al. ISBI 2015.

## 9. Biometric Person Recognition

### OVERVIEW

Head: Sébastien Marcel (PhD, University of Rennes, France, 2000)

### GROUP OVERVIEW

The Biometrics group investigates and develops novel image-processing and pattern recognition algorithms for face recognition (2D, 3D, and near-infrared), speaker recognition, vein recognition, anti-spoofing (presentation attack detection), and emerging biometric modes (EEG and remote photoplethysmography). The group is geared toward reproducible research and technology transfer, using its own signal-processing and machine-learning toolbox. The Biometric Person Recognition group in 2015 was composed of 1 head of group, 2 scientific collaborator, 4 postdocs, 2 PhD students, few interns and visitors.

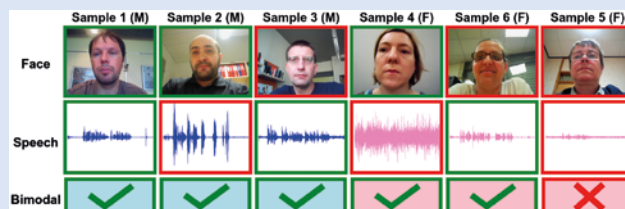
### KEY SCIENTIFIC OUTPUTS

The Biometrics group has been pioneering the work on mobile biometrics (face and speaker recognition) and on anti-spoofing in face recognition by sharing the first open databases, organizing the first International competitions and producing the first reproducible research studies in the domain. In mobile biometrics, the group demonstrated that the same probabilistic session variability modeling techniques can be applied both to speaker and to face recognition. The group proposed also an exact and scalable formulation of Probabilistic Linear Discriminant Analysis (PLDA), a state-of-the-art technique in speaker recognition, hence solving a major limitation. In face anti-spoofing, the group demonstrated that the current trend in discriminant-based anti-spoofing is prone to over-fitting hence resulting in a lack of generalization on unseen spoofing attacks.

### FACE AND SPEAKER RECOGNITION

We leveraged prior work on distribution modeling for part-based face recognition using session variability modeling techniques. Session variability modeling aims to explicitly

model and suppress detrimental within-class (inter-session) variation. We examined two techniques to do this, inter-session variability modeling (ISV) and joint factor analysis (JFA), which were initially developed for speaker recognition. Next, we explored Total Variability modeling (TV), so called i-Vectors originally proposed for speaker recognition, for the task of face recognition, and we reached boosted performance for the task of bimodal face and speaker recognition (Figure 24).



**Figure 24**

An illustration of face and speaker 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. However, it has one major drawback, at training time estimating the latent variables requires the inversion and storage of a matrix whose size grows quadratically with the number of samples for the identity (class). We developed a scalable derivation which is theoretically equivalent to the non-scalable solution and so obviates the need for a variational approximation. The main advantage of the proposed approach is a gain in performance when using more training samples per identity (class), which is made possible by the proposed scalable formulation of PLDA. As a consequence, on July 31 2014 the **biometrics group was ranked 1st** among 130+ participants to the NIST i-Vector Machine Learning challenge on Speaker Recognition ( [ivectorchallenge.nist.gov](http://ivectorchallenge.nist.gov)) using this novel PLDA combined with an efficient hierarchical clustering.

### ANTI-SPOOFING

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 threat is spoofing attacks: an action of outwitting a biometric sensor by presenting a counterfeit biometric evidence of a valid user. 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 spoofing attacks: a biometric system is usually designed to only recognize 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 anti-spoofing schemes for them has attracted much less attention. Recently, we advanced the state-of-the-art in anti-spoofing to printed photo attacks in face recognition by proposing public benchmarks and effective solutions based on texture modeling, motion detection or spatio-temporal pattern recognition. We also organized the first International Competition on the topic.

**REPRODUCIBLE RESEARCH**

Over the last year the group consolidated a major evolution of the Bob ([www.idiap.ch/software/bob/](http://www.idiap.ch/software/bob/)) software library to support its research activities. Bob is a free signal processing and machine learning toolbox. The toolbox is designed to meet the needs of researchers by reducing development time and efficiently processing data. Firstly, Bob provides a researcher-friendly Python environment for rapid development. Secondly, efficient processing of large amounts of multimedia data is provided by fast Cpp implementations of identified bottlenecks. The Python environment is integrated seamlessly with the Cpp library, which ensures the library is easy to use and extensible. Thirdly, Bob supports reproducible research through its integrated experimental protocols for several databases. Finally, a strong emphasis is placed on code clarity, documentation, and thorough unit testing. Bob is thus an attractive resource for researchers due to this unique combination of ease of use, efficiency, extensibility and transparency. Bob is an open-source library and an ongoing community effort.

**REPRODUCIBLE RESEARCH**

Within the EU BEAT project<sup>50</sup>, coordinated by Idiap, we built a biometry-independent platform for Biometrics research, development and certification. By making use of such a system, academic or industrial partners enable users to easily compare results from distinct algorithms or/and parameterizations with minimal interaction using one or potentially many biometric traits. Here are three practical examples:

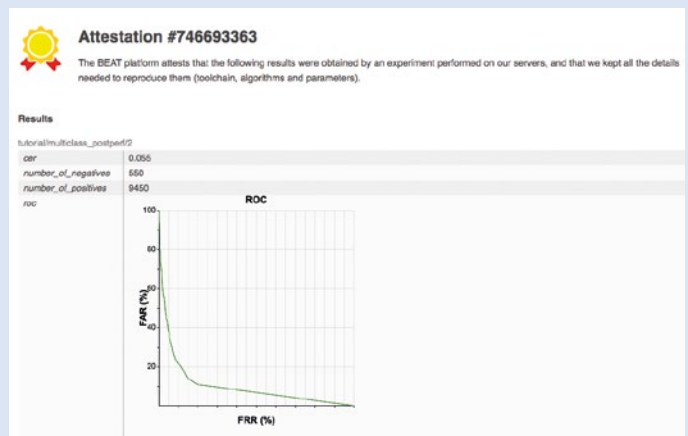
1. A teacher asks students to implement a specific illumination normalization algorithm in a preconfigured face recognition tool-chain. The teacher would like to receive the examination results from his group of students;
2. A researcher (PhD student or Professor) wishes to test new

ideas for pre-processing (i.e. to solve translation, rotation, and scaling problems due to different positions and downward pressure of the finger) in a fingerprint recognition system. She/He wishes to upload the new algorithm to an existing fingerprint-recognition tool-chain and measure its impact on the overall system performance;

3. A scientific challenge is organized in which users contribute with features that will be used in a pre-configured key-point localisation algorithm, for face recognition. The user wishes to upload his/her own feature-extraction algorithm to the contest and observe how that performs.

One can use the platform to generate attestation (a.k.a. certificates) to any produced experimental results. Attestation can be used to support formal certification processes or improve evidence on published academic articles (Figure 25).

XXIV



**Figure 25**  
A snapshot of a BEAT attestation.

**SWISS CENTER FOR BIOMETRICS RESEARCH AND TESTING**

In 2014, the Idiap Research Institute launched the "Swiss Center for Biometrics Research and Testing" ( [www.biometrics-center.ch](http://www.biometrics-center.ch) ), a competence center within the Institute following recent successes in coordinating International research projects in Biometrics (MOBIO, TABULA RASA and BEAT). The aim of this center is to serve as a legacy for these

<sup>50</sup> [www.beat-eu.org/](http://www.beat-eu.org/)



projects and to push for industry-driven research and testing in biometrics. The center will mainly promote two instruments: (1) the coordination of a cooperative research consortium, and (2) the deployment of the BEAT evaluation platform. These instruments were presented on November 2014 during a kick-off to researchers and stake-holders in biometrics (40+ participants). Consequently to this kick-off, we decided to elevate the visibility of the cooperative research consortium to a more European level. We concluded an agreement with the European Association for Biometrics (composed of 160 members) to transform the cooperative research consortium initiated in Switzerland into an European cooperative research consortium but still coordinated in Switzerland by the center. We expect to attract a critical mass of partners to fund the European cooperative research consortium. Regarding the BEAT evaluation platform, we were able to speed-up its development for a deployment late 2015. As a consequence, we have been contacted already by institutions and companies to use the platform for evaluation and hence we plan to monetize the platform by providing commercial licenses. Finally, all these activities related to the center attracted the attention of large companies and led to a license agreement with Safran Morpho, a research agreement with Google and an US IARPA project.

## 10. Natural Language Processing

### OVERVIEW

**Heads: 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 important 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 semantic relatedness, disambiguation of pronouns or connectives, keyword extraction, or sentiment analysis – and guides these improvements so that they become useful to the two applications, MT and IR. Regarding the first domain, the NLP group combines text-level processing techniques with phrase-based statistical MT systems to improve translation quality. Regarding the second task, the NLP group exploits content and sentiment analysis to improve document and multimedia recommendation, in realtime or not, over networked repositories.

XXV

The NLP group had in 2015 the following members: the head of the group; two postdoctoral students; four PhD students (including one visitor); and one intern. A PhD thesis was defended in autumn 2015 and another one is scheduled for spring 2016.

### KEY SCIENTIFIC OUTPUTS

We have demonstrated that semantic and coherence-based analyses of discourse connectives, verb tenses, noun phrases and pronouns in multi-sentence texts are beneficial to statistical MT, an innovation that has been met with interest from the community. We have integrated novel algorithms for keyword extraction and document ranking, which reward the diversity of results, into a system that recommends Wikipedia pages in real-time to conversation participants. We have shown that sentiment analysis of user comments improves recommendation, and that it is possible to predict individual sentiment aspects and relate them to text excerpts that explain the predictions using multiple instance regression; moreover, the learned

weights of the excerpts are predictive for segmentation and summarization of comments. These achievements have been demonstrated at competitive evaluations such as the ACM Multimedia Grand Challenge, the MediaEval hyperlinking task, and the EMNLP DiscoMT shared task on pronoun translation. They have also been used for technology transfer with startup companies.

### TEXT-LEVEL MACHINE TRANSLATION

We focus on a problem that is not often targeted in the current statistical machine translation (SMT) paradigm: 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), pronouns and noun phrases (our more recent focus), and verb tenses (studied in 2013–2014). Connectives and verb tenses have been mainly addressed in the COMTIS SNSF Sinergia project<sup>51</sup>, and completed in the first year of the MODERN SNSF Sinergia project<sup>52</sup>. The main focus of MODERN is on the coherent translation of noun phrases and pronouns. Both projects are coordinated by the head of the NLP group, with teams from Geneva, Zürich and Utrecht.

xxvi

The negative impact of wrong pronoun English/French translation is illustrated with the following example, translated by online SMT. In the source, the pronoun *it*<sub>3</sub> refers to the mouse, which is feminine in French, but is translated by the SMT into masculine *le*<sub>3</sub>, making the readers think that the cat (masculine in French) was hurt and that the writer intends to kill it.

#### Source text:

My cat brought home a mouse that he hunted, and *it*<sub>1</sub> was not dead but *it*<sub>2</sub> was mortally wounded. What is the best way to kill *it*<sub>3</sub> humanely?

#### Machine translation:

Mon chat a ramené à la maison une souris qui il a chassé, et *\*il*<sub>1</sub> était pas mort, mais *\*il*<sub>2</sub> a été mortellement blessé. Quelle est la meilleure façon de *\*le*<sub>3</sub> tuer humainement?

#### Correct:

Mon chat a ramené à la maison une souris qu'il a chassée, et *elle*<sub>1</sub> n'était pas morte, mais *elle*<sub>2</sub> était mortellement blessée. Quelle est la meilleure façon de *la*<sub>3</sub> tuer humainement?

In 2015, we addressed the problem of noun and pronoun coherence in translation as follows. (1) We enforced the consistency of translations of nouns that appear first in a compound and then alone, and demonstrated that this simple strategy improves their translation from German and Chinese into English. Later, we have generalized this approach with a classifier that hypothesizes, when the same noun is repeated in a source text, whether the first translation should replace the second one, or viceversa, or no such post-editing should be done. (2) We designed two systems for the DiscoMT 2015 shared task<sup>53</sup>. The system for pronoun prediction does not rely on coreference resolution, and uses instead a large set of source-side and target-side features to train a pronoun predictor. The system for pronoun-focused translation combines two knowledge sources: grammatical constraints from the hypothesized coreference links, and candidate translations from an SMT decoder. The latter system was ranked first among seven participants at the DiscoMT shared task. (3) Improving on this successful system, we designed a solution for coreference-aware decoding, which uses a coreference model with information about the anaphoric pronouns and their likely antecedents. Scores were thus improved on the DiscoMT 2015 task. (4) To evaluate automatically the translation of pronouns, we compared them automatically with those in a reference translation with the APT metric (Accuracy of Pronoun Translation), and showed that it correlates well with human judgments.

Moreover, to explore text-level SMT coherence even further, we conducted experiments that re-rank the hypotheses of an SMT system based on their coherence with the topics inferred from previous sentences, as modeled using a topic-aware keyword extraction method previously designed in the NLP group. Similarly, re-using previous work on sentiment analysis, we explored sentence-level and word-level constraints on the sentiment of target words to re-rank MT hypotheses. Finally, through the work of a visiting PhD student, we conducted experiments with speech-based editing of SMT output, showing that features such as ASR confidence and word length are relevant to decide which hypothesis (ASR or MT) to select.

<sup>51</sup> 2010–2013, see [www.idiap.ch/project/comtis](http://www.idiap.ch/project/comtis)

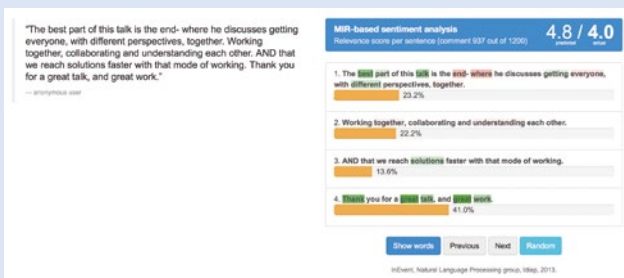
<sup>52</sup> 2013–2016, see [www.idiap.ch/project/modern](http://www.idiap.ch/project/modern)

<sup>53</sup> The 2nd Workshop on Discourse and Machine Translation (at EMNLP), co-chaired by the head of the NLP group.

TEXT ANALYSIS FOR MULTIMEDIA RECOMMENDATION

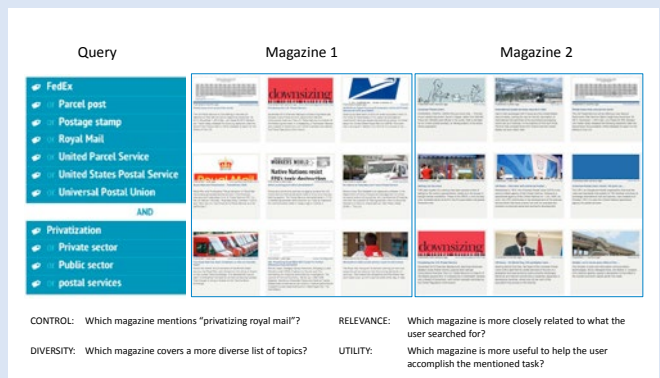
We developed methods for multimedia information retrieval and recommendation based on the analysis of content and on the networked structure of social media repositories. This work is supported by the InEvent EU project<sup>54</sup>, TheArk foundation<sup>55</sup>, and the CTI/KTI<sup>56</sup>. A PhD thesis started within the IM2 NCCR and the AMI/AMIDA EU projects<sup>57</sup> was defended in 2015, on diverse keyword extraction from conversations, and the use of these keywords to perform just-in-time document recommendation in conversations. The theoretical contributions have been integrated into an end-to-end system (ACLD, the Automatic Content Linking Device), for which a brainstorming task and evaluation metrics have been designed; the system was favorably rated for usability in evaluation experiments.

At the end of 2015, a second PhD thesis had a successful private defense at EPFL, mainly supported by the inEvent project. The thesis proposes novel methods based on machine learning to model user sentiment and preferences for information filtering systems. In 2015, the problem of modeling user attitude toward aspects of items was solved using a weighted multiple-instance learning model. The saliency weights learned by the model are useful features for review segmentation and summarization. Experiments were conducted on real-world data collected from online social networks or Web communities such as TED, Flickr, Vimeo, or Audible.



**Figure 26**  
Sentence relevance weights hypothesized using multiple-instance learning for predicting the sentiment of a user comment related to a TED talk.

Working on personalized search and recommendation for information watch with Faveeo SA, we designed implicit query expansion solutions to increase the relevance of the ranked list of results for complex queries. Within a hybrid semantic search framework, we proposed a two-step retrieval and reranking strategy and compared two methods for query expansion, one based on word co-occurrence in the document repository, and the other one using semantically-related expansion terms drawn from external resources. We evaluated our methods and compared them with the Faveeo baseline by crowdsourcing user-centric tasks on three evaluation dimensions: relevance, diversity, and utility, as illustrated in Figure 27. The results show that our proposed hybrid semantic search approach achieves better results in terms of relevance and overall utility than the initial baseline, with word co-occurrence outperforming semantic query expansion. In addition, we extended word co-occurrence methods to the recommendation of explicit keywords for query expansion to users, and also applied implicit query expansion to spoken queries in the ACLD framework.



**Figure 27**  
Sentence relevance weights hypothesized using multiple-instance learning for predicting the sentiment of a user comment related to a TED talk.

<sup>54</sup> www.inevent-project.eu

<sup>55</sup> Technology transfer project with Unono.net, a university social network.

<sup>56</sup> Technology transfer project with Faveeo SA, an information watch company.

<sup>57</sup> See www.im2.ch, www.amiproject.org and www.amidaproject.org, finished in 2013.

# Projects in progress during 2015

## PROJECTS FUNDED BY EUROPEAN AGENCIES

**Name** BEAT (Biometrics Evaluation and Testing)  
**Funding** FP7 – STREP – SEC  
**Coordinator** Idiap Research Institute  
**Duration** 01.03.2012 – 28.02.2016  
**Partner(s)** École Polytechnique Fédérale de Lausanne, Commissariat à l'Énergie Atomique, University of Surrey, Universidad Autónoma de Madrid, TÜViT, TUBITAK Informatics and Information Security, Advanced Technologies Research Center, Sagem Securite S.A., Katholieke Universiteit Leuven

**Name** EUMSSI (EUMSSI – Event Understanding through Multimodal Social Stream Interpretation)  
**Funding** FP7 – IP – ICT  
**Coordinator** Fundacio Barcelona Media  
**Duration** 01.12.2013 – 30.11.2016  
**Partner(s)** Video Stream Networks, S.L., Université du Maine, Gottfried Wilhelm Leibniz Universität Hannover, Idiap Research Institute, Gesellschaft zur Förderung der angewandten Informationsforschung EV, Deutsche Welle

**Name** DBOX (D-Box: A generic dialog box for multi-lingual conversational applications)  
**Funding** Eurostars Programme  
**Coordinator** Mi'pu'mi Games GmbH  
**Duration** 01.11.2012 – 31.10.2015  
**Partner(s)** Universität des Saarlandes, SIKOM Software GmbH, Koemei SA, Idiap Research Institute, Acapela

**Name** ROCKIT (Roadmap for Conversational Interaction Technologies)  
**Funding** FP7 – CSA – ICT  
**Coordinator** University of Edinburgh  
**Duration** 01.12.2013 – 30.11.2015  
**Partner(s)** Vodera Limited, Universität des Saarlandes, LT-Innovate, Idiap Research Institute, Europe Limited SA

**Name** DEXROV (Effective Dexterous ROV Operations in Presence of Communications Latencies)  
**Funding** H2020 – RIA – BG  
**Coordinator** Space Applications Services  
**Duration** 01.03.2015 – 31.08.2018  
**Partner(s)** Interuniversity Center of Integrated Systems for the Marine Environment, Jacobs University Bremen, Idiap Research Institute, Graal Tech s.r.l., COMEX SA.

**Name** SIIP (Speaker Identification Integrated Project)  
**Funding** FP7 – SRTEP – SEC  
**Coordinator** Verint System Ltd  
**Duration** 01.05.2014 – 30.04.2018  
**Partner(s)** Cassidian SAS, Laboratorio Di Scienze Della Cittadinanza, Synthema S.R.L, Idiap Research Institute, Police Service of Northern Ireland, Loquendo SpA, Singarlogic Anonymos Etairia Pliroforiakon Systimaton & Efarmogon Pliroforikis, International Biometric Group (UK) Limited, Ministério da Justiça, Lisboa, The International Criminal Police Organization, Sail Labs Technology AG, Green Fusion Limited (Data Fusion Internation), ok-2go, University of Warwick, Inov Inesc Inocacao, Instituto De Novas Tecnologias, Rijksuniversiteit Groningen



## PROJECTS FUNDED BY SWISS AGENCIES

**Name** **MCM-FF** (Multimodal Computational Modeling of Nonverbal Social Behavior in Face to Face Interaction)

**Funding** SNSF – Ambizione

**Coordinator** Idiap Research Institute

**Duration** 01.11.2011 – 28.02.2015

**Partner(s)**

**Name** **MAAYA** (Multimedia Analysis and Access for Documentation and Decipherment of Maya Epigraphy)

**Funding** SNSF – DACH

**Coordinator** Idiap Research Institute

**Duration** 01.06.2013 – 31.05.2016

**Partner(s)** University of Geneva, University of Bonn

**Name** **SIVI** (Situating Vision to Perceive Object Shape and Affordances)

**Funding** SNSF – DACH

**Coordinator** University of Bonn

**Duration** 01.11.2011 – 30.11.2015

**Partner(s)** Technische Universität Wien, RWTH Aachen, Idiap Research Institute

**Name** **A-MUSE** (Adaptive Multilingual Speech Processing)

**Funding** FP7 – IP – ICT

**Coordinator** Fundacio Barcelona Media

**Duration** 01.12.2013 – 30.11.2016

**Partner(s)** Video Stream Networks, S.L., Université du Maine, Gottfried Wilhelm Leibniz Universität Hannover, Idiap Research Institute, Gesellschaft zur Förderung der angewandten Informationsforschung EV, Deutsche Welle

**Name** **DASH** (Object Detection with Active Sample Harvesting)

**Funding** SNSF – Division II

**Coordinator** Idiap Research Institute

**Duration** 01.09.2012 – 31.08.2015

**Partner(s)**

**Name** **UNITS** (Unified Speech Processing Framework for Trustworthy Speaker Recognition)

**Funding** SNSF – Division II

**Coordinator** Idiap Research Institute

**Duration** 01.07.2015 – 30.06.2018

**Partner(s)**

**Name** **G3E** (G3E: Geometric Generative Gaze Estimation model)

**Funding** SNSF – Division II

**Coordinator** Idiap Research Institute

**Duration** 01.04.2014 – 31.01.2015

**Partner(s)**

**Name** **HFACE** (Heterogeneous Face Recognition)

**Funding** SNSF – Division II

**Coordinator** Idiap Research Institute

**Duration** 01.07.2014 – 30.06.2017

**Partner(s)**

**Name** **PHASER** (PHASER: Parsimonious Hierarchical Automatic Speech Recognition)

**Funding** SNSF – Division II

**Coordinator** Idiap Research Institute

**Duration** 01.06.2014 – 31.05.2016

**Partner(s)**

ACTIVE AND GRANTED PROJECTS IN 2015

**Name** I-DRESS (Assistive Interactive robotic system for support in DRESSing)  
**Funding** SNSF – ERA-NET Chist-Era  
**Coordinator** Agencia Estatal Consejo Superior de Investigaciones Científicas  
**Duration** 01.12.2015 – 30.11.2018  
**Partner(s)** Idiap Research Institute, University of the West of England

**Name** YOUTH@NIGHT (Youth@Night – A multi-disciplinary multi-method study of young people's outgoing and drinking behaviors)  
**Funding** SNSF – Interdisciplinary Project  
**Coordinator** Sucht Schweiz – Research Institute  
**Duration** 01.01.2014 – 31.12.2016  
**Partner(s)** Universität Zürich, Idiap Research Institute

**Name** SP2 (SCOPES Project on Speech Prosody)  
**Funding** SNSF – SCOPES  
**Coordinator** Idiap Research Institute  
**Duration** 01.04.2014 – 31.03.2016  
**Partner(s)** University of Novi Sad, Ss. Cyril and Methodius University, Budapest University of Technology and Economy

**Name** MODERN (Modeling discourse entities and relations for coherent machine translation)  
**Funding** SNSF – Sinergia  
**Coordinator** Fundacio Barcelona Media  
**Duration** 01.12.2013 – 30.11.2016  
**Partner(s)** Universiteit Utrecht, Universität Zürich, University of Geneva

**Name** SIWIS (Spoken Interaction with Interpretation in Switzerland)  
**Funding** SNSF – Sinergia  
**Coordinator** Idiap Research Institute  
**Duration** 01.12.2012 – 30.11.2016  
**Partner(s)** University of Geneva, University of Edinburgh, Eidgenössische Technische Hochschule Zürich

**Name** UBIMPRESSED (UBImpressed: Ubiquitous First Impressions and Ubiquitous Awareness)  
**Funding** SNSF – Sinergia  
**Coordinator** Idiap Research Institute  
**Duration** 01.01.2014 – 31.12.2016  
**Partner(s)** Université de Neuchâtel, Cornell University

**Name** WILDTRACK (Tracking in the Wild)  
**Funding** SNSF – Sinergia  
**Coordinator** Idiap Research Institute  
**Duration** 01.01.2014 – 31.12.2016  
**Partner(s)** Eidgenössische Technische Hochschule Zürich, École Polytechnique Fédérale de Lausanne

**Name** ADDG2SU (Flexible Acoustic Data-Driven Grapheme to Subword Unit Conversion)  
**Funding** Hasler Foundation  
**Coordinator** Idiap Research Institute  
**Duration** 01.03.2013 – 01.03.2016  
**Partner(s)**

**Name** COHFACE (CONTACTless Heartbeat detection for trustworthy FACE Biometrics)  
**Funding** Hasler Foundation  
**Coordinator** Idiap Research Institute  
**Duration** 01.09.2014 – 31.08.2015  
**Partner(s)**

**Name** DEEPSTD-EXT (Universal Spoken Term Detection with Deep Learning (extension))  
**Funding** Hasler Foundation  
**Coordinator** Idiap Research Institute  
**Duration** 01.12.2014 – 30.09.2015  
**Partner(s)**

ACTIVE AND GRANTED PROJECTS IN 2015

**Name** **GENEEMO** (Geneemo: An Expressive Audio Content Generation Tool)  
**Funding** Hasler Foundation  
**Coordinator** Idiap Research Institute  
**Duration** 01.04.2014 – 30.09.2015  
**Partner(s)**

**Name** **L-PASS** (Linguistic-Paralinguistic Speech Synthesis)  
**Funding** Hasler Foundation  
**Coordinator** Idiap Research Institute  
**Duration** 01.12.2015 – 28.02.2016  
**Partner(s)**

**Name** **MASH-2** (Massive Sets of Heuristics for Machine Learning II)  
**Funding** Hasler Foundation  
**Coordinator** Idiap Research Institute  
**Duration** 01.07.2013 – 30.06.2016  
**Partner(s)**

**Name** **RECAPP** (Making speech technology accessible to Swiss people)  
**Funding** Hasler Foundation  
**Coordinator** Idiap Research Institute  
**Duration** 01.01.2015 – 30.09.2016  
**Partner(s)**

**Name** **SESAME** (SEarching Swiss Audio MEmories)  
**Funding** Hasler Foundation  
**Coordinator** Idiap Research Institute  
**Duration** 01.01.2011 – 28.02.2015  
**Partner(s)**

**Name** **SODS** (Semantically Self-Organized Distributed Web Search)  
**Funding** Hasler Foundation  
**Coordinator** Idiap Research Institute  
**Duration** 01.01.2012 – 01.03.2015  
**Partner(s)**

**Name** **OMSI** (Objective Measurement of Speech Intelligibility)  
**Funding** Armasuisse  
**Coordinator** Idiap Research Institute  
**Duration** 01.07.2014 – 30.06.2015  
**Partner(s)**

**Name** **RECOD2014** (low bit-rate speech coding)  
**Funding** Armasuisse  
**Coordinator** Idiap Research Institute  
**Duration** 01.04.2014 – 31.03.2015  
**Partner(s)**

**Name** **SENSECITYVITY** (Mobile Sensing, Urban Awareness, and Collective Action)  
**Funding** EPFL – Coopération  
**Coordinator** Idiap Research Institute  
**Duration** 01.03.2014 – 28.02.2015  
**Partner(s)** Instituto Potosino de Investigacion Cientificay Tecnologica

**Name** **MIRROR** (Mobile Data to Improve Decision Making in African Cities)  
**Funding** EPFL – Coopération  
**Coordinator** Idiap Research Institute  
**Duration** 01.01.2014 – 31.12.2016  
**Partner(s)**

**Name** **VALAIS 2015** (Valais\*Wallis Digital)  
**Funding** Etat du Valais  
**Coordinator** Idiap Research Institute  
**Duration** 01.04.2013 – 31.12.2015  
**Partner(s)** Valais/Wallis Promotion, Migros Engagement, Société Coopérative Migros Valais, Médiathèque Valais et Archives de l'Etat, Ecole professionnelle des arts contemporains

## ACTIVE AND GRANTED PROJECTS IN 2015

**Name** DEMO-NAO (Demonstrateur NAO)  
**Funding** Loterie Romande  
**Coordinator** Idiap Research Institute  
**Duration** 01.09.2014 – 31.08.2015  
**Partner(s)**

**Name** VALAIS+ (Valais+ Une plateforme pour mieux connaître l'espace de vie du canton)  
**Funding** Loterie Romande  
**Coordinator** Idiap Research Institute  
**Duration** 01.09.2015 – 30.08.2016  
**Partner(s)** HES-SO Valais

### PROJECTS FUNDED BY INDUSTRIAL PARTNERS

**Name** BIOWATCH (Biowatch)  
**Funding** Fondation The Ark  
**Coordinator** Idiap Research Institute  
**Duration** 01.01.2015 – 30.04.2015  
**Partner(s)** Biowatch SA

**Name** BAXTER (Démonstrateur Baxter)  
**Funding** Fondation The Ark  
**Coordinator** Idiap Research Institute  
**Duration** 16.02.2015 – 15.04.2015  
**Partner(s)**

**Name** EYEWARE (Eyeware)  
**Funding** Fondation The Ark  
**Coordinator** Idiap Research Institute  
**Duration** 01.09.2015 – 30.04.2016  
**Partner(s)**

**Name** FAR (Face Alignment based on RGB-D Cameras)  
**Funding** Fondation The Ark  
**Coordinator** KeyLemon SA  
**Duration** 01.07.2014 – 28.02.2015  
**Partner(s)** Idiap Research Institute

**Name** HYBRID (Hybrid Recommender System)  
**Funding** Fondation The Ark  
**Coordinator** Unono  
**Duration** 01.08.2015 – 31.03.2016  
**Partner(s)** Idiap Research Institute

**Name** PUNK (Punktuation)  
**Funding** Fondation The Ark  
**Coordinator** recapp  
**Duration** 01.05.2015 – 30.04.2016  
**Partner(s)** Idiap Research Institute

**Name** SUVA (Recomed: Intégration de la transcription vocale dans le dossier patient informatisé CRR)  
**Funding** Fondation The Ark  
**Coordinator** Idiap Research Institute  
**Duration** 01.03.2014 – 31.03.2015  
**Partner(s)** Clinique romande de réadaptation, Cimark

**Name** BEEAPP (Beeapp)  
**Funding** CTI  
**Coordinator** Idiap Research Institute  
**Duration** 01.10.2015 – 31.12.2015  
**Partner(s)** Vergers des Essinges

**Name** DIGIT-ARENA (Real-Time Perimeter Board Content Digital Replacement)  
**Funding** CTI  
**Coordinator** Idiap Research Institute  
**Duration** 01.12.2014 – 30.07.2016  
**Partner(s)** E.S. Concept

**Name** FAVEO (Accelerating online information discovery through context-driven and behaviour-based personalization of search)  
**Funding** CTI  
**Coordinator** Idiap Research Institute  
**Duration** 01.09.2014 – 31.03.2015  
**Partner(s)** Faveeo, École Polytechnique Fédérale de Lausanne



## ACTIVE AND GRANTED PROJECTS IN 2015

**Name** **MULTIVEO** (High Accuracy Speaker-Independent Multilingual Automatic Speech Recognition System)

**Funding** CTI

**Coordinator** Idiap Research Institute

**Duration** 01.11.2013 – 31.05.2015

**Partner(s)** Veovox

**Name** **SCOREL2** (Automatic scoring and adaptive pedagogy for oral language learning)

**Funding** CTI

**Coordinator** Idiap Research Institute

**Duration** 01.06.2014 – 31.01.2016

**Partner(s)** SpeedLingua

**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** 01.12.2015 – 31.07.2017

**Partner(s)** HES-SO Vaud, DomoSafety S.A., Universität Bern

**Name** **VIDEOPROTECTOR** (Morphean VideoProtector)

**Funding** CTI

**Coordinator** Idiap Research Institute

**Duration** 01.06.2014 – 31.12.2015

**Partner(s)** HES-SO Fribourg

**Name** **ATAP** (Continuous Authentication Program)

**Funding** IND – Industrial Project

**Coordinator** Idiap Research Institute

**Duration** 15.01.2015 – 14.04.2015

**Partner(s)** Google

**Name** **DRACULA** (Detect and track people/object in order to deliver personalised movies.)

**Funding** IND – Industrial Project

**Coordinator** Idiap Research Institute

**Duration** 01.07.2015 – 29.02.2016

**Partner(s)** iRewind

**Name** **EMMA1** (Expression Mimics Marker Analysis)

**Funding** IND – Industrial Project

**Coordinator** Idiap Research Institute

**Duration** 01.03.2014 – 28.02.2015

**Partner(s)** DoctorWeb

**Name** **GOOGLE MOBILE** (Mobile Face and Voice Anti Spoofing)

**Funding** IND – Industrial Project

**Coordinator** Idiap Research Institute

**Duration** 15.05.2015 – 31.03.2016

**Partner(s)**

## PROJECTS AWARDED IN 2015 AND STARTING IN 2016

XXXIII

**Name** **MALORCA** (Machine Learning of Speech Recognition Models for Controller Assistance)

**Funding** H2020-SESAR-2015

**Coordinator** Deutsches Zentrum für Luft- und Raumfahrt EV

**Duration** 01.04.2016 – 31.03.2018

**Partner(s)** Idiap Research Institute, Austro Control Österreichische Gesellschaft für Zivilluftfahrt MBH, Řízení letového provozu České republiky, Universität des Saarlandes

**Name** **MUMMER** (MultiModal Mall Entertainment Robot)

**Funding** H2020-ICT-2015

**Coordinator** University of Glasgow

**Duration** 01.03.2016 – 28.02.2020

**Partner(s)** Kiinteistö Oy Ideapark AB, Centre national de la recherche scientifique, Teknologian Tutkimuskeskus VTT, ALDEBARAN Robotics, Idiap Research Institute

## ACTIVE AND GRANTED PROJECTS IN 2015

**Name** **SUMMA** (Scalable Understanding of Multilingual Media)

**Funding** H2020-ICT-2015

**Coordinator** University of Edinburgh

**Duration** 01.02.2016 – 31.01.2019

**Partner(s)** LETA, University College London, Idiap Research Institute, Qatar Computing Research Institute, Deutsche Welle, Priberam Informatica SA, British Broadcast Corporation

**Name** **SWAN** (Secure Access Control over Wide Area Network)

**Funding** Research Council of Norway

**Coordinator** Hogskolen / Gjovik

**Duration** 01.01.2016 – 31.12.2018

**Partner(s)** Bankenverband, Universitaet Oslo, Morpho, Idiap Research Institute, Zwiipe AS

**Name** **TESLA** (An Adaptive Trust-based e-assessment System for Learning)

**Funding** H2020-ICT-2015

**Coordinator** Fundacio Per A La Universitat Oberta de Catalunya

**Duration** 01.01.2016 – 31.12.2018

**Partner(s)** Idiap Research Institute, The Open University, Anadolu University, Instituto Nacional de Astrofisica, Optica y Electronica, Technical University of Sofia, European Quality Assurance Network for Informatics Education e.V., Open Universiteit Nederland, WFSW, SA, Institut Mines-Telecom, Télécom SudParis, European Association for quality assurance in higher education AISBL, LPLUS GmbH, Sofiiski Universitet sveti kliment ohridski, Imperial College London, Protos Sistemas de Información, S.L., Agència per a la Qualitat del Sistema Universitari de Catalunya, Jyvaskylan Yliopisto, Université de Namur ASBL

**Name** **LIVEHEART** (The Cellular Basis of Cardiac Development Revealed by Live Imaging)

**Funding** SNSF – ANR

**Coordinator** Institut de génétique et de biologie moléculaire et cellulaire

**Duration** 01.03.2016 – 28.02.2019

**Partner(s)** Universität Bern, École Polytechnique Paris, Idiap Research Institute

**Name** **ODESSA** (Online Diarization Enhanced by recent Speaker identification and Sequential learning Approaches)

**Funding** SNSF – ANR

**Coordinator** Centre national de la recherche scientifique

**Duration** 01.03.2016 – 28.02.2019

**Partner(s)** EURECOM, Idiap Research Institute

**Name** **PLATFORM MMD** (Platform for Reproducible Acquisition, Processing, and Sharing of Dynamic, Multi-Modal Data)

**Funding** SNSF – R'Equip

**Coordinator** Idiap Research Institute

**Duration** 01.01.2016 – 31.12.2016

**Partner(s)**

**Name** **COMETS-M** (Computational Methods for Temporal Super-resolution Microscopy)

**Funding** SNSF – Division II

**Coordinator** Idiap Research Institute

**Duration** 01.04.2016–30.03.2019

**Partner(s)**

ACTIVE AND GRANTED PROJECTS IN 2015

**Name** **TACT-HAND** (Improving control of prosthetic hands using tactile sensors and realistic machine learning)  
**Funding** SNSF DACH  
**Coordinator** Deutsches Zentrum für Luft- und Raumfahrt EV  
**Duration** 01.04.2016 – 31.03.2019  
**Partner(s)** Universitaet Bielfeld, Idiap Research Institute

**Name** **SMILE** (Scalable Multimodal sign language Technology for sign language Learning and assessmEnt)  
**Funding** SNSF Sinergia  
**Coordinator** Idiap Research Institute  
**Duration** 01.03.2016 – 28.02.2019  
**Partner(s)** University of Surrey, University of Applied Sciences of Special Needs Education

**Name** **ADDG2SU EXT.** (Flexible Acoustic data-driven Grapheme to Subword Unit Conversion)  
**Funding** Hasler Foundation  
**Coordinator** Idiap Research Institute  
**Duration** 01.03.2016 – 28.02.2017  
**Partner(s)**

**Name** **ELEARNING VALAIS 3.0** (eLearning-Valais 3.0)  
**Funding** Fondation The Ark  
**Coordinator** Formation universitaire à Distance  
**Duration** 01.03.2016 – 31.01.2018  
**Partner(s)** KLEWEL, Idiap Research Institute

**Name** **RECOMED2** (Intégration de la transcription vocale dans le dossier patient informatisé CRR)  
**Funding** Fondation The Ark  
**Coordinator** Idiap Research Institute  
**Duration** 01.02.2016 – 30.06.2016  
**Partner(s)** DPC Solutions, Clinique romande de réadaptation, recapp

**Name** **BIOWAVE** (A BIometric Watch Activated by VEins)  
**Funding** CTI  
**Coordinator** Centre Suisse d'Electronique et de Microtechnique  
**Duration** 01.02.2016 – 31.07.2017  
**Partner(s)** Idiap Research Institute, Biowatch SA

**Name** **ESGEM** (Enhanced Swiss German mEdia Monitoring)  
**Funding** CTI  
**Coordinator** Idiap Research Institute  
**Duration** 01.03.2016 – 30.10.2017  
**Partner(s)** recapp, ARGUS der Presse AG

**Name** **FARGO** (Convenient and Secure 3D Face Recognition based on RGB-D Cameras)  
**Funding** CTI  
**Coordinator** Idiap Research Institute  
**Duration** 01.05.2016 – 31.10.2017  
**Partner(s)** KeyLemon SA

## Books and Book Chapters

- [1] O. Aran, I. Ari, A. Kindiroglu, P. Santemiz, and L. Akarun, "Otomatik işaret dili tanıma ve türk işaret dili için bilgisayar uygulamaları," in *Türk İşaret Dili Araştırmaları - Research on Turkish Sign Language*, in Turkish, Koc University Press, 2015.
- [2] D. Bruno, S. Calinon, M. S. Malekzadeh, and D. G. Caldwell, "Learning the stiffness of a continuous soft manipulator from multiple demonstrations," in *Intelligent Robotics and Applications*, ser. Lecture Notes in Computer Science, H. Liu, N. Kubota, X. Zhu, R. Dillmann, and D. Zhou, Eds., vol. 9246, Best Paper Award Finalist at ICIRA'2015, Springer, 2015, pp. 185–195.
- [3] T. Kriyakerne and D. Ginsbourger, "Global optimization with sparse and local gaussian process models," in *Machine Learning, Optimization, and Big Data*, ser. Lecture Notes in Computer Science, P. Pardalos, M. Pavone, G. M. Farinella, and V. Cutello, Eds., vol. 9432, Springer International Publishing, 2015, pp. 185–196.
- [4] A. Lazaridis, B. Potard, and P. N. Garner, "DNN-based speech synthesis: Importance of input features and training data," in *International Conference on Speech and Computer*, SPECOM, ser. Lecture Notes in Computer Science, A. Ronzhin, R. Potapova, and N. Fakotakis, Eds., vol. 9319, Springer Berlin Heidelberg, 2015, pp. 193–200.
- [5] R. Lebret and R. Collobert, "Rehabilitation of count-based models for word vector representations," in *Computational Linguistics and Intelligent Text Processing*, ser. Lecture Notes in Computer Science, Alexander Gelbukh, vol. 9041, Springer International Publishing, 2015, pp. 417–429.
- [6] S. Marmin, C. Chevalier, and D. Ginsbourger, "Differentiating the multipoint expected improvement for optimal batch design," in *Machine Learning, Optimization, and Big Data*, ser. Lecture Notes in Computer Science, P. Pardalos, M. Pavone, G. M. Farinella, and V. Cutello, Eds., vol. 9432, Springer International Publishing, 2015, pp. 37–48.

- [7] B. Webber, M. Carpuat, A. Popescu-Belis, and C. Hardmeier, *Proceedings of the Second Workshop on Discourse in Machine Translation (DiscoMT 2015)*. Association for Computational Linguistics, Sep. 2015.

## Articles in Scientific Journals

- [1] M. Cernak, P. N. Garner, A. Lazaridis, P. Motlicek, and X. Na, "Incremental syllable-context phonetic vocoding," *IEEE/ACM Transactions on audio, speech, and language processing*, vol. 23, no. 6, Jun. 2015.
- [2] I. Chingovska and A. Anjos, "On the use of client identity information for face anti-spoofing," *IEEE Transactions on Information Forensics and Security*, Special Issue on Biometric Anti-spoofing, vol. 10, no. 4, pp. 787–796, 2015.
- [3] A. Dutta, M. Günther, L. El Shafey, S. Marcel, R. Veldhuis, and L. Spreeuwers, "Impact of eye detection error on face recognition performance," *IET Biometrics*, vol. 4, no. 3, pp. 137–150, Sep. 2015.
- [4] K. A. Funes Mora and J.-M. Odobez, "Gaze estimation in the 3D space using RGB-D sensors. towards head-pose and user invariance.," *International Journal of Computer Vision*, pp. 1–23, Nov. 2015.
- [5] D. Gatica-Perez, "Signal processing in the workplace," *IEEE Signal Processing Magazine*, vol. 32, no. 1, pp. 121–125, Jan. 2015. *Vision*, pp. 1–23, Nov. 2015.
- [6] M. Guillemot, J.-M. Odobez, A. Vinciarelli, and S. Ingram, "KleweL webcast: From research to growing company," *IEEE Multimedia*, vol. 22, no. 4, pp. 94–99, Dec. 2015.
- [7] M. Günther, S. Böhringer, D. Wiczorek, and R. P. Würtz, "Reconstruction of images from Gabor graphs with applications in facial image processing," *Journal of Wavelets, Multiresolution and Information Processing*, vol. 13, no. 4, pp. 1550019 1–25, 2015.



## LIST OF PUBLICATIONS IN 2015

- [8] M. Habibi and A. Popescu-Belis, "Keyword extraction and clustering for document recommendation in conversations," *IEEE/ACM Transactions on Audio Speech and Language Processing*, vol. 23, no. 4, pp. 746–759, 2015.
- [9] A. Hadid, N. Evans, S. Marcel, and J. Fierrez, "Biometrics systems under spoofing attack: An evaluation methodology and lessons learned," *IEEE Signal Processing Magazine*, vol. 32, no. 5, pp. 20–30, 2015.
- [10] R. Hu, G. Can, C. Pallan Gayol, G. Krempel, J. Spotak, G. Vail, S. Marchand-Maillet, J.-M. Odobez, and D. Gatica-Perez, "Multimedia analysis and access of ancient Maya epigraphy," *Signal processing magazine*, vol. 32, no. 4, pp. 75–84, 2015.
- [11] Y. E. Kara, G. Genc, O. Aran, and L. Akarun, "Modeling annotator behaviors for crowd labeling," *Neurocomputing*, vol. 160, no. C, pp. 141–156, Jul. 2015.
- [12] 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*, pp. 1–14, 2015.
- [13] T. Meyer, N. Hajlaoui, and A. Popescu-Belis, "Disambiguating discourse connectives for statistical machine translation," *IEEE/ACM Transactions on Audio, Speech and Language Processing*, vol. 23, no. 7, pp. 1184–1197, Jul. 2015.
- [14] P. Motlicek, D. Imseng, B. Potard, P. N. Garner, and I. Himawan, "Exploiting foreign resources for DNN-based ASR," *EURASIP Journal on Audio, Speech, and Music Processing*, no. 2015:17, pp. 1–10, Jun. 2015.
- [15] N. Pappas and A. Popescu-Belis, "Combining content with user preferences for non-fiction multimedia recommendation: A study on TED lectures," *Multimedia Tools and Applications, Special Issue on Content Based Multimedia Indexing*, pp. 1–23, Jan. 2015.
- [16] R. Rasipuram and M. Magimai-Doss, "Acoustic and lexical resource constrained ASR using language-independent acoustic model and language-dependent probabilistic lexical model," *Speech Communication*, vol. 68, pp. 23–40, Apr. 2015.
- [17] D. Sanchez-Cortes, S. Kumano, K. Otsuka, and D. Gatica-Perez, "In the mood for Vlog: Multimodal inference in conversational social video," *ACM Transactions on Interactive Intelligent Systems*, vol. 5, no. 2, 9:1–9:24, 2015.
- [18] A. Sapru and H. Bourlard, "Automatic recognition of emergent social roles in small group interactions," *IEEE Transactions on Multimedia*, vol. 17, no. 5, pp. 746–760, 2015.
- [19] B. Schuller, S. Steidl, A. Batliner, E. Nöth, A. Vinciarelli, F. Burkhardt, f. Weninger, F. Eyben, T. Bocklet, G. Mohammadi, and B. Weiss, "A survey on perceived speaker traits: Personality, likability, pathology and the first challenge," *Computer Speech and Language*, vol. 19, no. 1, pp. 100–131, Jan. 2015.
- [20] S. Sheikhi and J.-M. Odobez, "Combining dynamic head pose-gaze mapping with the robot conversational state for attention recognition in human-robot interactions," *Pattern Recognition Letters*, vol. 66, pp. 81–90, Nov. 2015.
- [21] A. Sizov, E. Khoury, T. Kinnunen, Z. Wu, and S. Marcel, "Joint speaker verification and antispoofing in the i-vector space," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 4, pp. 821–832, Apr. 2015.
- [22] N. Suditu and F. Fleuret, "Adaptive relevance feedback for large-scale image retrieval," *Multimedia wTools and Applications*, pp. 1–31, 2015.
- [23] M. J. Taghizadeh, A. Asaei, S. Haghighatshoar, P. N. Garner, and H. Bourlard, "Spatial sound localization via multipath euclidean distance matrix recovery," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 5, pp. 802–814, Aug. 2015.
- [24] V. Trivedi, T. V. Truong, L. A. Trinh, D. B. Holland, M. Liebling, and S. E. Fraser, "Dynamic structure and protein expression of the live embryonic heart captured by 2-photon light sheet microscopy and retrospective registration," *Biomedical Optics Express*, vol. 6, no. 6, pp. 2056–2066, Jun. 2015.

- [25]** A. Asaei, H. Bourlard, M. J. Taghizadeh, and V. Cevher, "Computational methods for underdetermined convolutive speech localization and separation via model-based sparse component analysis," *Speech Communication*, in press.
- [26]** A. Asaei, M. J. Taghizadeh, S. Haghghatshoar, B. Raj, H. Bourlard, and V. Cevher, "Binary sparse coding of convolutive mixtures for sound localization and separation via spatialization," *IEEE Transactions on Signal Processing*, in press.
- [27]** P. Dighe, A. Asaei, and H. Bourlard, "Sparse modeling of neural network posterior probabilities for exemplar-based speech recognition," *Speech Communication: Special Issue on Advances in Sparse Modeling and Low-rank Modeling for Speech Processing*, in press.
- [28]** N. Pappas and A. Popescu-Belis, "Adaptive sentiment-aware one-class collaborative filtering," *Expert Systems with Applications*, in press.
- [29]** O. Pertz, L. Fusco, R. Lefort, K. C. Smith, F. Benmansour, G. Gonzalez, C. Barilari, B. Rinn, F. Fleuret, and P. Fua, "Computer vision profiling of neurite outgrowth dynamics reveals spatiotemporal modularity of Rho GTPase signaling," *Journal of Cell Biology*, in press.
- [30]** R. Rasipuram and M. Magimai-Doss, "Articulatory feature based continuous speech recognition using probabilistic lexical modeling," *Computer Speech and Language*, in press.
- [31]** R. Ullmann and H. Bourlard, "Predicting the intrusiveness of noise through sparse coding with auditory kernels," *Speech Communication*, in press.

## Theses

- [1]** I. Chingovska, "Trustworthy biometric verification under spoofing attacks: Application to the face mode," PhD thesis, École Polytechnique Fédérale de Lausanne, Nov. 2015.
- [2]** K. A. Funes Mora, "3D gaze estimation from remote RGB-D sensors," PhD thesis, École Polytechnique Fédérale de Lausanne, Oct. 2015.
- [3]** M. Habibi, "Modeling users' information needs in a document recommender for meetings," PhD thesis, École Polytechnique Fédérale de Lausanne, Nov. 2015.
- [4]** L. S. Nguyen, "Computational analysis of behavior in employment interviews and video resumes," PhD thesis, École Polytechnique Fédérale de Lausanne, May 2015.
- [5]** A. Sapru, "Automatic social role recognition and its application in structuring multiparty interactions," PhD thesis, École Polytechnique Fédérale de Lausanne, Jun. 2015.
- [6]** M. J. Taghizadeh, "Enabling speech applications using ad-hoc microphone arrays," PhD thesis, École Polytechnique Fédérale de Lausanne, May 2015.
- [7]** S. H. Yella, "Speaker diarization of spontaneous meeting room conversations," PhD thesis, École Polytechnique Fédérale de Lausanne, Lausanne, Jan. 2015.

# Articles in Conference Proceedings

- [1] A. Asaei, M. Cernak, and H. Boulard, "On compressibility of neural network phonological features for low bit rate speech coding," in *Proceeding of Interspeech*, 2015.
- [2] A. Asaei, N. Mohammadiha, M. J. Taghizadeh, S. Doclo, and H. Boulard, "On application of non-negative matrix factorization for ad hoc microphone array calibration from incomplete noisy distances," in *IEEE 40th International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2015.
- [3] T. Bagautdinov, F. Fleuret, and P. Fua, "Probability occupancy maps for occluded depth images," in *Proceedings of the IEEE international conference on Computer Vision and Pattern Recognition*, 2015.
- [4] S. Calinon, "Robot learning with task-parameterized generative models," in *Proc. Intl Symp. on Robotics Research*, 2015.
- [5] M. Cernak and P.-E. Honnet, "An empirical model of emphatic word detection," in *Proc. of Interspeech*, Dresden, Germany: ISCA, Sep. 2015.
- [6] M. Cernak, B. Potard, and P. N. Garner, "Phonological vocoding using artificial neural networks," in *IEEE 40th International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brisbane, Australia: IEEE, Apr. 2015.
- [7] N. Chacko, K. G. Chan, and M. Liebling, "Intensity-based point-spread-function-aware registration for multi-view applications in optical microscopy," in *Biomedical Imaging (ISBI)*, 2015 IEEE 12th International Symposium on, Apr. 2015, pp. 306–309.
- [8] K. G. Chan and M. Liebling, "Estimation of divergence-free 3D cardiac blood flow in a zebrafish larva using multi-view microscopy," in *Biomedical Imaging (ISBI)*, 2015 IEEE 12th International Symposium on, IEEE, Brooklyn, NY, USA, Apr. 2015, pp. 385–388.
- [9] G. Chávez-Martínez, S. Ruiz-Correa, and D. Gatica-Perez, "International conference on mobile and ubiquitous multimedia," in *Happy and Agreeable? Multi-Label Classification of Impressions in Social Video*, Linz, Dec. 2015.
- [10] Y. Chen, Y. Yu, and J.-M. Odobez, "Head nod detection from a full 3D model," in *Proceedings of the ICCV 2015*, Dec. 2015.
- [11] P. Dighe, A. Asaei, and H. Boulard, "Dictionary learning for sparse representation of neural network exemplars in speech recognition," in *Proceedings of SPARS 2015: Workshop on Signal Processing with Adaptive Sparse Structured Representations*, 2015, 2015.
- [12] P. Dighe, A. Asaei, and H. Boulard, "Sparse modeling of neural network posterior probabilities for exemplar-based speech recognition," in *Proceedings of SPARS 2015: Workshop on Signal Processing with Adaptive Sparse Structured Representations*, 2015.
- [13] N. Erdogmus, M. Vanoni, and S. Marcel, "Within- and cross- database evaluations for gender classification via BeFIT protocols," in *International Workshop on Multimedia Signal Processing*, 2015.
- [14] T. de Freitas Pereira and S. Marcel, "Periocular biometrics in mobile environment," in *IEEE Seventh International Conference on Biometrics: Theory, Applications and Systems*, Arlington, USA: IEEE, Sep. 2015.
- [15] J. Gancet, D. Urbina, P. Letier, M. Ilzokvitz, P. Weiss, F. Gauch, G. Antonelli, G. Indiveri, G. Casalino, A. Birk, M. F. Pfungsthorn, S. Calinon, A. Tanwani, A. Turetta, C. Walen, and L. Guilpain, "Dexrov: Dexterous undersea inspection and maintenance in presence of communication latencies," in *IFAC Workshop on Navigation, Guidance and Control of Underwater Vehicles (NGCUV)*, 2015.
- [16] B. Gerazov and P. N. Garner, "An investigation of muscle models for physiologically based intonation modeling," in *Proceedings of the 23rd Telecommunications Forum*, Belgrade, Serbia, Nov. 2015.
- [17] B. Gerazov, P.-E. Honnet, A. Gjoreski, and P. N. Garner, "Weighted correlation based atom decomposition

## LIST OF PUBLICATIONS IN 2015

intonation modelling," in Proceedings of Interspeech, Sep. 2015.

- [18]** G. Gosztolya, T. Grósz, L. Tóth, and D. Imseng, "Building context-dependent DNN acoustic models using Kullback-Leibler divergence-based state tying," in Proceedings IEEE International Conference on Acoustics, Speech and Signal Processing, 2015.
- [19]** M. Habibi and A. Popescu-Belis, "Query refinement using conversational context: A method and an evaluation resource," in Proceedings of the NLDB 2015 (20th International Conference on Applications of Natural Language to Information Systems), 2015.
- [20]** I. Himawan, P. Motlicek, M. Ferras, and S. Madikeri, "Towards utterance-based neural network adaptation in acoustic modeling," in IEEE Automatic Speech Recognition and Understanding Workshop, 2015.
- [21]** I. Himawan, P. Motlicek, D. Imseng, B. Potard, N. Kim, and J. Lee, "Learning feature mapping using deep neural network bottleneck features for distant large vocabulary speech recognition," in IEEE International Conference on Acoustics, Speech, and Signal Processing, 2015.
- [22]** I. Himawan, P. Motlicek, S. Sridharan, D. Dean, and D. Tjondronegoro, "Channel selection in the short-time modulation domain for distant speech recognition," in Proceedings of Interspeech, 2015.
- [23]** P.-E. Honnet, B. Gerazov, and P. N. Garner, "Atom decomposition-based intonation modelling," in IEEE 40th International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brisbane, Australia: IEEE, Apr. 2015.
- [24]** A. Hyafil and M. Cernak, "Neuromorphic based oscillatory device for incremental syllable boundary detection," in Proc. of Interspeech, Dresden, Germany, Sep. 2015.
- [25]** B. Jou, T. Chen, N. Pappas, M. Redi, M. Topkara, and S.-F. Chang, "Visual affect around the world: A large-scale multilingual visual sentiment ontology," in ACM International Conference on Multimedia, 2015.
- [26]** E. Khan, P. Baqué, F. Fleuret, and P. Fua, "Kullback-Leibler proximal variational inference," in Proceedings

of the international conference on Neural Information Processing Systems, 2015.

- [27]** S. Kucur Ergunay, E. Khoury, A. Lazaridis, and S. Marcel, "On the vulnerability of speaker verification to realistic voice spoofing," in IEEE International Conference on Biometrics: Theory, Applications and Systems, Sep. 2015.
- [28]** I. Kuzborskij, F. Orabona, and B. Caputo, "Transfer learning through greedy subset selection," in Volume 9279 of the series Lecture Notes in Computer Science, 2015.
- [29]** N. Le, D. Wu, S. Meignier, and J.-M. Odobez, "EUMSSI team at the mediaeval person discovery challenge," in Working Notes Proceedings of the MediaEval 2015 Workshop, Wurzen, Germany, Sep. 2015.
- [30]** R. Lebrete and R. Collobert, "N-gram-based low-dimensional representation for document classification," in International Conference on Learning Representations, 2015.
- [31]** R. Lebrete, P. H. O. Pinheiro, and R. Collobert, "Phrase-based image captioning," in International Conference on Machine Learning (ICML), 2015.
- [32]** J. Legrand and R. Collobert, "Joint RNN-based greedy parsing and word composition," in Proceedings of ICLR 2015, 2015.
- [33]** N.-Q. Luong, L. M. Werlen, and A. Popescu-Belis, "Pronoun translation and prediction with or without coreference links," in EMNLP 2015 Workshop on Discourse in Machine Translation, Sep. 2015.
- [34]** S. Madikeri and H. Bourlard, "KL-HMM based speaker diarization system for meetings," in Proceedings of ICASSP 2015, Apr. 2015, pp. 4435-4439.
- [35]** S. Madikeri, I. Himawan, P. Motlicek, and M. Ferras, "Integrating online l-vector extractor with information bottleneck based speaker diarization system," in Proceedings of Interspeech 2015, 2015.
- [36]** S. Madikeri, P. Motlicek, and H. Bourlard, "Combining SGMM speaker vectors and KL-HMM approach for



## LIST OF PUBLICATIONS IN 2015

- speaker diarization," in Proceedings of ICASSP 2015, 2015, pp. 4834-4837.
- [37]** R. Metha, M. Günther, and S. Marcel, "Gender classification by LUT based boosting of overlapping block patterns," in Scandinavian Conference on Image Analysis, vol. 9127, Springer International Publishing, 2015, pp. 530-542.
- [38]** P. Motlicek, S. Dey, S. Madikeri, and L. Burget, "Employment of subspace gaussian mixture models in speaker recognition," in 2015 IEEE International Conference on Acoustics, Speech, and Signal Processing, IEEE, Brisbane, Australia, May 2015, pp. 4445-4449.
- [39]** L. S. Nguyen and D. Gatica-Perez, "I would hire you in a minute: Thin slices of nonverbal behavior in job interviews," in Proceedings of the ACM International Conference on Multimodal Interaction, 2015.
- [40]** C. Oertel, K. A. Funes Mora, J. Gustafson, and J.-M. Odobez, "Deciphering the silent participant. On the use of audio-visual cues for the classification of listener categories in group discussions," in International Conference on Multimodal Interaction, Seattle, Washington, USA: ACM, Nov. 2015.
- [41]** S. Okada, O. Aran, and D. Gatica-Perez, "Personality trait classification via co-occurrent multiparty multimodal event discovery," in Proceedings of the ACM International Conference on Multimodal Interaction, 2015.
- [42]** D. Palaz, M. Magimai.-Doss, and R. Collobert, "Analysis of CNN-based speech recognition system using raw speech as input," in Proceedings of Interspeech, 2015.
- [43]** D. Palaz, M. Magimai.-Doss, and R. Collobert, "Convolutional neural networks-based continuous speech recognition using raw speech signal," in International Conference on Acoustics, Speech and Signal Processing, Apr. 2015.
- [44]** P. H. O. Pinheiro and R. Collobert, "From image-level to pixel-level labeling with convolutional networks," in Computer Vision and Pattern Recognition (CVPR), 2015.
- [45]** P. H. O. Pinheiro, R. Collobert, and P. Dollár, "Learning to segment objects candidates," in Advances in Neural Information Processing Systems, 2015.
- [46]** X. Pu, L. Mascarell, A. Popescu-Belis, M. Fishel, N.-Q. Luong, and M. Volk, "Leveraging compounds to improve noun phrase translation from chinese and german," in Proceedings of the ACL-IJCNLP 2015 Student Research Workshop, Jul. 2015.
- [47]** R. Raghavendra, M. Avinas, C. Busch, and S. Marcel, "Finger vein liveness detection using motion magnification," in IEEE International Conference on Biometrics: Theory, Applications and Systems, Sep. 2015.
- [48]** D. Ram, A. Asaei, P. Dighe, and H. Bourlard, "Sparse modeling of posterior exemplars for keyword detection," in Proceedings of Interspeech, 2015.
- [49]** R. Rasipuram, M. Cernak, A. Nanchen, and M. Magimai.-Doss, "Automatic accentedness evaluation of non-native speech using phonetic and sub-phonetic posterior probabilities," in Proceedings of Interspeech, Sep. 2015.
- [50]** R. Rasipuram, M. Razavi, and M. Magimai.-Doss, "Integrated pronunciation learning for automatic speech recognition using probabilistic lexical modeling," in International Conference on Acoustics, Speech and Signal Processing, 2015.
- [51]** M. Razavi and M. Magimai.-Doss, "An HMM-based formalism for automatic subword unit derivation and pronunciation generation," in International Conference on Acoustics, Speech and Signal Processing, 2015.
- [52]** M. Razavi, R. Rasipuram, and M. Magimai.-Doss, "Pronunciation lexicon development for under-resourced languages using automatically derived subword units: A case study on scottish gaelic," in 4th Biennial Workshop on Less-Resourced Languages, 2015.
- [53]** L. Rozo, D. Bruno, S. Calinon, and D. G. Caldwell, "Learning optimal controllers in human-robot

LIST OF PUBLICATIONS IN 2015

cooperative transportation tasks with position and force constraints," in Proc. IEEE/RSJ Intl Conf. on Intelligent Robots and Systems (IROS), Hamburg, Germany, 2015.

**[54]** D. Santani and D. Gatica-Perez, "Loud and trendy: Crowdsourcing impressions of social ambiance in popular indoor urban places," in Proceedings of the 23rd ACM International Conference on Multimedia, ser. MM '15, Brisbane, Australia: ACM, Oct. 2015, pp. 211–220.

**[55]** D. Santani, J. Njuguna, T. Bills, A. W. Bryant, R. Bryant, J. Ledgard, and D. Gatica-Perez, "CommuniSense: Crowdsourcing road hazards in Nairobi," in Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services, ser. MobileHCI '15, Copenhagen, Denmark: ACM, Aug. 2015, pp. 445–456.

**[56]** D. Santani, S. Ruiz-Correa, and D. Gatica-Perez, "Looking at cities in Mexico with crowds," in Proceedings of the 2015 Annual Symposium on Computing for Development, ser. DEV '15, London, United Kingdom: ACM, Dec. 2015, pp. 127–135.

**[57]** B. Schuller, B. Vlasenko, F. Eyben, M. Wöllmer, A. Stuhlsatz, A. Wendemuth, and G. Rigoll, "Cross-corpus acoustic emotion recognition: Variances and strategies (extended abstract)," in Proceedings of ACII, 2015.

**[58]** I. Siegert, R. Boeck, B. Vlasenko, and A. Wendemuth, "Exploring dataset similarities using PCAbased feature selection," in Proceedings of ACII, 2015.

**[59]** I. Siegert, R. Boeck, B. Vlasenko, K. Ohnemus, and A. Wendemuth, "Overlapping speech, utterance duration and affective content in HHI and HCI – an comparison," in 6th IEEE Conference on Cognitive Infocommunications, 2015.

**[60]** J. Silverio, L. Rozo, S. Calinon, and D. G. Caldwell, "Learning bimanual end-effector poses from demonstrations using task-parameterized dynamical systems," in Proc. IEEE/RSJ Intl Conf. on Intelligent Robots and Systems (IROS), Hamburg, Germany, 2015.

**[61]** M. J. Taghizadeh, S. Haghghatshoar, A. Asaei, P. N.

Garner, and H. Boulard, "Robust microphone placement for source localization from noisy distance measurements," in IEEE 40th International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2015.

**[62]** P. Tome and S. Marcel, "On the vulnerability of palm vein recognition to spoofing attacks," in The 8th IAPR International Conference on Biometrics (ICB), May 2015.

**[63]** P. Tome and S. Marcel, "Palm vein database and experimental framework for reproducible research," in IEEE International Conference of the Biometrics Special Interest Group, Sep. 2015.

**[64]** P. Tome, R. Raghavendra, C. Busch, S. Tirunagari, N. Poh, B. H. Shekar, D. Gragnaniello, C. Sansone, L. Verdoliva, and S. Marcel, "The 1st competition on counter measures to finger vein spoofing attacks," in The 8th IAPR International Conference on Biometrics (ICB), May 2015.

**[65]** T. Tommasi, N. Patricia, B. Caputo, and T. Tuytelaars, "A deeper look at dataset bias," in German Conference on Pattern Recognition, 2015.

**[66]** R. Ullmann, M. Magimai-Doss, and H. Boulard, "Objective speech intelligibility assessment through comparison of phoneme class conditional probability sequences," in 40th IEEE Int. Conf. on Acoustics, Speech and Signal Processing, Brisbane, Australia, 2015.

**[67]** R. Ullmann, R. Rasipuram, M. Magimai-Doss, and H. Boulard, "Objective intelligibility assessment of text-to-speech systems through utterance verification," in Proceedings of Interspeech, 2015.

**[68]** J. Velasco, M. J. Taghizadeh, A. Asaei, H. Boulard, C. J. Martín-Arguedas, J. Macias-Guarasa, and D. Pizarro, "Novel GCC-PHAT model in diffuse sound field for microphone array pairwise distance based calibration," in IEEE 40th International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2015.

**[69]** B. Vlasenko and A. Wendemuth, "Annotators'

## LIST OF PUBLICATIONS IN 2015

agreement and spontaneous emotion classification performance," in Proceedings of Interspeech, 2015.

- [70]** M. Zeestraten, S. Calinon, and D. G. Caldwell, "Learned minimal intervention control synthesis based on hidden semi-Markov models," in Proc. of the 8th Intl Workshop on Human-Friendly Robotics, Munich, Germany, Oct. 2015, p. 17.

I — XLIV

**encarté scientifique**