



## ANNUAL REPORT 2013



## Publication details

Production: Martina Fellay and Léonore Miauton, Project Management Team

Drafting: Le fin mot Communication, Martigny

Translation and correction: Dave Brooks, ELCS

Graphic design: Atelier Grand, Sierre

Photographic credits: Céline Ribordy, Sion; Sedrik Nemeth, Sion; Idiap, Martigny

Printing: CRI, Centre rhodanien d'impression SA, Martigny

Print run: 1,800 copies

# CONTENTS

## Messages

|   |   |
|---|---|
| Audacity, hard work, and excellence<br>Olivier Dumas, President of the Foundation Council, Idiap              | 2 |
| From a national research center to a continuing quest for excellence<br>Prof. Hervé Bourlard, Director, Idiap | 3 |

## Research

|  |    |
|--|----|
| Profile: An institute on a human scale, but with an international reputation           | 5  |
| National Centre of Competence in Research – IM2  | 10 |
| Valais*Wallis Digital: The collective memory of Valais – by the people, for the people | 13 |
| Project SIIP: The voice – the signature at the very heart of security                  | 14 |
| International appraisals: Idiap's work recognized on the European stage                | 15 |

## Network

|  |    |
|--|----|
| Idiap - EPFL: A strategic alliance. Research and training – a fruitful partnership | 17 |
| International Create Challenge: The expressway to enterprise creation              | 18 |
| Start-up news: The three Ks have the wind in their sails                           | 19 |
| Computer infrastructure: At the service of global data-sharing                     | 21 |

## Faces

|   |    |
|---|----|
| Philip Garner, senior researcher: "Life isn't a serious business"   | 23 |
| Vasil Khalidov, senior development engineer:<br>"The quality of the infrastructure at Idiap impresses me" | 24 |
| David Imseng, postdoc:<br>"For a Valais-lover like me, working at Idiap is amazing"                       | 25 |
| Michel Salamin, French teacher:<br>"I travel thanks to my students!"                                      | 26 |
| Joining and leaving   | 27 |
| Distinctions  | 28 |
| Theses completed  | 29 |

## Finances

|   |    |
|---|----|
| Operating accounts                      | 31 |
| Sources of financing / Costs / Comments | 32 |
| Balance sheet                           | 33 |

## Organization

|                         |       |
|-------------------------|-------|
| Organization chart      | 35    |
| Employees               | 36-37 |
| Foundation Council 2013 | 38    |
| Advisory Board          | 39    |
| Main partners           | 40    |

## Scientific Inserts

|  |       |
|--|-------|
| Idiap Research Areas: Human and Media Computing          | II    |
| Scientific Progress Report                               | IV    |
| Selection of Idiap's key scientific achievements in 2013 |       |
| Main projects in progress                                | XVIII |
| Major publications / Conferences                         | XXVI  |

# A MESSAGE FROM THE PRESIDENT

---

## AUDACITY, HARD WORK, AND EXCELLENCE



**Olivier Dumas, President of the Foundation Council, Idiap**

Moving in the inner-circle of a research institute like Idiap can be a dizzying experience. Not because of the complexity of the research carried out here, nor even due to that magnificent stairway which soars up toward the broad, central light well in the heart of the building that houses, on four floors, the institute from Martigny...

No, the feeling of dizziness comes from elsewhere. From the tremendous energy that, for twenty years now, has kept Idiap moving forward, multiplying its budget by thirty, its head count by five, expanding the influence of its network far beyond national and continental frontiers, and positioning it as a major player in its chosen fields of expertise; fields situated at the heart of global issues linked to information technology. Today, this balance enables the institute to provide for around sixty percent of its budget thanks to the competitive research projects its researchers have secured at international and national levels. Balance, therefore, but also excellence.

### **Ten million from Bern for 2013-2016**

In February 2013, that excellence was also rewarded by the Confederation's decision to extend its confidence in and support of Idiap to the years 2013-2016, to the tune of CHF 10M, an increase of CHF 3.5M on the previous four-year period. And to think that, when it was first granted, that subsidy amounted to just CHF 150,000 — what a long way we have come. Today, thanks to this subvention from the State Secretariat for Training, Research, and Innovation (SERI), not to mention the substantial support of the Canton of Valais and the city of Martigny, Idiap is reinforcing its foundations and continuing to rise, but also to perpetuate its research activities and especially to further develop its technology-transfer activities, one of its three core missions.

This gesture from the Confederation also emphasizes the importance of the strategic alliance that has linked, since 2008, Idiap and EPFL, with the aim of developing scientific and academic activities of common interest, but also of making possible further synergies with the EPFL pole Valais/Wallis, particularly in the fields of energy, health, and security and risk management, fields in which Idiap has been active for several years.

### **Projects for Valais too**

Because even if Idiap's ambitions have placed it on the world stage, the institute also offers its expertise at the regional level. Partner of the Valais-paperless-parliament project and creator of the parliamentary-session indexing engine Mediaparl, in the spring of 2013 the institute convinced the jury of the bicentenary of Valais' entry into the Confederation with its "Valais\*Wallis Digital" project for digitizing the collective memory of our canton — a project that will be carried out in collaboration with the canton's multimedia library and Migros Valais (See page 13).

A real motor for innovation, Idiap is also involved in business incubation on the IdeArk site, where no fewer than twelve start-ups are currently active.

Certain of these start-ups have experienced a meteoric rise, as is the case of Koemei, which — in 2013 — signed a contract with Al Jazeera and was named by the World Economic Forum as a "Technology Pioneer", a title previously bestowed on small companies that were to become giants, such as Google (See page 20). In 2013, three Idiap start-ups could boast of their inclusion in the list of "50 start-ups to invest in", compiled by the business magazine *Bilan*.

Yes, Idiap makes you dizzy. Such audacity, desire, hard work, and excellence, all in one "small institution" — as certain media outlets describe us — can unsettle you. But sometimes losing yourself helps you to better see the stars.

A handwritten signature in blue ink, appearing to be the name of the author, written in a cursive style.

# A MESSAGE FROM THE DIRECTOR

## FROM A NATIONAL RESEARCH CENTER TO A CONTINUING QUEST FOR EXCELLENCE



**Prof. Hervé Bourlard, Director, Idiap**

The year 2013 marked, for Idiap, the conclusion of a formidable twelve-year adventure — the management of the National Centre of Competence in Research IM2 (Interactive Multimodal Information Management). When such a program comes to a close, one can be tempted to retreat into regret and nostalgia. But one can equally draw from such an experience a greater strength upon which the future can feed. And this will undoubtedly be the case for Idiap. For in those twelve years, Idiap has indeed become stronger, growing its national and international reputation and the scientific and technological potential that is now coveted by many of our partners.

### **IM2 — the cradle of Idiap**

In 2001, when the Swiss National Science Foundation (SNSF) — in an effort to maintain Switzerland's position as one of the world's most productive nations — chose, from over 300 project proposals, Idiap for the creation of a National Centre of Competence in Research (NCCR), our institute was barely a decade old. For the SNSF then, this was in some ways an investment in youth.

Idiap seized that opportunity with enthusiasm, aware of how very favorable that particular constellation of conditions was to its taking the next step in its quest for excellence. We succeeded in that step, and - year after year - the committee of experts appointed to analyze our work renewed our status as an NCCR. More than simply a source of additional funding for our projects, we saw in these renewals tremendous recognition and an inexhaustible source of encouragement. For, more than funding, this is all about confidence and motivation — creating a unique atmosphere of creativity and an environment in which we feel that anything is possible.

Yes, IM2 was the cradle of Idiap — a perfect launching pad. Without it, would the institute have reached the level at which it finds itself today? Probably, given the particularly promising and transversal nature of our research and the quality of our researchers. But such an evolution would certainly have taken place more slowly.

### **Ten percent of funding linked to technology transfer**

The closure of the IM2 chapter leaves us with a lasting and fertile heritage that will allow us, tomorrow, to go on making history. For when quality research is being conducted there really is no beginning or end, only a constant metamorphosis and controlled diversification. In a scientific world, which sometimes yields to the temptations of superficiality and fashionable themes, Idiap strives to maintain its course in its chosen fields of research. This also has the effect of maximizing potential impact in terms of technology transfer that can only succeed in a stable and visionary environment in the long term, steering clear of grandstanding and "opportunities" that may well be ephemeral. In this regard, the proportion of our budget that comes from industry-related projects, around 10%, is remarkable in comparison with institutions often cited as exemplary, such as MIT (Massachusetts Institute of Technology) with its figure of around 11%.

### **Publications: the quality challenge**

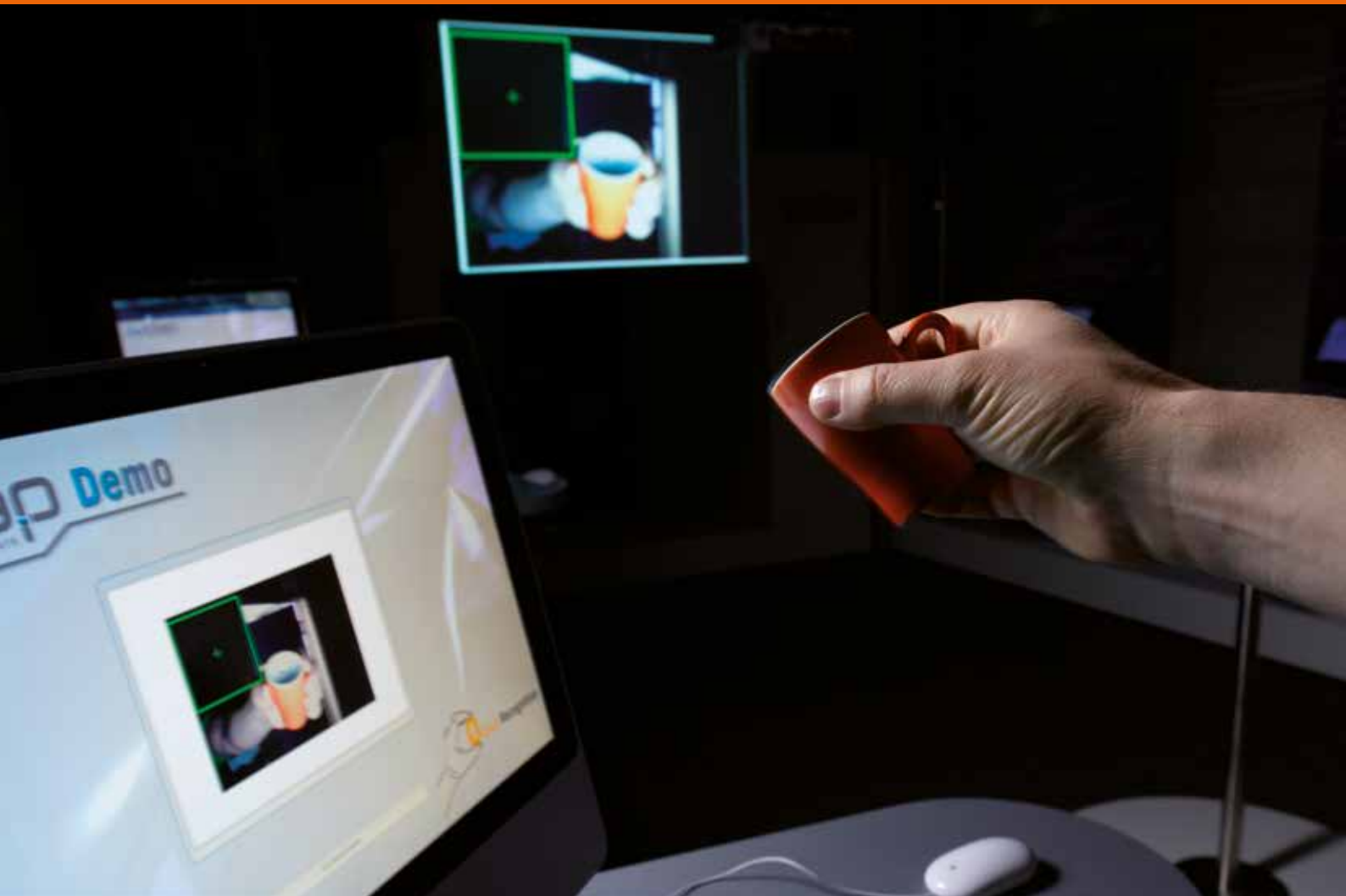
Our institute cultivates the quality not only of its research, but also of its top-level publications. In the scientific world, specialist journals are the yardstick. In 2013 we set a very ambitious target: for every three conference publications we would publish one paper in a scientific journal. The goal was so successfully achieved that I am now setting a new challenge for 2014: one publication in a specialist journal for every two conference proceedings published. While this new goal may seem beyond our reach, our latest statistics show that we have already taken a first step in this direction.

Our teams work tirelessly to meet new challenges. Again, I would like to congratulate and thank all our staff for their commitment to Idiap. To all, I wish to perpetuate the spirit of conviviality and creativity that drives us every day and creates such a pleasant work environment.

A handwritten signature in blue ink, reading "H. Bourlard".



RESEARCH



### AN INSTITUTE ON A HUMAN SCALE, BUT WITH AN INTERNATIONAL REPUTATION

With over one hundred employees, and research domains that address 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 21st 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 hand, 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 with the two federal institutes of technology, as well as 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 those in France, England, and Germany. Across the Atlantic, Idiap and the International Computer Science Institute (ICSI) in Berkeley, California, share a privileged partnership.

#### At a glance

**Structure** / The Idiap Research Institute is a non-profit foundation which specializes in the management of multimedia information and man-machine, multimodal interactions. The institute was founded in 1991 by the town of Martigny, the state of Valais, l'Ecole polytechnique fédérale de Lausanne (EPFL), the University of Geneva, and Swisscom. Although connected to EPFL via a joint development plan, Idiap remains an independent institution.

**Financing** / Of Idiap's CHF 10M budget, 60 percent is financed by competitively awarded research projects, and 40 percent from public funds (See "Distribution of sources of financing", page 32).

**Team** / As of 2013, Idiap employs more than one hundred people, 80 of whom are researchers (professors, senior researchers, researchers, postdoctoral students and research assistants.)

**Site** / Idiap has, since 2007, been located in the west wing of the Centre du Parc in Martigny, and currently occupies 2,500 m<sup>2</sup> on four floors.

### Missions

---

#### ■ Research

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 60 percent of the institute's financing.*

#### ■ Education

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, internally and at EPFL.

*One in two scientific contributors is a research assistant.*

#### ■ Technology transfer to industry

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 10 percent of Idiap's budget.*

*Thanks to the incubator The Ark at IdeArk, Idiap contributes to the creation of numerous successful start-ups.*



## FIVE RESEARCH THEMES & TEN APPLICATION DOMAINS

Research at Idiap is grouped into five research themes promoted via ten application domains.

### Five research themes

Since its foundation in 1991, Idiap has developed the core of its research around human-machine communication, and human-human communication facilitated by machines. Researchers at the institute therefore study the information that flows between these actors, and how it is — among other activities — perceived, understood, and processed. This topic is divided into five research themes:

#### 1 Perceptual and cognitive systems

(automatic speech recognition, computer vision, handwriting recognition, multimedia document processing, robotics, machine translation)

#### 2 Social/human behavior

(web and mobile communication, social interaction, social signal processing)

#### 3 Information interfaces and presentation

(multimedia information systems, personalization and contextualization)

#### 4 Biometric person recognition

(speaker recognition, face recognition)

#### 5 Machine learning

(statistical modeling, neural networks, mathematical models)



### Idiap in figures (2013)

#### Human resources

- 1 professor
- 3 senior scientists
- 13 permanent senior researchers
- 21 postdocs
- 40 research assistants
- 17 system engineers and development engineers
- 28 interns and visitors (average per year)
- 12 administrative staff
- 9 doctorates awarded
- 40 posts in IdeArk start-ups
- 34 nationalities represented

#### Scientific activities

- National Centre for Competence in Research — IM2 (Interactive Multimodal Information Management) from 2001 to 2013
- Participation in 44 research programs
- Project management in 12 consortia
- Participation in the economic development strategy of the Canton of Valais through the program The Ark and, in particular, the company IdeArk
- 273 scientific publications
- Participation in numerous international conferences



## Ten application domains

Whether through national or European projects in which it is engaged, or its close industrial partnerships, Idiap enhances its core competencies in application domains which are closely aligned to society's current preoccupations, including energy, security, mobile systems, and the exploitation of multimedia archives.

**1** Voice-controlled devices, voice-voice translation systems, navigation systems

### HUMAN-HUMAN, HUMAN-MACHINE & HUMAN-ROBOT INTERACTION

Semantic indexing, object recognition, audio-video content filtering, YouTube document analysis, analysis of cultural heritage media

### 2 EXPLOITATION OF RICH MULTIMEDIA ARCHIVES

**3** Smart meeting rooms, video conferencing, multimedia indexing and access, cross-lingual collaboration

### COLLABORATIVE AND CREATIVE SYSTEMS

Signal processing for mobile platforms, mobile social networks, mobile collaboration

### 4 MANAGEMENT OF MOBILE SYSTEMS

**5** Heterogeneous energy-networks, sensor networks, modeling of human activity to anticipate needs

### ENERGY MANAGEMENT

Smart patient (data-) management, prostheses (auditory, corporal, etc.), bio-systems modeling, interfaces for the disabled, capture and indexing of medical documents

### 6 HEALTH AND BIO-ENGINEERING

**7** Access controls (for physical or virtual spaces), speaker identification, face identification, video surveillance, natural-risk modeling, safety of public and private spaces

### SECURITY AND RISK MANAGEMENT

Ecology, environmental management, pollution reduction, traffic reduction, better use of road networks, noise reduction

### 8 CITIES OF THE FUTURE ("SMART CITIES")

**9** Multi-lingual gaming, remote-family games ("togetherness")

### ENTERTAINMENT AND GAMES

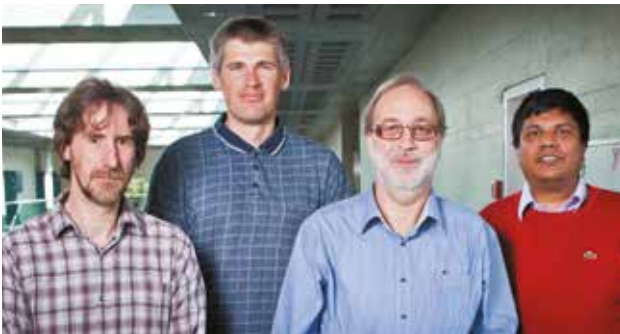
Multi-sensor activity analysis, human-behavior analysis, adaptation to human behavior (e.g. heating), home safety and security

### 10 HOME AUTOMATION (DOMOTICS)



## RESEARCH GROUPS

Eight 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 four senior researchers. The remaining groups are also headed by senior researchers.



### Speech & Audio Processing

**Prof. Hervé Bourlard** (*second from right*), **Dr. John Dines** (*not in the photo*), **Dr. Phil Garner**, **Dr. Petr Motlicek**, **Dr. Mathew Magimai Doss**

Speech processing has been one of the mainstays of Idiap's research portfolio for many years. Today it is still the largest group within the institute, and Idiap continues to be recognized as a 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).



### Perception & Activity Understanding

**Dr. Jean-Marc Odobez**

This group conducts research in human-human activity analysis from multi-modal 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 their interpretation in the forms of gestures, activities, behavior, or social relationships, 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.



### Computer Vision & Learning

**Dr. François Fleuret**

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, tracking of people and biological structures, and the understanding of images in general.



### Social Computing

**Dr. Daniel Gatica-Perez**

Social computing is an interdisciplinary domain that integrates theories and models from mobile and ubiquitous computing, multimedia, machine learning, and social sciences, 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.



### Artificial Cognitive Systems Dr. Barbara Caputo

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 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.



### Biometric Person Recognition Dr. Sébastien Marcel

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 (2D, 3D, 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.



### Applied Machine Learning Dr. Ronan Collobert

This group is interested in computer algorithms that can "learn" a behavior in order to perform a given task of interest, in contrast to algorithms with behavior constrained by handcrafted rules. Research is driven by real-world applications involving large amounts of data. Domains of interest include natural language processing, computer vision, and audio processing. A particular emphasis is placed on generic machine-learning tools which require minimum a priori knowledge of the data (such as deep-learning techniques), and on unsupervised learning techniques which can leverage inherent semantics from large-scale, structured data.



### Natural Language Processing Dr. Andrei Popescu-Belis

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 recommendation of information from networked multimedia repositories can be improved

by analyzing a network's structure and the information content of its nodes.



### IDIAP'S CRADLE AND GROWTH ACCELERATOR

Idiap led the National Centre of Competence in Research — IM2 (Interactive Multimodal Information Management) for twelve years from 2001 to 2013. An extremely positive experience for the institute.

It is the late '90s and the Swiss National Science Foundation (SNSF) launches its National Centres of Competence in Research (NCCR) program. Its aim: to strengthen Swiss research in areas of strategic importance — including science, economics, and society — by supporting major projects.

The country's research institutes respond to the foundation's call, submitting around three hundred proposals. Fifteen are selected, including that of Idiap. In its tenth-anniversary year, the institute kicks off IM2 (Interactive Multimodal Information Management), one of fourteen NCCRs funded in this first wave — an exceptional event for a non-university institution.

#### **An astonishing accelerator for Idiap**

The adventure will last twelve years — "Longer if one considers that our project preparations and the selection process took nearly two years", says Hervé Bourlard, director of Idiap. "One could also say that this period corresponds to three generations of doctoral students."

As the moment comes for taking stock, Bourlard is enthusiastic: "IM2 was a great experience and a wonderful opportunity. It was the cradle of Idiap. It acted as an accelerator allowing our institute to attain more rapidly the level at which it finds itself today."

Prestigious, and encouraging by virtue of the funding that comes with it, the direction of an NCCR is also extremely demanding. "Being granted an NCCR is difficult — keeping it equally so", says Bourlard. Each year, a committee of external experts evaluates the work carried out, and the mandate is granted only for a four-year period, after which it can be renewed — the maximum duration of an NCCR is fixed at twelve years.

#### **An expanded scientific community**

The benefits of IM2 are obvious in various domains. In research terms, the project further honed Idiap's expertise in the field of interactive multimodal information management (speech recognition, document indexing, recording and analysis of meetings, etc.) This interdisciplinarity, which itself represents a unique opportunity for the training of doctoral students, also helped build a large scientific community, forging links, for example, between American and European researchers. IM2 has helped to significantly raise Idiap's international profile and boost the institute's global reputation. Under its auspices, the International Create Challenge (ICC) contest was established in 2012. And IM2 encouraged the development of more than twenty start-ups. Born from Idiap technologies, some are now being sought out by prestigious clients and have experienced significant expansion — as is the case, particularly, for Klewel, Koemei, and KeyLemon.

And IM2's influence on Idiap will not stop in 2013 — Hervé Bourlard is convinced of that. "Idiap emerges from this adventure stronger than ever. What we have learned will help us tomorrow. It encourages us to stay on the course we've set for ourselves, in terms both of our chosen areas of research, and of technology transfer."



*Between 2001...*



*... and 2013, Idiap changed in scale.*



## THE FINAL EVENT — TAKING STOCK

Representatives from the fields of science, politics, and the economy participated in the "final event" — held on 17 and 18 October — that marked the closure of IM2. Two days to look back on the adventure and sketch out future prospects.

Hervé Bourlard rightly felt both pleasure and pride as he welcomed participants to the event marking the end of IM2 — Ildiap's twelve-year "thrilling adventure". "One door may well be closing, but others are certainly opening", Ildiap's director announced to the guests gathered together at EPFL.

Nearly 120 people took up the organizers' invitation: researchers, alumni, representatives of start-ups, and policy makers involved in IM2, but also individuals active in research areas similar to those pursued at Ildiap.

### A positive structural impact on the region

Besides providing participants with an opportunity to review the results of IM2's flagship activities, these two days of meetings and workshops also allowed them to reflect on the impact of major, long-term research projects. Their positive structural impact on a region or country was evoked by Steve Flinter, scientific programme manager of Science Foundation Ireland, the Irish agency charged with funding research in strategic fields such as biotechnology, information, and communication.

Sandy Ingram, head of R&D at Martigny-based Klewel SA, explained how IM2 had helped create the company and enabled it to grow inside Ildiap. François Foglia, deputy director of the institute, spoke of the International Create Challenge (ICC), the contest — launched in partnership with IM2 and IdeArk — which supports young researchers' entrepreneurial projects.

While these two days allowed participants to look back on IM2, they also provided an opportunity to think about the future. A future in which Professor Bourlard is confident: "Quality research always finds funding; and that will only continue."



### Twelve years of research in book-form

The conclusion of IM2 was also marked by the publication of a book — "Interactive Multimodal Information Management". Edited under the direction of Hervé Bourlard and Andrei Popescu-Belis, the volume outlines just some of the achievements of the project's twelve-year lifespan. Those who participated in the project comment on their work, while one entire chapter focuses on how those from outside the institute viewed the NCCR.

A copy was presented to all who attended the IM2 closing event.

*Interactive Multimodal Information Management*  
EPFL Press, 2013



## INSPIRED BY IDIAP RESEARCH — MARKETED BY APPLE

Inspired by a technology born at Idiap during the NCCR IM2, the multi-directional microphone *Microcone* is now sold at the Apple Store.

Former Idiap researcher, now head of an Australian start-up, Iain McCowan has reached an agreement with Apple to sell his multi-directional microphone *Microcone* through the Apple Store. The device, shaped like a small pyramid, contributes significantly to meetings involving a limited number of participants, simplifying their recording, marking, and analysis.

The company that developed this meeting microphone — Dev-Audio — may well be Australian, but it was at Idiap, where McCowan worked as a researcher from 2001 to 2005, that the first research into "smart meeting rooms" was carried out. Indeed, the institute's expertise in automatic language-processing led to its launching numerous research projects on the concept of the smart meeting room, in which a network of microphones makes it possible to determine who is speaking during a group discussion.

*Microcone* records up to six speakers on separate audio channels, capturing and automatically improving each voice while reducing background noise. Each word, whether spoken in a business meeting, interview, or podcast, is clearly captured.

Idiap is delighted with Iain McCowan's tremendous success, and is glad to see that its research has — once again — led to the creation of a device that makes life easier.



*The forerunner of Microcone, created at Idiap...*



*... the device currently sold by Apple.*

### IM2 IN FIGURES

- 1 funding source — the Swiss National Science Foundation (SNSF)
- 6 Swiss partners:
  - Idiap
  - Two federal institutes of technology - Lausanne and Zurich
  - Three universities - Geneva, Fribourg, and Bern
- 12 years of research, training, and technology transfer
- 88.1 million Swiss francs in funding over twelve years (CHF 33.8M from the SNSF and CHF 54.3M in matching funds from partners)
- 1.6 Swiss francs invested by partners for each franc granted by the SNSF
- 2,168 publications (more than 180 per year)
- 57 contributions to European or American projects
- 23 start-ups created, thirteen of which are still active
- 72 industrial partners

### The association AIM2 — extending IM2

Founded in March 2012, AIM2 (the Association for Interactive Multimodal Information Management) aims to ensure that IM2's technological and economic achievements endure. Chaired by Idiap's deputy director, François Foglia, the association is pursuing that goal by focusing on three specific domains: research, technology transfer, and training.

AIM2 brings together five academic partners: Idiap, EPFL, ETHZ, and the universities of Geneva and Fribourg.

## THE COLLECTIVE MEMORY OF VALAIS — BY THE PEOPLE, FOR THE PEOPLE

Conceived by Idiap to celebrate the 200th anniversary of the Canton of Valais' entry into the Swiss Confederation, the project "Valais\*Wallis Digital" is fun and brings people together. Archiving and sharing memories will soon be child's play.

An Internet platform, a smartphone app, and a card game — just some of the tools that Idiap and its partners are using to build the Canton of Valais' collective memory. The ambitious Valais\*Wallis Digital project was selected by the canton in the context of its bicentennial celebrations. No need to be a technology wizard to get started. A simple Internet connection is all that's required to upload and share documents via the application or the platform.

Imagine a young girl and her grandfather: "Say, what was school in your village like when you were small?" The old man's answer, filmed on a smartphone, can be uploaded to the Valais\*Wallis Digital platform, which — thanks to Idiap technology — can then index the contents of the file, filter the information according to its relevance, and create new links. Voice- and face-recognition tools developed in related research projects can identify content elements and multiply the connections between them. For example, a photo of the Matterhorn entitled "Zermatt" could be added to the stock of images dedicated to the mythical Valais mountain. Individuals, municipalities, schools, and associations will be invited to submit their records and documents from 2015 on. The plan is for this collective memory initiated by Idiap to be, in part at least, taken on by the Canton of Valais' multimedia library and cantonal archives, partner institutions for the project.



**François Foglia, deputy director of Idiap and initiator of the project "Valais\*Wallis Digital"**

### Cards to collect

To encourage natives of Valais of all generations to participate in this online heritage project, a card game depicting 200 events that took place between 1815 and 2015 in "the Old Country" will be distributed in Migros stores. The illustrations on the cards were created by young artists from the Professional School of Contemporary Arts (EPAC) at Saxon. Thus, by bringing together partners from all backgrounds around a common challenge, Idiap has found the perfect way of putting its technology at the service of the public and the canton in a sustainable manner. The institute will call on its national and European networks to launch new research projects linked to the challenges of Valais\*Wallis Digital (e.g. the development of image-analysis content-detection algorithms).



**Damian Elsig, Director of the Valais multimedia library, project partner**

### "This collaboration opens new horizons for us"

*The mission of the Valais multimedia library is to preserve and promote the Canton of Valais' audiovisual heritage. What will you do with the data from Valais\*Wallis Digital?*

We don't know yet. The project is a fascinating "laboratory", the results of which we will discover during the process of its creation. The project invites us to launch ourselves into the world of social networks, and at the same time to stop and think about the electronic heritage of this canton. As a first step, we'll define selection criteria that will enable Idiap to automatically pre-sort the data. From 2015, we'll see what kind of material the people of Valais are uploading and what parts of it we want to save. We may also be able to use this opportunity to consult the people of Valais on the subject of other research themes, for example.

*This project also provides you with an opportunity to work quite concretely with Idiap. What do you expect from these new information-processing technologies?*

This is a new dimension opening up for us, a dimension with tremendous potential for our activities. Currently, image processing — from scanning to cataloging — is very time-consuming. Idiap's technologies could help us to much more rapidly put together photographic archives, with pre-sorting and a pre-description. So I'm very interested in what comes out of this new collaboration with Idiap. The pace of social networks is only going to accelerate further; we need help capturing and sorting interesting content.



# PROJECT SIIP

## THE VOICE — THE SIGNATURE AT THE VERY HEART OF SECURITY

Idiap is the only Swiss institute engaged in a major European project to identify suspicious individuals from their voices. Interpol, which provides the project's researchers with the necessary data, will be among the first organizations to test the system.

In autumn 2012, the European Commission launched one of the final calls of its Framework Programme FP7. Among the many projects submitted by the world's most renowned research centers, the commission selected that of Idiap. In spring 2013, Hervé Bourlard was in Brussels to present his institute's proposal, entitled SIIP (Speaker Identification Integrated Project). Its scope: security.

"The goal of the project", explains Petr Motlicek, "is to develop a new system for the identification of suspects by their voices." Motlicek works on the project alongside Idiap's director, and one of the institute's senior researchers Sébastien Marcel.

### Combining several feature in one device

The future system will not only identify who is speaking, but also determine which language is being used and the person's gender, and estimate the speaker's age. It will also be able to identify words that recur frequently in the person's speech. "Even if the suspect uses a code word instead of words like 'bomb' or 'cocaine', the device will detect that something is suspect", says Motlicek, a specialist in speech- and audio processing.

Although applications capable of performing one of these operations do currently exist, project SIIP will improve on such capabilities and, more importantly, combine these different features in one device. It should also be able to establish whether the speaker is trying to disguise his or her voice (by pinching the nose, putting a hand over the mouth, putting on an accent, etc.) or mimic that of another person (generated by a machine). "Today, these problems remain to be solved, as do those of noise caused, for example, by the passage of a train or a car, which parasitizes the system."

### Thousands of hours of recordings "listened to" by machines

Idiap researchers will work on data provided by Interpol (the International Criminal Police Organization), a body that groups together 190 member countries. A corpus will contain thousands of hours of recordings to be processed for the purposes of identification (find a given voice in different files) and verification (do two voices belong to the same person?)

"Speech analysis doesn't provide us with results that are as reliable as fingerprints or DNA. But, by using this device, we can attain a confidence level that can be very high", says Motlicek. Man is not entirely excluded from the process however. Imagine that the device identifies three instances of the same voice, with probabilities of 98%, 80%, and 20% respectively.

### Seventeen partners, ten countries

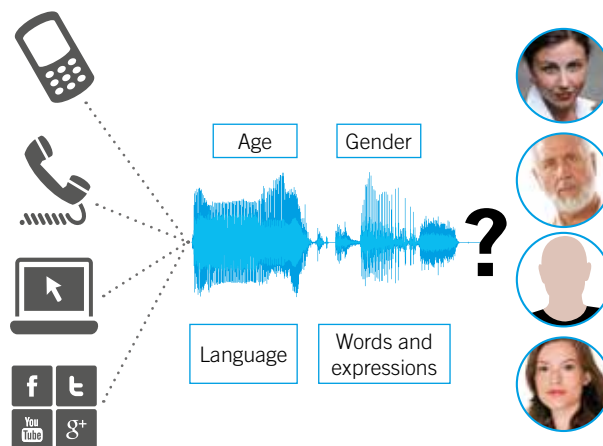
Project SIIP brings together seventeen partners (universities, companies, police organizations) in ten European countries including France, Portugal, and the UK. The project is coordinated by Verint, an Israeli company specializing in the field of security.

The project is expected to begin in spring 2014 and run for four years.

It is now the turn of people to listen to these three recordings in order to conclude if the voice belongs to the same speaker or not. "The human ear is capable of a finer analysis. But people can't listen to thousands of hours of recordings and their memories aren't infallible: a voice is quickly forgotten."

### From banks to homes — a system with a range of applications

Petr Motlicek is already imagining the uses to which banks could put the future system — the client's voice becoming the passport that guarantees access to that client's accounts. Home automation would also benefit, for example with a TV that could only be switched on by the voices of parents, and not those of children. "Ultimately, we hope to adapt the software to systems like Skype, Google Talk, etc. (VoIP [Voice over IP]), and messages written either on social networks or by e-mail."



*The system will not only identify who is speaking, but also determine which language is being used and the person's gender, and estimate the speaker's age; it will also identify individuals from the messages they write, either on social networks or by e-mail.*





# NETWORK



## RESEARCH AND TRAINING — A FRUITFUL PARTNERSHIP

In February 2013, Idiap and EPFL renewed their existing collaborative agreement putting their respective signatures to a new four-year joint development plan. Between the Lausanne academic giant and the dynamic institute from Martigny, this is a strategic alliance.

The Ecole polytechnique fédérale de Lausanne (EPFL), as its French name suggests, trains engineers. The Idiap Research Institute, according to the same logic, employs researchers. Yet EPFL also conducts research and Idiap contributes to the education of doctoral students: research and training, naturally, have interests in common.

### Training the promise of the new generation

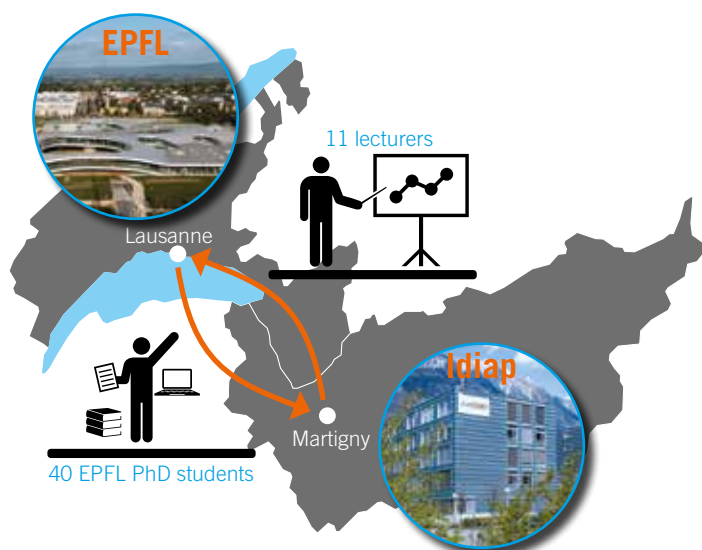
Knowledge is not an inert material. Mankind's knowledge, particularly in the technical sciences, changes at the pace of new discoveries. Training itself can only gain by not losing touch with the reality on the ground, while researchers — by transmitting their know-how — help train quality new generations to come.

In 2008, with this strongly held belief in mind, Idiap and EPFL signed a joint development plan, including — for Idiap, and following the logic of interdisciplinarity — a strategic alliance with the domain of the Swiss federal Institutes of technology in general. On 8 February 2013, the two institutions committed to continuing their collaboration and joint development of scientific and academic activities of mutual interest for a further four years.

### Academic recognition for researchers

For Idiap, this agreement is important as it provides a form of academic recognition to the institute's researchers, hence reinforcing the attractiveness of its research positions. Today, several of the institute's scientists have EPFL titles and lecturing responsibilities, while simultaneously supervising PhD students writing their theses at Idiap. As for Idiap's director, Hervé Bourlard, for many years now an EPFL professor, he directs the EPFL faculty of Science and of Engineering's (STI)

LIDAP laboratory. It is to this branch of Idiap at the heart of EPFL that senior scientists (MERs) François Fleuret, Daniel Gatica-Perez, and Jean-Marc Odobez — all three senior researchers at Idiap — are affiliated.



"Today", explains Bourlard "around a third of our staff are PhD students who will be awarded their qualifications by EPFL. They have the opportunity, by carrying out their practical work with us, to work in our world-class research teams, and our institute — in turn — forges privileged links with the scientists of the future. Indeed, our reputation is such that many students now enroll at STI in order to be able work on their PhD with us."

### Ten million in federal grants for 2013–2016

Following the signing of the renewed Idiap-EPFL joint development plan, Idiap saw the Secretariat of State for Training, Research, and Innovation (SERI) illustrate its continued confidence in the institute by approving a grant of CHF 10M for the period 2013–2016; CHF 3.5M more than for the previous four-year period.

This grant will enable the institute to further strengthen its financial base and to maintain its research, development, and technology-transfer activities.

"All lights are on green for the next four years", explains Hervé Bourlard. "On the condition of an increase in federal funding, the Canton of Valais agreed — in late 2013 — to increase its own support by CHF 0.5M per year. Great news for us, especially as we're now running two projects in the canton, and are engaging more and more in technology-transfer projects operating at the regional level."



# INTERNATIONAL CREATE CHALLENGE

## THE EXPRESSWAY TO ENTERPRISE CREATION



The 2<sup>nd</sup> edition of the International Create Challenge (ICC), an accelerator program launched by Idiap in 2012, was a resounding success: 80 proposals received, three winning projects. Luma7, POWZY, and ReMeeting may, tomorrow, be changing the way we go about our business.

The call for proposals was published in spring 2013: "Turn your idea into a start-up!" Participation criteria: propose an innovative

project of high scientific quality, preferably in a research field close to those pursued at Idiap. By mid-June, the institute had received no fewer than 80 proposals from across the world, proposals delivered by creative young people, keen to take advantage of this unique opportunity. A total of ten teams, comprised of 33 people from 15 different countries, made it through to the final on 21 September in Martigny. "You have all already won", François Foglia deputy director of Idiap told them, "because you will all finally be able to focus exclusively on seeing your projects progress."

### Idiap congratulated by IBM, Nokia, and Thalès

Initiated in 2012 by Idiap, the ICC offers successful applicants a three-week, 24–7, total-immersion, all-expenses-paid acceleration program — participants can concentrate exclusively on moving their idea from project to start-up. During this period, the young entrepreneurs can take advantage of Idiap's infrastructure and benefit from the support not only of the institute's administrative staff and researchers, but also that of scientists from across the country, and can as well consult an

array of top-level professionals (including coaches and venture capital specialists). For the occasion, Idiap expressly invited some of its prestigious industrial partners including Nokia, IBM, and Thales, all of whom mentioned in their presentations Idiap's flexibility, dynamism, and creativity.

### Significant regional support

On October 11, as the program drew to a close, Luma7, POWZY, and ReMeeting each received CHF 10,000 while Luma7 — on the condition that its creators launch a start-up in Valais — also clinched a three-year support deal at the incubator IdeArk, worth approximately CHF 100,000. For, while technology transfer is one of Idiap's missions, the institute also wishes to create jobs in the region by ensuring that such technology transfer generates new businesses. The Idiap initiative was further encouraged by the support of the city of Martigny, the Valais-romand Regional Office (l'Antenne Régions Valais romand), and the Ark foundation via the incubator IdeArk, as well as that of Eurofin Ventures, IMD (Lausanne), and the Hotel Vatel (Martigny).

It is worth noting that, among the participants in the 2012 edition, five projects are still running and under development.

[www.createchallenge.org](http://www.createchallenge.org)



### Luma7, the idea generator

Graphic designer Nelio Barros dreamed up Luma7, an automatic Mind-Map generator. By simply entering, for example, *Napoleon*, the system proposes a galaxy of ideas from *Elba* through *Joan of Arc* to *submarine*.

[luma7.com](http://luma7.com)



### POWZY, non-intrusive mobile publicity

The idea behind POWZY: bring to business the pleasure of game playing and boost smartphone-user loyalty by providing users with offers that are linked to what they are doing on their screen (play, find their way, etc.)

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



### ReMeeting, taking meeting minutes

ReMeeting is a mobile app that incorporates a personalized cloud service that enables users to record, annotate, analyze, transcribe, and archive meetings intelligently. Two natives of Valais are involved in the project, one of whom is David Imseng, a postdoc at Idiap.

[www.remeeting.com](http://www.remeeting.com)

## THE THREE KS HAVE THE WIND IN THEIR SAILS

We knew they were growing, but now they are ready to blossom and bear fruit. These young offshoots of Idiap technologies are among the "50 start-ups to invest in" according to *Bilan* magazine. Overview of a year rich in projects and contracts for Klewel, KeyLemon, and Koemei.

Idiap start-ups were among the top 50 Swiss start-ups selected by business magazine *Bilan*. In its February 2013 edition, *Bilan* encouraged investors to buy into these innovative, young companies that need funds in order to grow — "With patience, you can expect a five- to ten-times return on investment in three to five years." "Business-angel" clubs have been formed in western Switzerland to advise investors on how to place their capital intelligently. In 2013, the three Idiap spin-offs have overcome significant hurdles and now, together, employ around twenty members of staff.

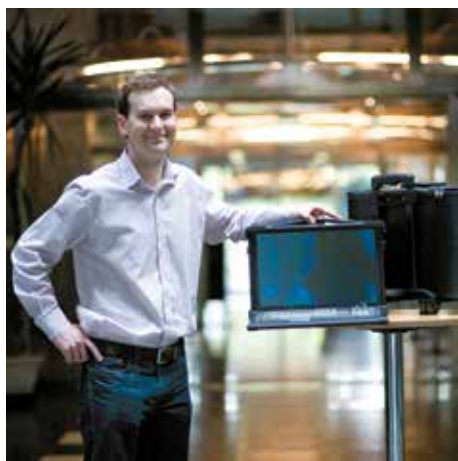


## Klewel

### Klewel — Serving audiovisual professionals

Triskel is now a finished product, ready to conquer the market. The launch version of this webcasting tool has, over the last few years, demonstrated its advantages to major international clients. The start-up has also managed to self-finance thanks to the services it provides to conference organizers. Product development has continued via participation in national and European research projects, concerning — among others challenges — the integration of speech recognition and automatic content recommendation and the creation of a conference web-platform.

Klewel finalized Triskel in 2013, creating a quick and easy-to-use turnkey product covering everything from event recording to online publication. Its founder, Maël Guillemot, welcomes this new step: "We're very pleased with the result. We now have to convince audiovisual agencies to use the product, because it's they who manage the technical aspects of conferences. It'll take time, and we're going step by step. Instead of wasting days on editing conferences, those agencies could synchronize their audio recordings, videos, and presentation slides instantly, and publish content in just a few minutes. "



**The start-up at a glance:** since 2007, Klewel has provided innovative solutions for the recording, archiving, and web-distribution of presentations and conferences. Its main tool, Triskel, is a turnkey webcasting solution composed of a suitcase-size audiovisual-recording station. This is linked to a web platform that allows users to index, edit, and publish audio and video of a speaker's presentation (slides and other supports). Anyone can then consult the presentation on the organization's website, and search for specific content using keywords. In 2008, Klewel's solution was awarded the European Seal of Excellence at the CeBIT international trade fair; in 2010 it was selected as a finalist for the International Association of Congress Centres' (AIPC) Innovation Award. Today, Klewel regularly provides conference-recording services to well-known organizations including Nestlé, Vaud Cantonal Bank (BCV), and Shire Pharmaceuticals International, and is a partner of the Geneva International Conference Centre (ICCG).

[www.klewel.com](http://www.klewel.com)





### KeyLemon — Investor confidence

Thanks to a spectacular fund-raising effort in 2013, KeyLemon is taking off in the global marketplace. With the help of Debiopharm Investment and Swisscom Venture, the start-up was able to secure CHF 1.5M in investment. These funds will not only enable KeyLemon to develop new technologies, they will also help establish the company's credibility. "It shows that people believe in us and that our product has great potential. This is a very strong signal of confidence for future investors," says Gilles Florey, CEO of KeyLemon, already preparing for a second round of funding in 2014. The start-up has hired five new staff members to reinforce product development and sales. It is now represented in the United States, Asia, and Europe. KeyLemon is



currently working in e-learning and the automotive industry, and on processor algorithms. Its face- and voice-recognition technology will, for example, enable clients to verify if the right student sits the right exam, or alert drivers falling asleep at the wheels of their vehicles.

**The start-up at a glance:** created in 2008 by an Idiap graduate student, an economist, and a professor of entrepreneurship, KeyLemon developed facial-recognition software that allows users to open a session on their computer by presenting their face to the screen. The application has over three million users and KeyLemon has already licensed its technology to a company that supplies banking solutions throughout South America, and to a European medtech company. In 2012, KeyLemon signed a partnership with Fujitsu, and the start-up's technology will be pre-installed on an entire range of the Japanese giant's computers.

[www.keylemon.com](http://www.keylemon.com)

## KOEMEI

### Koemei — A technology pioneer

2013 was a great year for Koemei. The start-up active in the field of voice transcription was selected by the World Economic Forum as a Technology Pioneer. With this prestigious award, Koemei takes its place in the "community of pioneers" alongside Google (2001), Mozilla (2007), and Dropbox (2012). For Temitope Ola, founder of Koemei, inclusion in the 2014 list of the 36 most innovative companies in the world is an extraordinary springboard. In 2013, the start-up opened up to the US market by moving to San Francisco. It managed to raise an additional CHF 0.5M of capital, bringing its funding to CHF 1.3M. Further



incoming investment is expected in 2014. Koemei has launched a solution for online video training, providing students and teachers with information-retrieval and collaboration tools. We knew of the enormous potential of voice transcription: investors and clients alike are now willing to trust Koemei.

**The start-up at a glance:** By 2015, ninety percent of web content will be in video format. There is, however, no reference system for such content. Koemei offers a voice-transcription solution that distinguishes between different speakers and allows search engine access. Koemei's solution is a favorite of higher education institutions and has been adopted, for example, by the University of Geneva, Lausanne's IMD business school, and the University of California at Berkeley. Koemei is also present in the market for closed-captioning for the deaf and hard of hearing.

[www.koemei.com](http://www.koemei.com)

# COMPUTER INFRASTRUCTURE

## AT THE SERVICE OF GLOBAL DATA-SHARING

Each year, 13 terabytes of data leave Idiap on their way to the global scientific community. Even if sharing is extremely common in the research milieu, it still requires a certain framework, as Frank Formaz, head of the institute's System and Infrastructure department, explains.

### *What data is exchanged between Idiap and the scientific community?*

Our institute is particularly active in developing systems capable, for example, of recognizing your face when you switch on your computer or of finding one word in an entire video. People learn to recognize an image or a sound by acquiring, with the help of their senses, huge quantities of data. Machines, on the other hand, need to be supplied with that data. Thus our computers are programmed to handle very large quantities of data — sound, images, video, text, etc. These interest other researchers elsewhere in the world, and our institute also receives other data from its partners. What's more, our scientists develop and use customized software applications in their work. These applications are also shared with the scientific community, often in open source formats.

### *Data and software are being exchanged constantly?*

Indeed, this is the principle of sharing that prevails within the global scientific community, which naturally enables research to progress more rapidly. For us then, it's important to facilitate this exchange, while at the same time respecting the legal framework in place. It's also worth noting that the volume of data being exchanged constitutes only a minuscule proportion of the total data housed at Idiap.

### *Why a legal framework?*

The data we work on are, in the legal sense, considered to be personal and sensitive. They include — for example — voice recordings, videos, and photos. Their use is subject to the Federal Act on Data Protection (FADP) and is regulated by specific contracts and licenses that impose controls on access and usage. We therefore have to guarantee these aspects while ensuring the fewest possible constraints for our researchers.

### *What does such a constant exchange entail for you?*

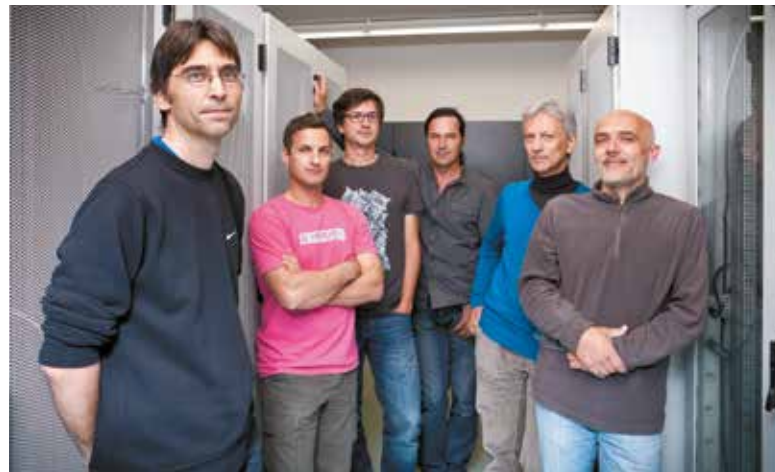
Idiap has around a hundred employees who between them request no fewer than 4,000 hours of computations each day on these data. If a hundred people want to do the same thing at the same time, they have to be able to do so by making the best use of available computing resources. It's our job therefore to provide a storage and computational infrastructure that meets these needs and the tools necessary for exploiting it as effectively as possible.

### *Specifically, how is this infrastructure set up?*

At Idiap facilities are shared in order to allow everyone access to the maximum measure of our resources, both for storage and for computation. Our servers, equivalent to approximately 200 computers and some 400 terabytes of storage, are concentrated in ten computer cabinets. This infrastructure is the heart of the institute's activities and hence has to function 24–7.

### *And if the power fails?*

The effects have to be minimal. This is why we have system redundancy. Whatever happens, everything is set up so that the system never stops and so — even if we have a serious problem — all data is preserved.



*Frank Formaz (left) and his team provide Idiap researchers with topflight computation infrastructure.*



FACES







## PHILIP GARNER, SENIOR RESEARCHER

### "LIFE ISN'T A SERIOUS BUSINESS"

With his British humor and his tousled hair, one can't miss Phil Garner at Idiap. In Martigny for seven years now, the researcher — specialist in speech processing — appreciates the *bon vivant* feeling of living in Valais.

Phil Garner doesn't have a hobby. Even the guitar lying around in his office can't distract him from algorithms. "I've been trying to play for 20 years — it's a waste of time. All audio/video engineers are failed musicians!" For Phil, artistic expression is less important than how the instrument itself works. Born on the outskirts of Manchester (UK), Phil grew up building and fixing things with his father, a telecommunications engineer. It's a passion that has never left him. Over the years and thanks to his academic studies, he added a scientific dimension to his activities. "The researcher tries to understand what already exists; the engineer tries to solve technical problems. I have both these sides to me. I spend my time solving problems and finding new ones. But basically, I like that. I mean, life isn't a serious business, is it?"

Evenings, while his pianist wife practices her scales, he's still programming. He has kept, from his English background, a microwave oven and the pleasure of going out to bars. "People from Valais are never so nice as when they're serving drinks and clinking glasses. I've discovered the world of wineries and winemakers. It's fabulous, even if wine is a far more dangerous alcohol than beer... it doesn't, anyway, help me improve my French."

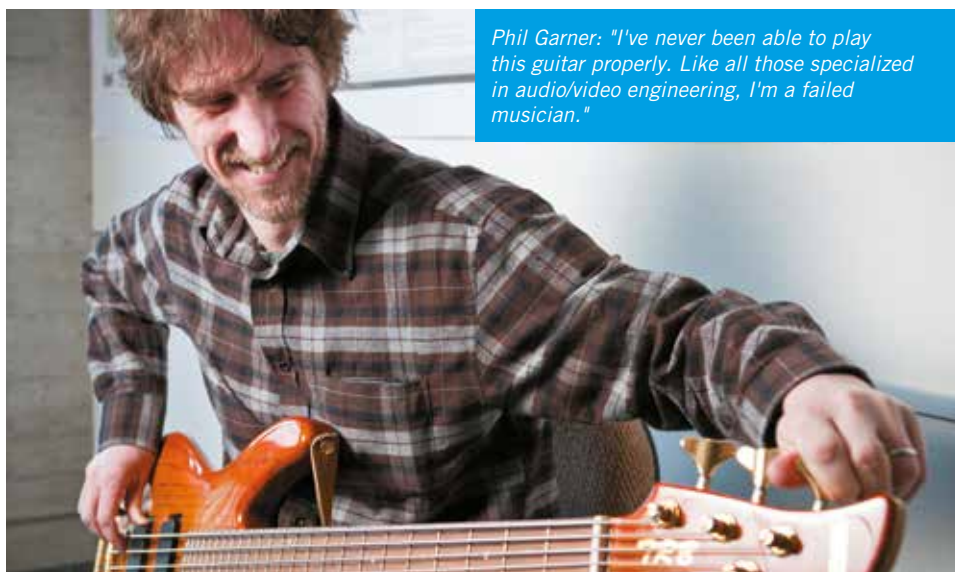
#### Talking 'Cat'

Phil readily confesses, not without a hint of guilt, his lack of enthusiasm for the French language. There are, however, extenuating circumstances. He arrived in Martigny in 2007 after six years working for Canon in Tokyo. While there he learned Japanese and continues to speak it on a daily basis with his wife. At Idiap, English is the common language of researchers, even when it comes to having a drink together after work.

But French is there the moment he steps out of the house or his office. In the end, for this speech specialist at least, communication isn't always as simple as it seems. "Sometimes I find it more relaxing to talk to my cats", he jokes. With his dry wit and his sharp view of the world, Phil Garner always gets his way. "It's enough to ask someone from Vaud Canton 'Do you speak German?' or someone from Zurich 'Do you speak French?' to have them both happily speaking English."

#### A friendly atmosphere

The 44-year-old Brit enjoys the quiet life and the sun of Martigny. Apart from his regular trips to EPFL in Lausanne where he teaches, Phil spends most of his time in the friendly atmosphere of the institute. "The time spent with colleagues at the break or downtown is very important. The wealth of our contact is the wealth of our projects."



*Phil Garner: "I've never been able to play this guitar properly. Like all those specialized in audio/video engineering, I'm a failed musician."*



## VASIL KHALIDOV, SENIOR DEVELOPMENT ENGINEER

### "THE QUALITY OF THE INFRASTRUCTURE AT IDIAP IMPRESSES ME"

The attractiveness of a state-of-the-art research institute; the charm of a small town. Vasil Khalidov, 30-year-old senior development engineer, has found in Martigny everything he needs for his professional and personal fulfillment.

*You're from Saint Petersburg. Was it there where you learnt French?*

I learned French in Grenoble where I did my PhD in applied mathematics. The French language isn't essential in the world of research, but I wanted to learn it for personal reasons and to facilitate my integration. When I arrived at Grenoble's Inria Research Center, I asked my colleagues to speak to me only in French. My team members were patient enough to do so, even when I wasn't able to say two words in the language.

*Did you teach them basic Russian in return?*

No, but when I organized a trip to Saint Petersburg for them, I taught them a few expressions. Just some short, useful phrases so they could avoid being scammed as some tourists are.

*You've been working at Idiap for three years now. How did you first come to learn about the institute?*

While I was in Grenoble, I saw a lot of interesting work coming out of Idiap and I met a few of its researchers. Thanks to a European project in which Inria and Idiap both participated I made the acquaintance of Jean-Marc Odobez, who was *rapporteur* for my thesis. When a postdoctoral opportunity at Idiap presented itself in November 2010, I didn't hesitate for a moment.

*What struck you most upon your arrival at Idiap?*

I was particularly impressed by the quality of infrastructure here. In addition to the level of research conducted at Idiap, research which is in the vanguard of its field, I found the atmosphere here very friendly. On a more general note, I was struck by the proximity of the mountains, even bigger than in Grenoble, which is great for hiking. I didn't choose Idiap because of the mountains ... but I'm very happy to be living in such a mountainous region.

*How do you feel about Martigny, where you chose to live?*

My wife Olga and I met in Grenoble; in Russia that would have been an unlikely meeting as we lived almost 4,000 miles apart. Today we're very happy in Martigny, a town which — despite its modest size — is extremely well equipped. You can feel the love that the inhabitants have for their city. I'm thinking, for example, of how the Place Centrale has been remodeled, or of the Pierre Gianadda Foundation that we visit regularly for exhibitions.

*You went to music school in Russia. Do you still play piano?*

Yes, I do. Music gives my head a rest from my work. I often go to see the operas screened at Martigny's cinema and I sing in the TouTanKhanon choir. There are twelve of us — three per register — and we interpret French variety artists like Joe Dassin, who I already knew of in Russia, or Gainsbourg, who I discovered here. In general, I'm always fascinated by new things, whatever the field.



#### The pleasure of presenting your work to a wider public

In his daily duties as a senior development engineer, Vasil Khalidov can draw on his knowledge both of computing and of applied mathematics. "I'm very happy in my work because our research should help to improve people's quality of life." With this in mind, Vasil takes great pleasure in bringing the fruits of his labors to a wider public, for example by attending public demonstrations of the humanoid robot NAO. "It's interesting to see how people, whatever their age, are impressed by its capabilities. I remember a demonstration at *Hérisson sous gazon*, the children's festival organized at Charrat: the kids were amazed!"

## DAVID IMSENG, POSTDOC

### "FOR A VALAIS-LOVER LIKE ME, WORKING AT IDIAP IS AMAZING"

It was while studying at EPFL under Hervé Bourlard, Idiap's director, that *Haut-Valaisan* David Imseng found his future path. Interview with a passionate young researcher.

"I love my life." If the statement is decidedly positive, it is only a just reflection of the man who makes it. In David Imseng, 31-year-old postdoc at Idiap, there is enthusiasm and modesty in equal measure. A passionate young man and enthusiastic scientist, he has kept his taste for life's simple pleasures.

#### **Improve the community's well-being through research**

This is how this keen speech-recognition researcher sees his work, primarily as a way of improving the lives of others. "Think about what speech recognition can bring to the hearing impaired, for example. We must always strive to realize new features that might make people's lives easier — that's what we're working for."

It's illustrative how David refers to his doctorate in electrical engineering, obtained in May 2013: "What makes me happy is having been able to do it; it's not about being able to tell everyone that I'm now a 'doctor'". Indeed, apart from his co-workers, only his close friends are aware of his new title.

#### **A fife-playing doctor... in the USA**

As David says, smiling, "in the field of fifes and drums, being a doctor doesn't mean anything." Fife-and-drum bands — a particularly strong tradition in Upper Valais — are another of David's passions. Piper in the Raron-Sankt German company, he teaches the juniors and sits on the regional and federal technical committees. "Socially speaking, I'm very much from Valais. That's another reason why I'm so happy to be able to work at Idiap and practice my profession of researcher in my own country. When I'm at music rehearsals once every week, I take great pleasure from being with my friends and clearing my head."

This summer, David will take part in the Federal Fife and Drum Festival in Frauenfeld before heading to Deep River, near New York, along with around sixty other Swiss musicians including his wife Erika, also a fife player. The group is taking up an invitation from their American counterparts to give a series of concerts in the US. "I love my canton, but I also like to leave it from time to time; here the mountains can some-

times limit one's horizons a little. I'm always happy to get to know other people and other cultures, whether that's in a festive context or as part of my job, attending conferences, or following an internship, like the one I spent at the International Computer Science Institute (ICSI) in Berkeley, California, where I stayed for ten months."

If his work ever led him into exile, this *Haut-Valaisan* knows two things for sure: "It would be a difficult decision to make. And, ultimately, I'm sure I'd come back to Switzerland." By doing so, David would find once more the mountains in which he loves to hike: "To find myself at altitude, with a beautiful view, that gives me something of the same feeling I have when I solve a problem — joy and peace."



#### **ReMeeting — a meeting-facilitation tool**

David Imseng is among the winners of the International Create Challenge (ICC 2013 — See page 18), the second edition of which took place in autumn 2013. Baptized ReMeeting, his project not only makes it possible to identify people who speak at a meeting and determine when they do so, but also to summarize their comments, identify keywords used, and automatically record any new deadlines in the agendas of the meeting's different participants. The sum awarded by the ICC — CHF 10,000 — will enable him to continue his research and to set up a start-up to ensure the successful transfer of this technology to industry.



## MICHEL SALAMIN, FRENCH TEACHER

### "I TRAVEL THANKS TO MY STUDENTS!"

For the past twelve years, Michel Salamin has given French lessons to Idiap researchers — with an engagement and enthusiasm beyond the bounds of vocabulary and grammar.

"What excites me about my work here is the people." This is the secret of Michel's success; the reason for which, even after twelve years of collaboration with Idiap, he's still so enthusiastic. "My students come from all over the world. When they talk about their culture and their countries, it's like being there." It would be an understatement to say that this dynamic native of Martigny enjoys the time he spends each week at the institute. Indeed, when Martigny's Business and General Culture School, for which he teaches, offered him additional work, he declined the offer in order to continue working with the researchers at Idiap. "It was really the contact I have with these wonderful people that made up my mind to stay here. I admire the energy and the application they put into learning French, even though they don't use the French language in their work. It's a funny situation when you think about it: these researchers have extraordinary IQs and I teach them to count in French!"



#### **Brisolée, tobogganing, Mardi Gras, ...**

Among other delights, Michel's students have the *via ferrata Farinetta*, *brisolée*, the Valais Fair, moonlight tobogganing and fondues, and the Bourg carnival to look forward to. "It's nothing special. I just do with them what the average *Valaisan* does. I'm happy to show them that one can have a good life here, so close to nature." He admits that he sometimes reprimands his students if they talk about going to live in Montreux or Lausanne rather than staying in Valais.

Through the years and these shared activities, strong friendships have been forged with students, some of whom have come to know the region better than Michel himself. Job well done for the man who has taken on the mission of being a sort of "Mr. Integration".

#### **Favoring conversation skills**

With a Bachelor's degree and a *maître de gymnase* diploma that qualify him to teach French literature and the French and Italian languages, Michel adapts his courses to his audience by focusing on conversation skills. "Above all I try to convey simple, practical things that facilitate their daily lives." Have a short conversation, shop, be able to communicate with a real estate agent or a hairdresser. "I realized that my students feel really stressed if they can't explain how they want their hair cut."

The attention that Michel gives to his students' integration isn't limited to French lessons alone, or even to his role as a writing consultant, even if that proves indispensable when one of his students has to write a letter. Recognizing that language alone is not enough to ensure integration, he organizes numerous outings to which not only the researchers, but also their spouses, are invited.

#### **A real Tower of Babel**

With staff and students of over thirty nationalities and almost as many mother tongues, Idiap is a real Tower of Babel. The institute's researchers currently come from the following countries: Australia, Belgium, Brazil, China, Costa Rica, Spain, the USA, France, Great Britain, Greece, Hungary, India, Indonesia, Iran, Italy, Kenya, Lebanon, Lithuania, Macedonia, Mexico, Pakistan, Portugal, Slovakia, the Czech Republic, Romania, Russia, Switzerland, Turkey, and Vietnam.

## JOINING AND LEAVING

In 2013, nineteen members of staff left and fourteen joined the team. The new arrivals include two development engineers, one invited professor, four postdocs, six research assistants, and one industrial relation manager.

### JOINING US IN 2013

*First name, last name, position, country of origin*

|   |  |
|---|--|
| Sara Bahaadini Beigy, research assistant, Iran  | Renato Martins, development engineer, Portugal |
| Gulcan Can, research assistant, Turkey          | Srinivas Narayanan, invited professor, USA     |
| Samuel Gaist, development engineer, Switzerland | James Newling, research assistant, England     |
| Rui Hu, postdoc, China                          | Francisco Pinto, postdoc, Portugal             |
| Cijo Jose, research assistant, India            | Blaise Potard, postdoc, France                 |
| Srikanth Madikeri, postdoc, India               | Marzieh Razavi, research assistant, Iran       |
| Laurent Marti, industrial relation, Switzerland | Matthias Vanoni, research assistant, France    |

### MOVING ON IN 2013

*First name, last name, position, country of origin, joined Idiap, (new employer)*

Céline Aymon Fournier, public relations, Switzerland, 2004, Valais/Wallis Promotion  
John Dines, researcher, Australia, 2003  
Charles Dubout, research assistant, Switzerland, 2009  
Rémi Emonet, postdoc, France, 2010, University Jean Monnet, Saint-Etienne, France  
Arjan Gijsberts, postdoc, Netherlands, 2011  
Najeh Hajlaoui, postdoc, France, 2011  
Dinesh Babu Jayagopi, postdoc, India, 2007, International Institute of Information Technology (IIIT) Bangalore, India  
Samuel Kim, postdoc, South-Korea, 2011  
Riwal Lefort, postdoc, France, 2011, Groupe de Recherche en Informatique, Image, Automatique et Instrumentation de Caen (GREYC), France  
Adolfo Lopez Mendez, postdoc, Spain, 2012, start-up fezoo  
Laurent Marti, industrial relation, Switzerland, 2013  
Renato Martins, development engineer, Portugal, 2013, Koemei  
Edgar Roman Rangel, research assistant, Mexico, 2008, University of Geneva  
Dairazalia Sanchez-Cortes, research assistant, Mexico, 2009  
Nicolae Suditu, research assistant, Romania, 2008, SwissLitho AG  
Romain Tavenard, research assistant, France, 2011, University of Rennes 2  
Tatiana Tommasi, research assistant, Italy, 2008  
Alessandro Vinciarelli, senior research scientist, Italy, 1999  
Majid Yazdani, research assistant, Iran, 2008, University of Geneva



# DISTINCTIONS

## 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 2012 the Research Prize was awarded to **Laurent El Shafey** for his record of outstanding publications and for the excellence of his research into the topic "reproducible machine learning applied to face and speaker recognition". The Publication Prize went to **Leonidas Lefakis** for his outstanding scientific article entitled "Reservoir Boosting: Between Online and Offline Ensemble Learning".



*Laurent El Shafey*



*Leonidas Lefakis*

### External

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

#### **Vasil Khalidov, Florence Forbes, and Radu Horaud**

Best Paper Award Nomination at the IEEE Int. Workshop on Multimedia Signal Processing (MMSp) held in Pula (Sardinia, Italy), September/October 2013

***Alignment of Binocular-Binaural Data Using a Moving Audio-Visual Target***

#### **Alexander Heili and Jean-Marc Odobez**

Best Paper Award at the Workshop on Performance Evaluation of Tracking and Surveillance (PETS), January 2013

***Parameter Estimation and Contextual Adaptation for a Multi-Object Tracking CRF Model***

#### **C. Bhatt, Andrei Popescu-Belis, Maryam Habibi, Sandy Ingram, F. McInnes, S. Masneri, Nikolaos Pappas, and O. Schreer**

Winner of the Grand Challenge on Lecture Segmentation and Annotation at the ACM Multimedia 2013, Barcelona, October 2013

***Multi-factor Segmentation for Topic Visualization and Recommendation: the MUST-VIS System***

#### **C. Bhatt, Nikolaos Pappas, Maryam Habibi, and Andrei Popescu-Belis**

Winner of the hyperlinking subtask at the MediaEval 2013, Barcelona, October 2013

***Idiap at MediaEval 2013: Search and Hyperlinking Task***

## THESES COMPLETED

Nine students completed their thesis in 2013: Nicolae Suditu, Gelareh Mohammadi, Dairazalia Sanchez-Cortes, Tatiana Tommasi, Joan-Isaac Biel, Afsaneh Asaei, David Imseng, Charles Dubout et Majid Yazdani.

- ***Learning to Learn by Exploiting Prior Knowledge***  
**Tatiana Tommasi**, 2013  
Thesis directors: Prof. Hervé Bourlard, Dr. Barbara Caputo  
Members of the thesis committee: Prof. Dario Floreano, Prof. Jean-Philippe Thiran, Prof. Jim Little, Dr. Vittorio Ferrari
- ***Mining Conversational Social Video***  
**Joan-Isaac Biel**, 2013  
Thesis director: Dr Daniel Gatica-Perez  
Members of the thesis committee: Prof. Sabine Susstrunk, Prof. Pascal Frossard, Dr. Bernardo Huberman, Dr. Mathias Mehl
- ***Similarity Learning Over Large Collaborative Networks***  
**Majid Yazdani**, February 28, 2013  
Thesis directors: Prof. Hervé Bourlard, Dr. Andrei Popescu-Belis  
Members of the thesis committee: Prof. Karl Aberer, Dr. Samy Bengi, Dr. Jean-Cédric Chappelier, Prof. Andreas Krause
- ***Computational Methods for Audio-Visual Analysis of Emergent Leadership in Teams***  
**Dairazalia Sanchez-Cortes**, March 1, 2013  
Thesis director: Dr. Daniel Gatica-Perez  
Members of the thesis committee: Prof. Jean-Philippe Thiran, Prof. Pierre Dillenbourg, Prof. Fabio Pianesi, Prof. Marianne Schmid Mast
- ***Model-based Sparse Component Analysis for Multiparty Distant Speech Recognition***  
**Afsaneh Asaei**, March 21, 2013  
Thesis director: Prof. Hervé Bourlard  
Members of the thesis committee: Prof. Jean-Philippe Thiran, Prof. Christian Jutten, Prof. Dietrich Klakow, Prof. Pierre Vandergheynst
- ***Adaptive Relevance Feedback for Large-scale Image Retrieval***  
**Nicolae Suditu**, April 19, 2013  
Thesis director: Dr. François Fleuret  
Members of the thesis committee: Prof. M. Grossglauser, Dr. M. Ferecatu, Dr. S. Marchand-Maillet, Prof. J.-Ph. Thiran
- ***Automatic Personality Perception: Inferring Personality Traits from Nonverbal Vocal Behavior***  
**Gelareh Mohammadi**, June 17, 2013  
Thesis directors: Prof. Hervé Bourlard, Dr. Alessandro Vinciarelli  
Members of the thesis committee: Dr. Jean-Marc Vesin, Prof. Jean-Philippe Thiran, Dr. Fabio Pianesi, Prof. Björn Wolfgang Schuller
- ***Multilingual speech recognition: A posterior based approach***  
**David Imseng**, June 20, 2013  
Thesis director: Prof. Hervé Bourlard  
Members of the thesis committee: Prof. J.-Ph. Thiran, Prof. N. Morgan, Prof. T. Schultz, Dr. J.-M. Vesin
- ***Object Classification and Detection in High Dimensional Feature Space***  
**Charles Dubout**, December 17, 2013  
Thesis director: Dr François Fleuret  
Members of the thesis committee: Prof. Mark Pauly, Prof. Pascal Fua, Prof. Gilles Blanchard, Prof. Frédéric Jurie



# FINANCES





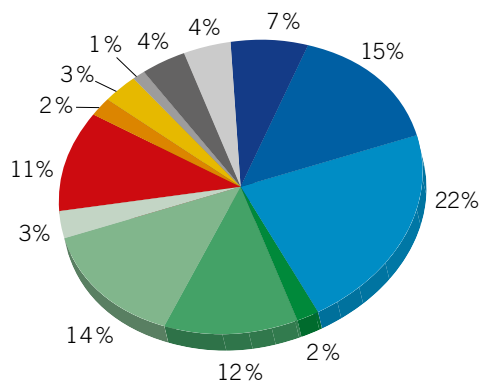
## OPERATING ACCOUNTS

| (Swiss francs)                             | 2012              | 2013              | %              |
|--|-------------------|-------------------|----------------|
| <b>INCOME</b>                              |                   |                   |                |
| City of Martigny                           | 650,000           | 700,000           | 6.58%          |
| Canton of Valais                           | 1,120,000         | 1,620,000         | 15.23%         |
| Swiss Confederation                        | 2,393,000         | 2,400,000         | 22.56%         |
| Loterie Romande                            | 250,000           | 250,000           | 2.35%          |
| NCCR IM2 projects                          | 764,071           | 1,273,628         | 11.97%         |
| Swiss National Science Foundation projects | 1,203,153         | 1,489,392         | 14.00%         |
| Hasler Foundation projects                 | 347,913           | 314,717           | 2.96%          |
| European Commission projects               | 2,219,526         | 1,132,694         | 10.65%         |
| The Ark projects                           | 55,628            | 189,288           | 1.78%          |
| CTI projects                               | 251,168           | 328,864           | 3.09%          |
| EPFL contribution                          | 72,000            | 72,000            | 0.68%          |
| Industrial Financing                       | 342,766           | 473,367           | 4.45%          |
| Other income / extraordinary income        | 669,449           | 394,213           | 3.71%          |
| <b>TOTAL INCOME</b>                        | <b>10,338,674</b> | <b>10,638,163</b> | <b>100.00%</b> |
| <b>EXPENSES</b>                            |                   |                   |                |
| Personnel expenses                         | 7,608,490         | 7,709,161         | 72.5%          |
| Education and travel                       | 528,042           | 458,200           | 4.3%           |
| Third-party expenses                       | 92,469            | 93,062            | 0.9%           |
| Computer equipment and maintenance         | 249,146           | 241,310           | 2.3%           |
| Administrative costs                       | 178,872           | 193,520           | 1.8%           |
| Promotion and communication                | 77,007            | 60,413            | 0.6%           |
| Rent                                       | 889,306           | 817,621           | 7.7%           |
| Depreciation                               | 281,777           | 184,563           | 1.7%           |
| Exceptional expenses                       | 421,340           | 97,881            | 0.9%           |
| Provisions                                 | —                 | 750,000           | 7.1%           |
| <b>TOTAL EXPENSES</b>                      | <b>10,326,449</b> | <b>10,605,731</b> | <b>99.7%</b>   |
| <b>OPERATING PROFIT / LOSS</b>             | <b>12,225</b>     | <b>32,432</b>     | <b>0.3%</b>    |



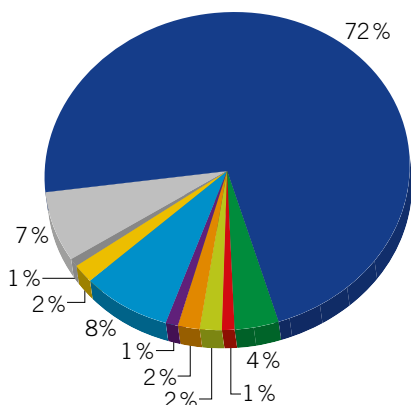
# SOURCES OF FINANCING / COSTS / COMMENTS

## Distribution of sources of financing



- City of Martigny
- Canton of Valais
- Swiss Confederation
- Loterie Romande
- NCCR IM2 projects
- Swiss National Science Foundation projects
- Hasler Foundation projects
- European Commission projects
- The Ark projects
- CTI projects
- EPFL contribution
- Industrial financing
- Other income / Extraordinary income

## Distribution of costs



- Personnel expenses
- Education and travel
- Third-party expenses
- Computer equipment and maintenance
- Administrative costs
- Promotion and communication
- Rent
- Depreciation
- Exceptional expenses
- Provisions

## Management comments

For the second consecutive year, the highly symbolic 10 million mark was exceeded. Idiap continues its momentum and closes its accounts with a profit of CHF 32,432.

The public share of funding remains within our strategic objective, at — for this year — 44.4%. The Canton of Valais and the city of Martigny reinforced their financial contributions by CHF 500,000 and CHF 50,000 respectively.

The volume of projects funded by the EU has decreased significantly. This is a direct consequence of the move from the union's 7<sup>th</sup> Framework Programme to H2020; a transition during which the institute has been faced with fewer calls for proposals and, therefore, fewer funding opportunities. The NCCR IM2 and our continuous development of relations with industry have partly offset this shortfall.

## Federal, cantonal, and municipal subsidies

(In thousands of Swiss francs)

| YEAR          | 2011  | 2012  | 2013  | 2014* |
|---------------|-------|-------|-------|-------|
| Confederation | 2,357 | 2,393 | 2,400 | 2,550 |
| Canton        | 900   | 1,120 | 1,620 | 1,720 |
| Municipality  | 650   | 650   | 700   | 700   |

\* Budget

# BALANCE SHEET

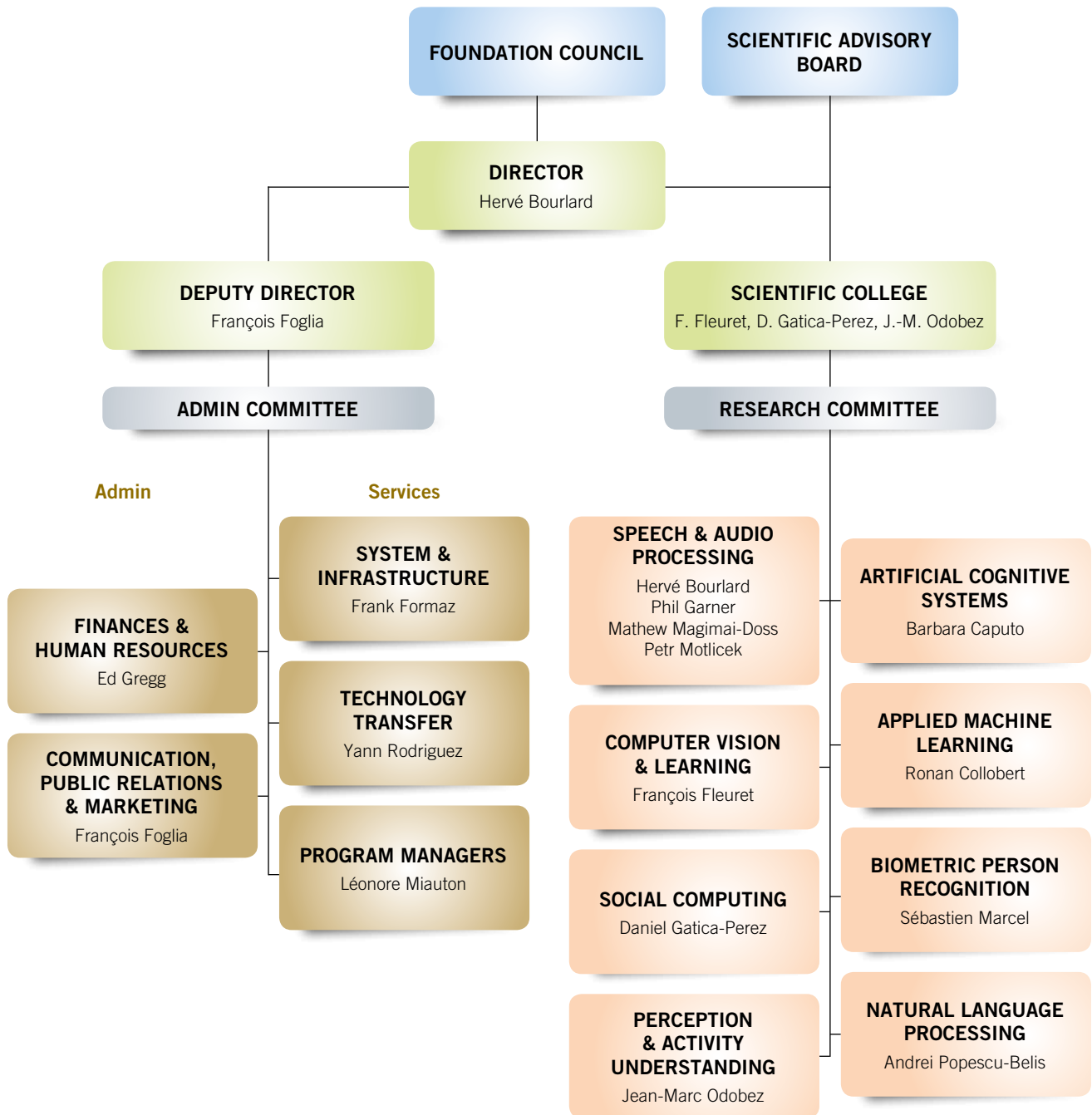
(Swiss francs)

|                                 | 31.12.2012          | 31.12.2013          |
|---------------------------------|---------------------|---------------------|
| <b>ASSETS</b>                   |                     |                     |
| Cash                            | 2,104,427.07        | 3,614,085.58        |
| Accounts receivable             | 287,382.35          | 220,492.98          |
| Accrued income and other        | 2,536,046.69        | 985,382.87          |
| <b>TOTAL CURRENT ASSETS</b>     | <b>4,927,856.11</b> | <b>4,819,961.43</b> |
| Equipment                       | 504,960.99          | 340,658.79          |
| Financial assets                | 10,000.00           | 10,000.00           |
| <b>TOTAL NON-CURRENT ASSETS</b> | <b>514,960.99</b>   | <b>350,658.79</b>   |
| <b>TOTAL ASSETS</b>             | <b>5,442,817.10</b> | <b>5,170,620.22</b> |
| <b>LIABILITIES</b>              |                     |                     |
| Accounts payable                | 365,838.69          | 307,725.05          |
| Accrued expenses                | 3,918,238.84        | 2,921,723.22        |
| Provisions                      | 0.00                | 750,000.00          |
| <b>TOTAL FOREIGN FUNDS</b>      | <b>4,284,077.53</b> | <b>3,979,448.27</b> |
| Share capital                   | 40,000.00           | 40,000.00           |
| Special reserve                 | 1,000,000.00        | 1,000,000.00        |
| Retained earnings               | 106,513.11          | 118,739.57          |
| Net income                      | 12,226.46           | 32,432.38           |
| <b>TOTAL OWN FUNDS</b>          | <b>1,158,739.57</b> | <b>1,191,171.95</b> |
| <b>TOTAL LIABILITIES</b>        | <b>5,442,817.10</b> | <b>5,170,620.22</b> |



# ORGANIZATION







# EMPLOYEES

## Scientists

*First name, last name, position, country of origin, (place of residence), joined Idiap*

Oya Aran Karakus, research fellow, Turkey, 2009  
Afsaneh Asaei, research assistant, Iran, 2008  
Sara Bahaadini Beigy, research assistant, Iran, 2013  
Chindansh Bhatt, postdoc, India, 2012  
Joan-Isaac Biel, research assistant, Spain, 2008  
Hervé Bourlard, director, Switzerland, Saxon, 1996  
Gulcan Can, research assistant, Turkey, 2013  
Olivier Canévet, research assistant, France, 2012  
Barbara Caputo, senior research scientist, Italy, 2005  
Ivana Chingovska, research assistant, Macedonia, 2011  
Ronan Collobert, research fellow, France, 2010  
Trinh-Minh-Tri Do, postdoc, Vietnam, 2009  
Elie El Khoury, postdoc, Lebanon, 2011  
Laurent El Shafey, research assistant, France, 2010  
Rémi Emonet, postdoc, France, 2010  
Nesli Erdogmus, postdoc, Turkey, 2012  
Marc Ferras Font, postdoc, Spain, 2011  
François Fleuret, senior research scientist, France, 2007  
Marco Fornoni, research assistant, Italy, 2010  
Kenneth Funes Mora, research assistant, Costa Rica, 2011  
Philip Garner, senior research scientist, England, 2007  
Daniel Gatica-Perez, senior research scientist, Mexico, 2002  
Paul Gay, research assistant, France, 2012  
Manuel Günther, postdoc, Germany, 2012  
Maryam Habibi, research assistant, Iran, 2011  
Alexandre Heili, research assistant, France, 2010  
Pierre-Edouard Honnet, research assistant, France, 2012  
Rui Hu, postdoc, China, 2013  
David Imseng, research assistant, Switzerland, Rarogne, 2009  
Cijo Jose, research assistant, India, 2013  
Dinesh Babu Jayagopi, postdoc, India, 2007  
Ilja Kuzborskij, research assistant, Lithuania, 2012  
Alexandros Lazaridis, postdoc, Greece, 2012  
Rémi Lebret, research assistant, France, 2012  
Leonidas Lefakis, research assistant, Greece, 2010  
Joël Legrand, research assistant, France, 2012  
Srikanth Madikeri, postdoc, India, 2013  
Mathew Magimai Doss, research scientist, India, 2007  
Sébastien Marcel, senior research scientist, France, 2000  
Thomas Meyer, research assistant, Switzerland, Martigny, 2010  
Gelareh Mohammadi, research assistant, Iran, 2009  
Petr Motlicek, research scientist, Czech Republic, 2005  
Srinivas Narayanan, invited professor, USA, 2013  
James Newling, research assistant, England, 2013  
Laurent Nguyen, research assistant, Switzerland, Lausanne, 2011  
Jean-Marc Odobez, senior research scientist, 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  
Francisco Pinto, postdoc, Portugal, 2013  
Andrei Popescu-Belis, senior research scientist, France/Romania, Switzerland, 2007  
Blaise Potard, postdoc, France, 2013  
André Rabello Dos Anjos, postdoc, Brazil, 2010  
Ramya Rasipuram, research assistant, India, 2010  
Marzieh Razavi, research assistant, Iran, 2013  
Lakshmi Saheer, research assistant, India, 2008  
Darshan Santani, research assistant, India, 2012  
Ashtosh Sapru, research assistant, India, 2011  
Samira Sheikhi, research assistant, Iran, 2010  
Nicolae Suditu, research assistant, Romania, 2008  
Gyorgy Szaszak, postdoc, Hungary, 2012  
Mohammad Javad Taghizadeh, research assistant, Iran, 2011  
Raphaël Ullmann, research assistant, Switzerland, Lausanne, 2012  
Matthias Vanoni, research assistant, France, 2013  
Sree Harsha Yella, research assistant, India, 2010

## Development engineers

Philip Abbet, senior dev. engineer, Switzerland, Conthey, 2006  
Olivier Bornet, senior dev. engineer, Switzerland, Nendaz, 2004  
Milos Cernak, senior dev. engineer, Slovakia, 2011  
Samuel Gaist, dev. engineer, Switzerland, Ardon, 2013  
Salim Kayal, dev. engineer, Switzerland, Vevey, 2011  
Vasil Khalidov, dev. engineer, Russia, 2010  
Christine Marcel, dev. engineer, France, 2007  
Florent Monay, dev. engineer, Switzerland, Choëx, 2008  
François Moulin, dev. engineer, Switzerland, Vollèges, 2009  
Alexandre Nanchen, dev. engineer, Switzerland, Martigny, 2008  
Flavio Tarsetti, dev. engineer, Switzerland, Martigny, 2008

## Administrative staff

*First name, last name, position, country of origin, (place of residence), joined Idiap*

Céline Aymon Fournier, public relations, Switzerland, Fully, 2004  
Antoine Dorsaz, financial assistant, Switzerland, Fully, 2012  
Christophe Ecoeur, program manager, Switzerland, Collombey, 2010  
Martina Fellay, junior program manager, Austria, 2012  
François Foglia, deputy director, Switzerland, Saxon, 2006  
Edward-Lee Gregg, financial manager, United States, 2004  
Léonore Miauton, senior program manager, Switzerland, Chexbres, 2012  
Sylvie Millius, administrative assistant, Switzerland, Vétroz, 1996  
Yann Rodriguez, industrial relations, Switzerland, Martigny, 2006  
Nadine Rousseau, administrative assistant, Switzerland, Saxon, 1998

## System engineers

Bastien Crettol, system administrator, Switzerland, Sion, 2005  
Norbert Crettol, system administrator, Switzerland, Martigny, 2002  
Cédric Dufour, system administrator, Switzerland, Verbier, 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.

Tauseef Ali, Pakistan, University of Twente (NLD)  
Krysten Ansermoz, Switzerland, EPFL, Lausanne  
Afroze Ibrahim Baqapuri, Pakistan, EPFL, Lausanne  
Julia Biolley-Franzetti, Switzerland, University of Lausanne  
Maxime Bourlard, Switzerland, University of Neuchâtel  
Aleksandra Cerekovic, Croatia, University of Zagreb (HRV)  
Simon Coquoz, Switzerland, EPFL, Lausanne  
Pranay Dighe, India, Indian Institute of Technology, Kanpu (India)  
Matthieu Duval, France, Polytech Nice-Sophia, Sophia Antipolis (FRA)  
Alexandre Ferchaud, Switzerland, Ecole de commerce, Sion  
Manon Fournier, Switzerland, Lycée-Collège de l'Abbaye, St-Maurice  
Javier Galbally, Spain, Autonomous University of Madrid (ESP)  
Hainan Hu, China, EPFL, Lausanne  
Nicolas Jacquemin, Switzerland, EPFL, Lausanne  
César Laurent, Switzerland, EPFL, Lausanne

Sharid Loaiciga Sanchez, Costa Rica, University of Geneva  
Alvaro Marcos, Spain, University of Alcalá (ESP)  
Rahesh Mehta, India, Tampere University of Technology (FIN)  
Braid Meyer, Switzerland, ISME, St. Gallen  
Xingyu Na, China, Beijing Institute of Technology, Beijing (CHN)  
Billy Okal, Kenya, Jacobs University, Bremen (DEU)  
Naoki Peter, Switzerland, EPFL, Lausanne  
Jérémie Rappaz, Switzerland, EPFL, Lausanne  
Sandrine Revaz, Switzerland, University of Fribourg  
Theresa Schwery, Switzerland, EPFL, Lausanne  
Abhishek Sharma, India, CNRS Paris (FRA)  
Kelly Tiraboschi, Italy, Lycée-Collège de l'Abbaye, St-Maurice  
Sarah Anna Vomsattel, Switzerland, University of Lausanne

## Visitors

*First name, last name, country of origin, 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.

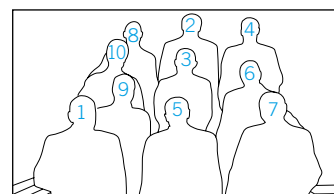
Abhijit Zipru Bendale, India, University of Colorado (USA)  
Abhishek Dutta, Nepal, University of Twente (NLD)  
Adrian Penate Sanchez, Spain, Polytechnic University of Catalonia, Barcelona (ESP)



# FOUNDATION COUNCIL 2013



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 — ensures Idiap's development and defends its interests.



- 1 **Olivier Dumas**, President  
Former President of the city of Martigny
  
- Jean-Daniel Antille**, Vice President (*not pictured*)  
Manager of "Antenne Régions Valais romand"
  
- Prof. Philippe Gillet**, Vice President (*not pictured*)  
Vice President for academic affairs  
Ecole polytechnique fédérale de Lausanne (EPFL)  
*Represented by:*
- 2 **Prof. Karl Aberer**  
Vice President for Information Systems  
Ecole polytechnique fédérale de Lausanne (EPFL)
- 3 **Marc-André Berclaz**  
President of the Board of Directors of HES-SO Valais Wallis
- 4 **M. Stefan Bumann**  
Head of Higher Education, State of Valais
  
- Adrienne Corboud Fumagalli** (*not pictured*)  
Vice President for industrial relations  
Ecole polytechnique fédérale de Lausanne (EPFL)
  
- Marc-Henri Favre** (*not pictured*)  
President of the city of Martigny
  
- 5 **Prof. Christian Pellegrini**  
Senior Member of the Foundation Council of Idiap  
Honorary Professor in the Faculty of Sciences  
at the University of Geneva

- 6 **Jean-René Germanier**  
Member of the Swiss parliament
- 7 **Patrick Furrer**  
Vice Director of Euresearch
- 8 **Jordi Montserrat**  
Regional Director of venturelab and venture kick
  
- Dominique Perruchoud** (*not pictured*)  
President of the Board of Cimark SA
  
- Walter Steinlin** (*not pictured*)  
Swisscom University Relations  
President of the Commission for Technology  
and Innovation (CTI)

## Not in the Foundation Council

- 9 **Prof. Hervé Bourlard**  
Director of Idiap
- 10 **François Foglia**  
Deputy Director of Idiap



## 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 counsel 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. Anil K. Jain**

Distinguished Professor, Department of Computer Science  
& Engineering  
Michigan State University, USA

**Dr. John Makhoul**

Chief Scientist, Speech and Signal Processing, BBN Technologies  
Cambridge, MA, USA

**Prof. Kenji Mase**

Professor, Graduate School of Information Science  
Nagoya University, Japan

**Prof. Nelson Morgan**

Deputy Director (and former Director) of the International Computer  
Science Institute (ICSI)  
Berkeley, CA, USA

**Prof. Klaus-Robert Müller**

Professor for Computer Science, TU Berlin  
Director, Bernstein Focus on Neurotechnology,  
Berlin, Germany

**Dr. David Nahamoo**

Senior Manager, Human Language Technologies, IBM Research  
Yorktown Heights, NY, USA

**Prof. Gerhard Sagerer**

Rector, University of Bielefeld  
Germany

**Prof. Bernt Schiele**

Max-Planck-Director, MPI Informatics  
Professor at Saarland University,  
Saarbrücken, Germany

**Prof. Bayya Yegnanarayana**

Professor and Microsoft Chair, International Institute  
of Information Technology (IIIT)  
Hyderabad, India



# 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.com](http://www.swisscom.com)



**Swiss Power Group.**

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



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



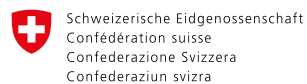
[www.theark.ch](http://www.theark.ch)



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



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



Innovation Promotion Agency CTI

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

**HASLERSTIFTUNG**

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

S C I E N T I F I C   I N S E R T S

# IDIAP RESEARCH AREAS: HUMAN AND MEDIA COMPUTING

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 web site [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 a few keywords.

| 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: 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. After the end of the MOBIO EU project, this leadership was confirmed in 2011 by the granting of new large EU projects, like the ICT project TABULA RASA and the SECURITY project BEAT. Based on this expertise, Idiap is currently initiating a “Swiss Center for Biometric Security Research and Testing”.

- **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. Real-life applications include end-to-end spoken-term detection (see the Deep-STD project), Natural Language Processing with a particular interest in semantic analysis (Sem-Tex project, in collaboration with NEC Research, USA) and large-scale distributed web search (SODS).



## 1. Speech and Audio Processing

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 group has certain core technology, built around the connectionist approach to automatic speech recognition (ASR), and continues to contribute in that area. However, the expertise of the group has diverged to encompass hidden Markov model technology, text to speech synthesis (TTS) and generic audio processing.

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 database has been released<sup>3</sup>; other contributions such as ISS and SSP are on the GitHub site<sup>4</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 that we are well placed to capitalise on given our geographical location. The group is also very active in contributing to KALDI – open-source ASR toolkit<sup>5</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 (funded through EC Eureka project D-Box<sup>6</sup>), Idiap is actively carrying research in several ASR directions, including:

- **Robust parametrization and acoustic modeling**  
We are still investigating new features (e.g., posterior-based features) and new acoustic models (new forms of hidden Markov models, such as KL-HMM, or artificial neural networks) that are more robust to noise and acoustic environments, as well as to speaker variability (e.g., accented speech). Over the last 12 months, we mainly focused on several novel acoustic modeling approaches that can ef-

fectively exploit acoustic and linguistic resources of a different language or multiple languages (funded through SNSF projects like ICS-2010<sup>7</sup> and FlexASR<sup>8</sup>).

- **Cross-lingual and multi-lingual speech recognition**  
In addition to the above, and in the context of an Industry sponsored project (Samsung, Inc. South Korea<sup>9</sup>), we are also investigating new fast acoustic model adaptation techniques (currently referred to as “subspace Gaussian modelling”) in cross-lingual and multi-lingual scenarios.
- **Swiss-languages**  
As part of a project with the Valais (and probably Vaud) Parliament (MediaParl project), we are developing state-of-the-art ASR systems for Swiss languages. We have made very good progress towards Swiss German and Swiss French speech recognition. In particular, recent advances in speech technology allow training on both languages simultaneously, improving the recognition rate for both.
- **Grapheme-based ASR**  
State-of-the-art ASR systems typically use phones as the standard subword units. However, phones are not always well defined, not all languages have well developed phonetic lexicon, and phonetic transcription of the lexicon words do not always match the actual pronunciation (especially in the case of accented speech). We are thus investigating the use of alternate subword unit representations such as graphemes (funded through the SNSF project FlexASR and HASLER Foundation project AddG2SU<sup>10</sup>).
- **Template-based ASR**  
Given the availability of very large databases, and the need to relax some of the HMM statistical assumptions (e.g., to better model the temporal dynamics of the speech signal), there is renewed interest in “template-based ASR” (also referred to as “episodic model”), replacing the training of stochastic models by simply storing full speech templates and performing recognition through large scale dynamic programming (dynamic time warping). In the context of projects like SNSF funded A-MUSE<sup>11</sup> and CTI funded

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

2 <http://juicer.amiproject.org/juicer/>

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

4 <https://github.com/idiap>

5 <http://kaldi.sourceforge.net>

6 [www.idiap.ch/scientific-research/projects/d-box-a-generic-dialog-box-for-multilingual-conversational-applications](http://www.idiap.ch/scientific-research/projects/d-box-a-generic-dialog-box-for-multilingual-conversational-applications)

7 [www.idiap.ch/scientific-research/projects/interactive-cognitive-systems](http://www.idiap.ch/scientific-research/projects/interactive-cognitive-systems)

8 [www.idiap.ch/scientific-research/projects/flexible-grapheme-based-automatic-speech-recognition](http://www.idiap.ch/scientific-research/projects/flexible-grapheme-based-automatic-speech-recognition)

9 [www.idiap.ch/scientific-research/projects/domain-adaptation-using-subspace-models](http://www.idiap.ch/scientific-research/projects/domain-adaptation-using-subspace-models)

10 [www.idiap.ch/scientific-research/projects/flexible-acoustic-data-driven-grapheme-to-subword-unit-conversion](http://www.idiap.ch/scientific-research/projects/flexible-acoustic-data-driven-grapheme-to-subword-unit-conversion)

11 [www.idiap.ch/scientific-research/projects/adaptive-multilingual-ASR-TTS](http://www.idiap.ch/scientific-research/projects/adaptive-multilingual-ASR-TTS)



MultiVEO<sup>12</sup>, we have recently shown that using posterior-based features and the appropriate local distance (based on KL-divergence) was yielding similar or better performance than full fledged HMM systems.

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 in continuous speech recognition, our ASR research also focuses on the development and improvement of state-of-the-art keyword spotting systems for audio mining. Keyword spotting (spoken term detection) is also addressed in the Applied machine Learning group, as discussed in Section 6, page XII.

### 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>13</sup> (Spoken Interaction With Interpretation in Switzerland) and RECOD<sup>14</sup>, on low bit rate speech coding (see below).

The TTS work at Idiap is largely aligned with the trend in the field towards statistical synthesis, which uses the same technology as ASR, placing us on a firm technical footing. Initial work was done under the FP7 EMIME<sup>15</sup> project, a project in speech to speech translation where the synthetic translated voices were adapted cross language to sound like the original speakers. This theme is continued in SIWIS which aims to do the same in Swiss languages, including the concept of accents and (to some extent) dialects. Another 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.

SIWIS and EMIME 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.

TTS is also pushed towards light-versions to be integrated into open-gaming platform through the EC Eureka project D-BOX<sup>16</sup>

12 [www.idiap.ch/scientific-research/projects/high-accuracy-speaker-independent-multilingual-automatic-speech-recognition-system](http://www.idiap.ch/scientific-research/projects/high-accuracy-speaker-independent-multilingual-automatic-speech-recognition-system)

13 [www.idiap.ch/project/siwis/](http://www.idiap.ch/project/siwis/)

14 [www.idiap.ch/scientific-research/projects/low-bit-rate-speech-coding](http://www.idiap.ch/scientific-research/projects/low-bit-rate-speech-coding)

15 [www.emime.org/](http://www.emime.org/)

16 [www.idiap.ch/scientific-research/projects/d-box-a-generic-dialog-box-for-multilingual-conversational-applications](http://www.idiap.ch/scientific-research/projects/d-box-a-generic-dialog-box-for-multilingual-conversational-applications)

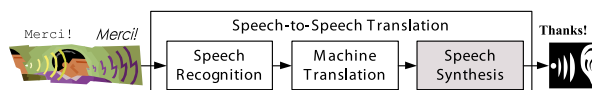


Figure 2: Typical speech-to-speech translation system architecture developed at Idiap.

- 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. Under the RECOD project funded by armasuisse<sup>17</sup>, the technology lends itself to clear and secure communication. Innovative aspects draw on those of SIWIS: the system will be multi-lingual, and will model prosody to give acceptable latency; i.e., the synthetic speech will be produced as the speaker is talking word by word rather than sentence by sentence.

### Audio Processing

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

- Sparse structured representations

Idiap continues in 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 far-field microphones.

17 [www.ar.admin.ch/internet/armasuisse/en/home.html](http://www.ar.admin.ch/internet/armasuisse/en/home.html)



- **Microphone arrays**

Idiap further performs studies on audio recorded from microphone arrays, especially those organized in “ad-hoc” geometries. 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.

- **Speaker diarization**

In speaker diarization — the task of annotating temporal regions of audio recordings with labels indicating “who spoke when” — Idiap focuses on processing of spontaneous overlapping speech through the use of multiple sources of information investigated in the context of various projects, including the EU project inEvent<sup>18</sup> and the SNSF project RODI<sup>19</sup>.

- **Audio indexing**

In the context of the Hasler Foundation project SESAME<sup>20</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.

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

Over the last two 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<sup>21</sup>. The technique exploits the opportunity of having technology in both ASR and TTS, allowing information to be transmitted at the word or phoneme level rather than acoustic level, whilst still being perceived as speech.

- **Perceptual background noise analysis for telecommunication systems**

Today, communication noises in mobile networks are often suppressed to improve speech intelligibility. But newer speech network generations can now transmit much higher quality signals, hence the potential to better exploit the information about a talker’s environmental context. New international ITU standards will soon be defined along that

trend. In the CTI project PANDA<sup>22</sup>, and in collaboration with SwissQual AG, we aim at exploiting leading edge speech processing technique to automatically extract and perceptually classify environmental noises towards those new standards.

## 2. Computer Vision and Learning

The main objective of the Computer Vision and Learning group is the development of novel machine-learning techniques, with a strong emphasis on their algorithmic efficiency, and applied mostly to the processing of images and video streams.

Our 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.

### *Generic Learning*

- **Joint feature selection**

When dealing with very large feature space, it is often desirable to apply a pre-processing step during learning which selects a “good” sub-set of features. Many criteria have been developed for that task, but while it is easy to measure how good a single feature is, they often fail at taking into account the joint quality of a set of features.

We have developed new methods which relies on a Gaussian model of the 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.

- **Efficient sample sampling and Reservoir learning**

To improve the extraction and the use of samples in the context of large scale learning, two issues are at hand: The extraction of novel samples, usually incurring a cost linear with the number of samples, and the processing of the said samples, also of linear cost.

The SNSF DASH project addresses the former. It can be seen as a form of unsupervised active learning to identify good “sources” of samples such as web sites, data-bases, or videos according to their ability to improve a predictor, and focus the extraction of additional training data from there, leading to a sub-linear cost.

<sup>18</sup> [www.inevent-project.eu/](http://www.inevent-project.eu/)

<sup>19</sup> [www.idiap.ch/scientific-research/projects/role-based-speaker-diarization](http://www.idiap.ch/scientific-research/projects/role-based-speaker-diarization)

<sup>20</sup> [www.idiap.ch/scientific-research/projects/searching-swiss-audio-memories](http://www.idiap.ch/scientific-research/projects/searching-swiss-audio-memories)

<sup>21</sup> [www.idiap.ch/scientific-research/projects/low-bit-rate-speech-coding](http://www.idiap.ch/scientific-research/projects/low-bit-rate-speech-coding)

<sup>22</sup> [www.idiap.ch/scientific-research/projects/perceptual-background-noise-analysis-for-the-newest-generation-of-telecommunication-systems-1](http://www.idiap.ch/scientific-research/projects/perceptual-background-noise-analysis-for-the-newest-generation-of-telecommunication-systems-1)





- **Learning in large dimension feature space**

In the context of the MASH project<sup>23</sup> we have designed new Boosting procedures to learn with hundreds of families of thousands or millions of features, by either sampling a few features from each family before the learning starts, and to use this features to estimate at every Boosting step the most promising feature family, or by optimizing on-the-fly the number of samples and features to look at every boosting step.

Our current work addresses the design of artificial neural networks in that context. We investigate novel algorithmic organizations of the computation coupled with sparse learning to reduce the cost for their evaluation.

- **High-dimension density similarity measures**

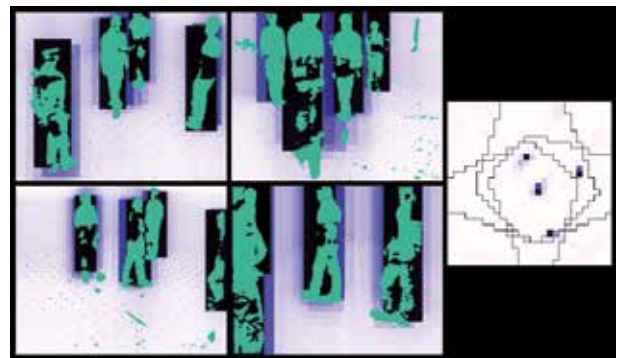
We study the clustering and classification of distributions of points in high-dimension spaces. To leverage the remarkable ability of random forests to cope with such situations, we have proposed a novel metric, which looks at the Entropy of the empirical distribution of a group of samples in the leaves of a forest build for another group of points.

If the two groups follow the same distribution, this entropy should be similar to that of the initial group, hence maximum. We applied this metric to the study of the growth dynamic of neurons, and its modulation by their genotypes in the framework of a collaboration with the University of Geneva, the University of Basel, and EPFL.

### **Scene understanding**

- **Fast object detection**

In the SNSF project VELASH, we have investigated the use of very large feature spaces for image analysis, and in particular the speed-up of state-of-the-art object detectors based on the use of linear operators at multiple scales and locations. To speed-up this process, we proposed a careful organization of the computation to allow the use of the classical Fast Fourier Transform. Our implementation of the image parsing is close to one order of magnitude faster than the best pre-existing algorithms, and available under the open-source GPL3 license.



*Figure 3: Our fast version of the Deformable Part Model object detector (left) is close to one order of magnitude faster than other existing implementations. The Probabilistic Occupancy Map (right) estimates probabilities of presence on the ground, given the result of a background subtraction procedure in individual views.*

We have also developed a Boosting variant to train jointly a cascade of classifiers for detection. This procedure relies on a stochastic interpretation of the classifier responses. The resulting algorithm pushes all the classifiers to respond properly on the positive samples, and pushes the classifiers which are “already good” to get even better on each negative samples.

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

Related to our work on the stationary features, we have started a project in collaboration with the Universitat Politècnica de Catalunya, whose underlying philosophy is to leverage high-quality multi-million pixels training image, to build models able to cope with extremely degraded test images of a few hundreds pixels. 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 we want to check, and train a predictor from these measurements.

<sup>23</sup> <http://mash-project.eu>



- **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 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.

### 3. Social Computing

Social computing is an interdisciplinary domain that integrates theory and models from mobile and ubiquitous computing, multimedia, machine learning, and social sciences to analyze and interpret human and social behavior in everyday life, and to create devices and systems that support interaction and communication. The Social Computing group in 2013 was composed of 1 head of group, 3 postdoctoral researchers, 1 SNSF Ambizione Research Fellow, 4 PhD students, and 2 visiting researchers (1 postdoc and 1 PhD student). Research lines that were investigated in 2013 included analysis of ubiquitous face-to-face interaction, behavioral analysis of social video, and urban data mining from smartphone and location-based social network data.

#### **Face-to-face interaction**

Collaborating with social psychologists, we are studying 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. In the dyadic case, in the context of organizational behavior situations, we have investigated connections between hireability, personality, and performance and nonverbal cues automatically measured from audio and video streams.



*Figure 4: Typical scenario of the NISHA project, an NTT-Idiap Social Behavior Analysis Initiative funded by NTT Communication Science Laboratories, Japan.*

In the small-group case, we are studying links between behavioral cues and variables like emergent leadership and personality traits using transfer learning approaches. This research was conducted in the framework of two SNSF projects (SONVB<sup>24</sup>: Sensing and Analyzing Organizational Behavior, Sinergia Interdisciplinary Program; and SOBE<sup>25</sup>, an Ambizione Research Fellowship) and the NISHA<sup>26</sup> project (NTT-Idiap Social Behavior Analysis Initiative) funded by NTT Communication Science Laboratories, Japan. Our research in ubiquitous face-to-face interaction was featured in Swiss National Television<sup>27</sup>.

#### **Social video**

The work in this line developed VlogSense, an approach to characterize video bloggers (vloggers) in social media sites like YouTube through the automatic extraction of behavioral features including prosody, body motion, and facial expressions, which enables high-level social analysis (Figure 5). We have examined the problem of interpersonal perception in vlogging, i.e., how vloggers are perceived by their online audiences. We have shown several significant associations between automatically extracted nonverbal and verbal cues and first impressions of personality traits, attractiveness, and mood. We are using video crowdsourcing with platforms like Amazon Mechanical Turk to collect interpersonal judgments; this is a powerful technique that scales up to large amounts of data while relying on demographically diverse viewers. This research was funded by the SNSF National Center of Competence in Research on Interactive Multimodal Information Management (NCCR IM2) through the VlogSense project (Modeling Conversational Be-

<sup>24</sup> [www.idiap.ch/project/sonvb](http://www.idiap.ch/project/sonvb)

<sup>25</sup> [www.idiap.ch/scientific-research/projects/multimodal-computational-modeling-of-nonverbal-social-behavior-in-face-to-face-interaction-1](http://www.idiap.ch/scientific-research/projects/multimodal-computational-modeling-of-nonverbal-social-behavior-in-face-to-face-interaction-1)

<sup>26</sup> [www.idiap.ch/project/nisha/](http://www.idiap.ch/project/nisha/)

<sup>27</sup> <http://la1.rsi.ch/home/networks/la1/cultura/ll-giardino-di-Albert/2013/05/28/gesto-rivelatore.html#Video>



havior in Social Media). As a follow-up, a PhD student graduate in the group received an EPFL Innogrant Award to support the creation of a social video start-up during 2014.

### Urban data mining

Our work in this domain spans two research lines: large-scale smartphone sensing and mobile social network data analysis (Figure 5). For the first line, after co-leading the successful Nokia Mobile Data Challenge 2012 initiative (which attracted participation from several hundred researchers around the world to develop new ideas in mobile computing using the Lausanne smartphone dataset – 180 volunteers in French-speaking Switzerland contributing everyday-life sensor and log data over one year), we undertook the roles of MDC Chair and exclusive worldwide distributor of the MDC data. During 2013, several dozen institutions requested access to this unique research resource<sup>28</sup>.

Regarding mobile social networks, supported by the SNSF HAI<sup>29</sup> project (Human Activity and Interactivity), we are investigating the use of Twitter and FourSquare data to characterize phenomena at city scale, including characterization of urban regions according to language usage. This work received press coverage at the national level<sup>30</sup>. We also initiated and are currently co-leading the MIRROR multidisciplinary initiative at EPFL (Interrogating Big Data to Improve Decision Making in African Cities), which involves several EPFL laboratories with the goal of developing approaches to characterize and understand large-scale phenomena in African cities.

## 4. Perception and Activity Understanding

The Perception and Activity Understanding group conducts research in human activities analysis from multi-modal data. This includes the development of principled and efficient algorithms for the characterization of people states from multiple data sources (mainly vision and audio) and the interpretation of sequential data in different scenarios. During 2013, the group was composed of five post-doctoral members, 4 PhD students, and one master student. The year was marked by ending projects (EU VANAHEIM and HUMAVIPS projects, SNSF Promovar) with increased efforts towards the building of real-time demonstrators, and the finalization of journal publications (6 journals published in 2013). Some of the main research carried out this year are described below.

<sup>28</sup> [www.idiap.ch/project/mdc/](http://www.idiap.ch/project/mdc/)

<sup>29</sup> [www.idiap.ch/scientific-research/projects/human-activity-and-interactivity-modeling](http://www.idiap.ch/scientific-research/projects/human-activity-and-interactivity-modeling)

<sup>30</sup> [www.computerworld.ch/news/kommunikation/artikel/social-media-schweiz-schreibt-englisch-63967/](http://www.computerworld.ch/news/kommunikation/artikel/social-media-schweiz-schreibt-englisch-63967/)

### Tracking in surveillance scenarios

In the context of the VANAHEIM project, we have continued our investigation towards enhanced tracking and behavior recognition for surveillance applications. In particular, several improvements were made within our Conditional-Random-Field (CRF) tracking-by-detection framework. This include the exploitation of image-based motion cues to improve the association of human detections and keep the optimization at the detection level rather than relying on the error prone generation of tracklets through frame-to-frame detection association to build trajectories; the derivation of multi-cue association reliability factors; and further refinements of the unsupervised learning of association parameter distribution parameters. State-of-the-art results were obtained on several benchmark data of the field. The work received the best paper award at the annual IEEE Performance Evaluation of Tracking and Surveillance (PETS) workshop.

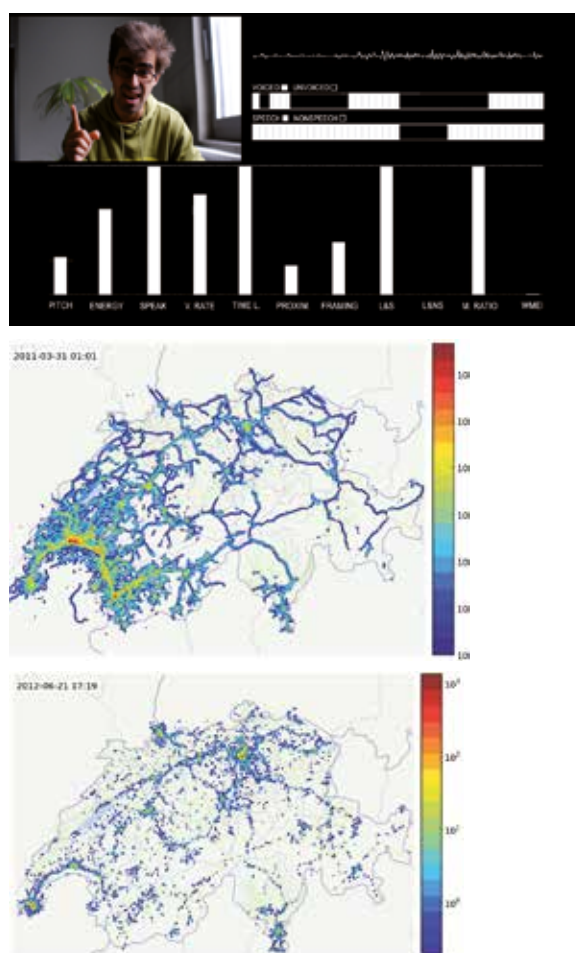


Figure 5: Top: The VlogSense concept: Online vloggers are characterized by a variety of nonverbal behavioral features. Center: Heatmap of the Lausanne smartphone location data in Switzerland. Bottom: Heatmap of check-ins in Switzerland using six months of Foursquare data.



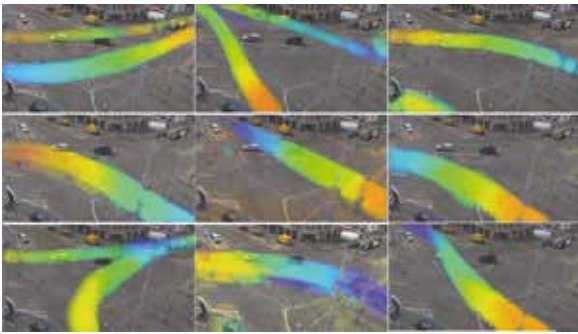


Figure 6: Top: the Idiap human sensing system working on the Nao robot. Bottom : automatic activity discovery from video logs. Each image displays an extracted activity motif that correspond to a sequence of observations that regularly co-occur over time. The color gradient indicates the timing since the start of the motif (from blue to red).



Figure 7: Top: Gaze tracking rendering during a job interview interaction. Bottom: Setup used for the EYEDIAP gaze dataset, the first public database for eyetracker evaluation.

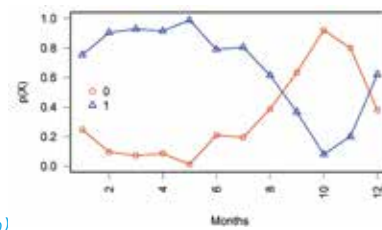
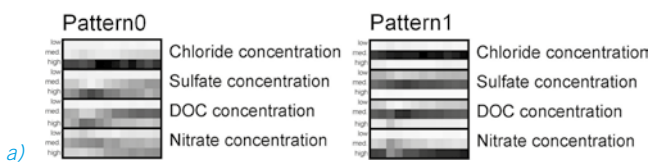


Figure 8: Water quality analysis during and after rainfall events using our topic models. The water quality is characterized by a set of solute-concentration parameters quantified in three different classes: low, medium, high. Water quality time series are decomposed as a mixture of  $K$  topic (temporal patterns) represented by the probabilities of observing solute concentration levels a given day (from 2 days before to 1 after the rainfall event). (a) Temporal patterns discovered when  $K=2$  (black means high probability). (b) The monthly pattern distribution (averaged over 12 years, month 1 = september). For instance, it resorts that the 0th pattern is typical of the summer season

### Non-verbal behaviour analysis in Human-Robot interaction scenarios

Our long-standing work on human interaction modelling has been continued in the context of the European HUMAVIPS project that seeks to endow humanoid robots with basic social skills necessary to deal with small groups of people, and where Idiap developed different sensing components (see left image of Fig. 6). Building upon our multiperson realtime head tracking and pose estimation algorithm, we have improved the recognition of the Visual focus of attention (VFOA) of people by using different contextual models exploiting the robot dialog state (what he is speaking about, who he is addressing) as a strategy for disambiguating amongst looking at different VFOA alternatives in a given situation. Furthermore, we have investigated how context could increase the recognition of addressee (to whom a person is speaking, in particular in HRI, is a person speaking to the robot) whether directly or indirectly through improving VFOA. In particular, we showed that considering only potential addressee targets within the VFOA recognition module rather than considering all VFOA targets (eg. including paintings) was producing much better results.

### Multi-modal 3D Gaze tracking using depth sensor data

In the TRACOME project funded by SNSF, we continued our investigation towards the extraction of non-verbal behavior cues and more specifically gaze from cheap RGB-D (depth) cameras like Kinect. Building on our head-pose free gaze estimation framework, we have proposed extensions that accommodate unknown users and can couple the gaze estimation from the two eyes. In a second thread, we have exploited the system and demonstrated its effectiveness at estimating precise mutual gaze events between two persons involved in natural interactions (job interviews) (see Fig. 7). Finally, given the lack of common benchmarks for the evaluation of eye gaze tracking systems, we have recorded a significant large dataset that covers most difficulties of gaze estimation (different persons, sessions and sensing conditions towards 2D -screen- or 3D targets). The dataset is made public and will be presented at the Eye Tracking and Research for Applications conference.



## **Activity analysis**

*Motif discovery and stream selection.* We have continued our work on the design of mining algorithms for the unsupervised discovery of recurrent activity pattern in multivariate time series, where observed values are caused by the superposition of multiple phenomena that can occur concurrently and with no synchronization. The method can be applied to typical situations where one or multiple sensor record the activities of multiple objects/people, like multimodal sensors (proximity, water, light sensors, etc.) in home automation (domotics) applications (see Fig. 8 right for an example in video).

In particular, in the EU Vanaheim project, these models were used this year to perform abnormality detection. More specifically, in collaboration with ethologists from the University of Vienna, we have investigated the selection of video streams to be displayed in control rooms of large public spaces and direct the attention of operators towards interesting information. The rankings of video streams according to the unusualness or interestingness of their content made by real human operators were shown to be in good agreement with those resulting from the automatic abnormality rating methods. A real-time demonstrator of the approach applied to video feeds from the Paris metro was live-demonstrated in a public show in Paris in front of more than 60 professionals of the field, along with other Idiap algorithms (tracking, left-luggage detection).

*Water chemistry analysis in rainfall events.* In collaboration with INRA (Institut National de Recherche en Agronomie), France, we investigated the use of our temporal probabilistic topic framework to analyze hydrology time series, and more precisely the impact of flood-induced variability on water chemistry to verify INRA's assumption that stream chemistry reacts differently to rainfall events according to the season and the weather of the hydrological year. The mixture aspect of our models and robustness to missing data proved to be interesting to explain the observed measures that depend on a mixture of environmental factors. The discovered chemistry motifs and their occurrences provided a good insight about what was going on (chemical wise) at a given water place, and matched the expert view on the data (see Fig. 8).

## **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 for personal data collection or broadcast programs. This is the challenge addressed in the french REPERE evaluation campaign, in which Idiap is participating as part of the SODA project funded by the french research agency ANR. This year, we have worked on a CRF framework for the audio-visual association and diarization of faces and speakers in broadcast data that accounts to rely on context and cluster analysis to

identify the role of people in the program (guest, anonymous person in the public, etc). Furthermore, leveraging on the text overlaid on videos that often announce speaking or appearing people, we have shown that it is possible to name and tag the most relevant persons of a program.

## **5. Artificial Cognitive Systems**

The Artificial Cognitive Systems group works on the development of multi-modal learning algorithms for enabling artificial agents to act autonomously in realistic setting, with a special emphasis on the ability to autonomously detect knowledge gaps and fill them autonomously with open ended learning strategies. The focus of the work is on designing algorithms that are principled, computationally efficient and that provides robust performance in very realistic settings while having at the same time theoretical guarantees on the expected behavior. Over the last years, the research activity has focused on two key projects: semantic spatial understanding of indoor places and advanced hand prostheses.

### ***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 and ICS projects, funded by SNSF. The VISION@HOME project aims at providing models and methods to detect, recognise and categorise the 3D shape of everyday objects and their affordances in homes. 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. Complementary to this, the main goal of the ICS project is to develop a system able to visually learn semantic concepts that characterize rooms and indoor environment, such as names referring to the activities normally performed in them (the fitness room) and the objects they contain (the bedroom). This would make it possible then to take advantage of such knowledge also in working scenarios which differ from the original ones. Taking inspiration from biological models of human perception, we have identified two



main components for the representation of indoor scenes: (1) a description of the global appearance of the image in term of image features, and (2) a description of the local landmarks present in some regions of the image. From a computational point of view the two representations could be regarded as a global appearance description of the scene, for example by means of statistics of visual features, and as a statistical representation of the co-occurrence of local concepts and scene categories. The design and integration of computational models of these two perceptual components, suitable for indoor place categorization, constitutes the core of our research.

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 global feature representation encoding at the same time task-driven and data driven spatial information. The combination of these two contributions has led us to obtain the state of the art in the field, as measured on reference benchmark databases. We also developed a new online transfer learning algorithm for leveraging over prior semantic spatial models in a dynamic, open-ended fashion, that allows to modulate the contributions of different prior sources in a principled manner.

### **Advanced Hand Prostheses**

The work on advanced hand prostheses has been initiated in 2011 with the SNSF funded NINAPRO<sup>31</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). Dr. B. Caputo is the coordinator of the project, and the principal investigator in the themes 2 and 4. 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 conducted a thorough benchmark evaluation using the most popular and effective feature descriptors proposed so far in the literature, used as input to the current off the shelf state of the art classifiers available. The evaluation was conducted on the NINAPRO database, that for the number of subjects and hand postures it contains, represents a very significant advance with respect to any previous study of this type. Our results show clearly that existing solutions that worked well on more limited settings are not able to provide the accuracy required by amputated subjects. At the same time, combining different features with advanced multi cue methods seem to obtain promising results, especially when the cues are acquired by different modal sensors.

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 pursue this vision in the large-margin classifiers framework, developing a transfer learning algorithm across multiple subjects that assumes that all prior models and the new models to be learned by the new subject all contain the same number and type of postures. In such conditions, it is possible to show that leveraging over priors significantly boost performance with a highly reduced number of repetitions by the new user. This work has been accepted for publication in the international journal *IEEE Transaction of Robotics*.

### **6. Applied Machine Learning**

Our group is interested in computer algorithms which can *learn* a behavior to achieve a given task of interest, in contrast to algorithms whose behavior is constrained by hand-crafted rules. Our research is driven from real-world applications involving a large amount of data. Domains of interest include natural language processing, computer vision, and audio processing fields. A particular emphasis is given to generic machine learning tools which require minimum a priori knowledge on the data (such as deep learning techniques), as well as unsupervised learning techniques which can leverage inherent semantic from large-scale structured data.

<sup>31</sup> [www.idiap.ch/project/ninapro/](http://www.idiap.ch/project/ninapro/)



## Deep Learning

Many state-of-the-art Natural Language Processing, Computer Vision or Audio Processing systems address a task by applying simple statistical models to carefully hand-crafted features. In other words, *the researchers themselves discover intermediate data representations* relevant for the task of interest, by engineering task-specific features. In contrast, deep learning algorithms aim at learning several layers of features representing raw data, with an increasing level of abstraction. Our group investigates both new generic deep learning algorithms, as well as algorithms specific to the following applications:

- **Natural Language Processing (NLP)**

In an ongoing collaboration with NEC Laboratories America<sup>32</sup>, we investigate new generic machine learning algorithms for various NLP tasks. We have a particular interest in *learning* generic semantic word representations, chunk of word representations, or sentence representations trained on *large unlabeled text corpora*. Recently, these semantic representations allowed us to design a single architecture which outputs tags for several NLP tasks ranging from syntax analysis (Part of Speech tagging, chunking, syntactic parsing) to semantic analysis (Name Entity Recognition, Semantic Role Labeling). Our architecture is state-of-the-art both in accuracy and speed performance.

- **Scene Parsing**

Scene parsing is a technique that consist on giving a label to all pixels in an image according to the class they belong to. Most systems rely on costly graphical models to take a global decision from local classifiers. We currently investigate new efficient end-to-end systems based on recurrent convolutional neural networks (see Figure 9), alleviating the need of any hand-crafted feature and allowing the modeling of complex spatial dependencies with a low inference cost.

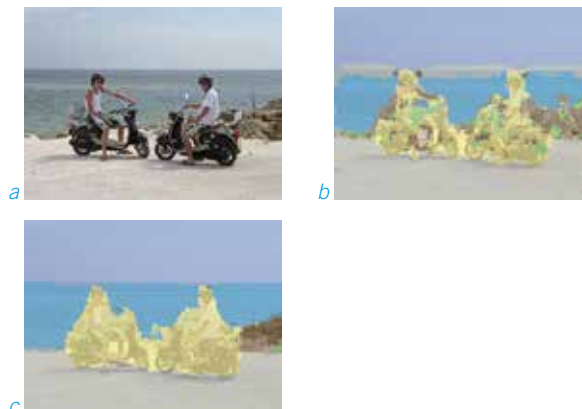


Figure 9: Our recurrent scene parser over original image (a) segments the image as shown in (b). Due to its recurrent nature, it can be fed again with its own prediction (b) and the original image (a), which leads to (c): most mistakes are corrected.

- **Spoken Term Detection**

The DeepSTD project, funded by the Swiss Hasler Stiftung<sup>33</sup>, is concerned about applying new deep learning techniques to audio processing, with a particular interest in fast spoken term detection. The year 2012 and 2013 were devoted to estimating phoneme class conditional probabilities directly from raw speech signal, using convolutional neural networks, as well as investigation ways to automate speech segmentation in these models.

### Distributed Semantic Search

Centralized search raises several issues, including ethical problems, inefficiencies due to the need of “copying” the web, and poor scaling capabilities. In the SODS project, funded by the Swiss Hasler Stiftung<sup>34</sup>, we wish to develop a new type of search engine distributed over available web servers (in the same spirit than YaCy<sup>35</sup>), in contrast to most existing search engines centralized at a single company site. Our web search engine will leverage state-of-the-art Natural Language Processing techniques, as well as distributed machine learning algorithms, in order to organize web pages, and guarantying an efficient search.

<sup>32</sup> [www.nec-labs.com/research-departments/machine-learning/machine-learning-home](http://www.nec-labs.com/research-departments/machine-learning/machine-learning-home)

<sup>33</sup> [www.haslerstiftung.ch](http://www.haslerstiftung.ch)

<sup>34</sup> [www.haslerstiftung.ch](http://www.haslerstiftung.ch). The SODS project is part of the SmartWorld special call.

<sup>35</sup> <http://yacy.de>



## Technology Transfer

Our group puts a strong emphasis on the reproducibility of all our research findings, such that our tools can be integrated easily in more complex systems or products. Tools we developed include:

- **Face Technologies**

This year we started a collaboration with the KeyLemon<sup>36</sup> startup, through the CLEAR CTI project. We are in charge of developing state-of-the-art algorithms for multi-pose face detection, head pose detection, facial feature detection, and gender detection. In 2013, we delivered a state-of-the-art version of our multi-pose face and gender detection system (see Figure 10), which runs in real time with good performance even in adverse conditions.



Figure 10: Our real time face detector detects multi-pose faces and labels gender (pink for female, bluish for male) in adverse conditions (left). It also outputs 22 precise facial features, in multi-pose conditions (right).

- **SENNA, a Natural Language Processing (NLP) tagger**

SENNA<sup>37</sup> is a standalone software distributed under a non-commercial license, that we develop through our ongoing collaboration with NEC Laboratories America<sup>38</sup>. It outputs a host of Natural Language Processing predictions, at blazing speed while holding state-of-the-art accuracy performance. Predictions include part-of-speech tags, chunking, name entity recognition, semantic role labeling and syntactic parsing.

- **TORCH, an Open-Source Machine Learning Platform**

We are maintaining Torch<sup>39</sup> an open-source (BSD license) fast and scalable Matlab-like system, which serve us as an efficient platform to implement all our machine learning algorithms. Torch leverages LuaJit, an extremely fast scripting language, which allows us to quickly develop complex demos and prototypes. Torch is widely spread, present in many international academic and private institutions. Thanks to its open-source nature, external contributors wrote various packages for a large number of application domains.

36 <https://www.keylemon.com>

37 <http://ml.nec-labs.com/senna>

38 [www.nec-labs.com/research-departments/machine-learning/machine-learning-home](http://www.nec-labs.com/research-departments/machine-learning/machine-learning-home)

39 [www.torch.ch](http://www.torch.ch)

## 7. Biometric Person Recognition

The Biometric Person Recognition group develops novel algorithms for face recognition (2D, 3D and Near-Infrared), speaker recognition, anti-spoofing (presentation attack detection) and emerging biometric modes (EEG and vascular). The group is geared towards reproducible research using its own signal-processing and machine learning toolbox.

### Face Processing

The group investigated the use of a new range of visual descriptors based on Local Binary Patterns (LBP) or on primitive binary features for the tasks of face detection, facial feature localisation and face recognition. We proposed initially to consider the histogram of LBP code as a probability distribution that can be adapted through a learning mechanism, but we also proposed a simple and effective approach to consider LBP as a pre-processing that filters out illumination variations from the face image (Figure 11).

We also explored an alternative search technique to the typical sliding window approach for the task of face detection. The proposed alternative search technique, based on primitive binary features, can improve the detection speed without decreasing for the detection rate.

More recently, we proposed a novel unified boosting framework, applied to several face processing tasks (face detection, facial feature localisation and pose classification), based on the same boosting algorithm and the same pool of features (Local Binary Patterns). This is in contrast with the standard approaches that make use of a variety of features and models, for example AdaBoost, cascades of boosted classifiers and Active Appearance Models.

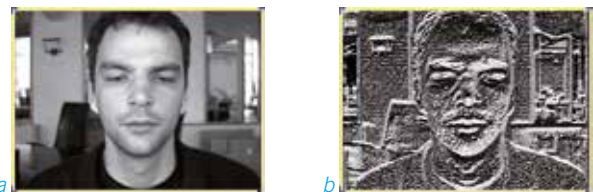


Figure 11: Illustration of the LBP feature map (b) for the original image (a).

### Face and Speaker Recognition

We leveraged prior work on distribution modelling for part-based face recognition (Figure 12) using session variability modelling techniques. Session variability modelling aims to explicitly model and suppress detrimental within-class (inter-session) variation. We examined two techniques to do this, inter-session variability modelling (ISV) and joint factor analysis (JFA), which were initially developed for speaker recognition.





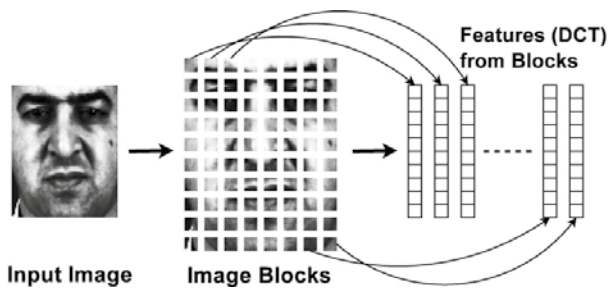


Figure 12: Illustration of the LBP feature map (b) for the original image (a).

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.

### 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 recognise identities without concern whether the sample comes from a live person or not. Despite the existence of very sophisticated biometric systems nowadays, the task of implementing 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 modelling, motion detection or spatio-temporal pattern recognition. We also organised the first International Competition on the topic.

### Reproducible Research

Over the last year the group developed various software libraries to support its research activities. Two of these libraries have been released open source starting with Torchvision (<http://torch3vision.idiap.ch/>) and finally with Bob (<http://www.idiap.ch/software/bob/>) a major evolution.

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.

### Biometrics Evaluation and Testing

Within the EU BEAT project<sup>40</sup>, coordinated by Idiap, we aim at building 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 parameterisations 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 normalisation 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 organised 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.

<sup>40</sup> [www.beat-eu.org/](http://www.beat-eu.org/)



## 8. Natural Language Processing

The Idiap Natural Language Processing group carries out research in semantic-level language analysis with two principal applications: machine translation and information retrieval, more specifically multimedia recommendation. The NLP Group reached in 2013 its maximal extension to date, with eight people. Its research is aimed at improving text analysis performance by combining robust semantic and pragmatic analysis methods with state-of-the-art, data-driven techniques for the two applications: respectively, phrase-based statistical machine translation, and content-based or collaborative filtering algorithms for multimedia recommendation over networked data.

### **Machine Translation**

Work on this topic focuses on a problem that is not so often targeted in the current statistical machine translation (SMT) paradigm: the translation of words that depend on long-range relationships between sentences. The work is mainly supported by the SNSF through its Sinergia program, first in the COMTIS project<sup>41</sup> and then in the MODERN project<sup>42</sup>, both coordinated by Idiap. We are collaborating with two teams from the University of Geneva, in linguistics and computational learning, a team from the University of Zurich in machine translation, and one from the University of Utrecht in linguistic modeling and experimentation.

We have analyzed large collections of various types of dependencies between sentences which are problematic for current MT engines, specifically rhetorical relations signaled by discourse connectives, and verb tenses which must be coherently translated across a text. We have proposed computational methods for the automatic disambiguation of discourse connectives, using classifiers such as maximum entropy or decision trees based on surface features, but also using more semantic features such as synonymy, word polarity, or temporal markers. The accuracy of our classifiers for discourse connectives (e.g. distinguishing between temporal and causal uses of the English word *since*) is competitive against the state of the art. The classifiers have been combined with phrase-based and hierarchical SMT systems, using methods such as concatenated labels or factored models. The systems were trained on large corpora (e.g. Europarl) that were automatically labeled with connectives, with English as the source language, and French, German, Italian and Arabic as target languages. We have demonstrated that the combination of connective classifiers and SMT brought improvement to the translation of discourse connectives.

Similarly, we dealt with verb tenses, which also depend on relations across sentences, and raise numerous challenges

when translating from English to French. We collected data and performed classification experiments on the translation of English past tenses into French, in collaboration with the University of Geneva, and through an internship at Idiap. A classifier using surface features was designed to predict a linguistic feature ('narrativity') in unseen texts, and the predictions were combined with SMT. Moreover, using a precise alignment of EN/FR verb phrases, we trained a classifier predicting directly the translated tense (e.g. for EN simple past into either FR *passé composé*, *passé simple*, or *imparfait*). We have demonstrated that the translation of verb phrases is improved by this method.

### **Similarity and Recommendation over Multimedia Repositories**

The NLP Group develops methods for multimedia information retrieval based on the analysis of content, including sentiment analysis, and taking advantage of the networked structure of social media repositories. This research is grounded in work on multimodal meeting analysis and meeting browsers, in the context of the past AMI<sup>43</sup> and AMIDA<sup>44</sup> EU projects, and the IM2 Swiss NCCR<sup>45</sup> finished at the end of 2013. We also are also supported by the InEvent<sup>46</sup> EU project and the REMUS Hasler project, and receive SNSF funding for technology transfer in collaboration with Klewel SA, the Idiap spin-off dedicated to conference recording and broadcasting.

In the past year, a PhD thesis was completed on learning similarity metrics over networked data, including multimedia data represented mainly through its metadata fields. A probabilistic random walk model was used to define the concept of Visiting Probability (VP) from a graph node to another node, and a distance measure was defined based on it. Moreover, several embeddings of items, offering compact representations, quicker computation and better generalization capacities were learned from VP. These similarity learning models over graphs were applied to NLP tasks such as word similarity or document clustering, and to information retrieval ones, including learning-to-rank.

Our first application framework is the Automatic Content Linking Device (ACLD), which is a multimedia retrieval system with spoken queries (explicit or implicit), running in real-time to enrich a conversation by making suggestions of potentially relevant documents. We have proposed a method, using a submodular optimization criterion, to extract from a conversation fragment, with high precision, a diverse set of keywords that reflect the diversity of topics hypothesized in the fragment. We have shown using human subjects recruited through a crowdsourcing plat-

<sup>41</sup> 2010-2013, see [www.idiap.ch/project/comtis](http://www.idiap.ch/project/comtis)

<sup>42</sup> 2013-2016, see [www.idiap.ch/project/modern](http://www.idiap.ch/project/modern)

<sup>43</sup> [www.amiproject.org](http://www.amiproject.org)

<sup>44</sup> [www.amidaproject.org](http://www.amidaproject.org)

<sup>45</sup> [www.im2.ch](http://www.im2.ch)

<sup>46</sup> [www.inevent-project.eu](http://www.inevent-project.eu)



form that our keywords were judged as more relevant than competing methods, and that they allowed the retrieval of more relevant recommendations for the conversation fragment, especially at the top levels. A method for diverse retrieval was also designed, with application to implicit but also explicit queries. We have studied the problem of retrieving multimedia networked data from the perspective of recommender systems. We have used as a benchmark the TED dataset of about 1100 lectures, with lists of user favorites and ground truth generic recommendations – we have collected its metadata from TED and made it publicly available<sup>47</sup>. We compared a number of approaches on this data set: content-based, collaborative filtering, and combined ones, and identified the most useful metadata features. We have shown that performing sentiment analysis on user comments improves the accuracy of recommendations. We have considered aspect-based sentiment analysis and proposed an innovative method based on multiple-instance learning to determine the contribution of each sentence in a review to aspect-related judgments, demonstrating improvement over the state-of-the-art on seven data sets (including TED).

The NLP Group participated in the MediaEval 2013 campaign, specifically in the Search and Hyperlinking task. The search sub-task required finding a determined segment of a show (from 1260 hours of BBC TV material) based on a test query looking for a known item, while the hyperlinking sub-task required finding items related to the known items. We proposed a unified approach to both sub-tasks, using content-based recommendation techniques to compute the most similar segments to a given text query or to another segment, based on words but also on visual concepts. For the search task, we ranked third out of seven institutions, while for the hyperlinking task (the closest to our expertise from inEvent) our team was ranked first, reaching a mean average precision above 50%.



Figure 13: The MUST-VIS navigation graph (bottom) and video/slide player (top). Each lecture is represented either with a key-frame or with keyword clouds (the two modes are switched by clicking). The lecture in focus, center, is surrounded by lectures with related segments.

In collaboration with our partners in the inEvent project, and in particular with Klewel, the NLP Group was the winner of one of the five ACM Multimedia 2013 Grand Challenge Solutions, specifically the one for lecture segmentation and retrieval. Our system, represented in Figure 13, was called MUST-VIS for Multi-factor Segmentation for Topic Visualization and Recommendation, and allowed users to visualize a lecture as a series of segments, with relations to other similar lectures and segments, computed using recommendation techniques.

<sup>47</sup> [www.idiap.ch/dataset/ted](http://www.idiap.ch/dataset/ted)



# MAIN PROJECTS IN PROGRESS

## Projects funded by European agencies

- Name** **BBFOR2** (Bayesian Biometrics for Forensics)  
**Funding** FP7 - Marie Curie Actions (ITN)  
**Coordinator** Radboud University Nijmegen - Stichting Katholieke Universiteit  
**Duration** 01.01.2010-31.12.2013  
**Partner(s)** Katholieke Universiteit Leuven, Idiap Research Institute, Högskolan i Halmstad, University of York, Universiteit Twente, Universidad Autónoma de Madrid, Politecnico di Torino, Netherlands Forensic Institute
- Name** **BEAT** (Biometrics Evaluation and Testing)  
**Funding** FP7 - STREP  
**Coordinator** Idiap Research Institute  
**Duration** 01.03.2012-28.02.2016  
**Partner(s)** University of Surrey, Universidad Autónoma de Madrid, Tüvit, Tubutak Informatics and Information Security, Advanced Technologies Research Center, Sagem Sécurité S.A., Katholieke Universiteit Leuven, École Polytechnique Fédérale de Lausanne, Commissariat à l'Énergie Atomique
- Name** **EUMSSI** (Event Understanding through Multimodal Social Stream Interpretation)  
**Funding** FP7 - IP  
**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** (A generic dialog box for multilingual conversational applications)  
**Funding** EUREKA's 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** **HUMAVIPS** (Humanoids with auditory and visual abilities in populated spaces)  
**Funding** FP7 - STREP  
**Coordinator** Institut de Recherche en Informatique et en Automatique  
**Duration** 01.02.2010-31.01.2013  
**Partner(s)** Jean-Marc Odobez, Hervé Boulard, François Fleuret, François Fleuret
- Name** **INEVENT** (Accessing Dynamic Networked Multimedia Events)  
**Funding** FP7 - STREP  
**Coordinator** Idiap Research Institute  
**Duration** 01.11.2011-31.10.2014  
**Partner(s)** University of Edinburgh, Radvision LTD, Klewel, IBM Israel LTD, Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V.
- Name** **MASH** (Massive Sets of Heuristics for Machine Learning)  
**Funding** FP7 - STREP  
**Coordinator** Idiap Research Institute  
**Duration** 01.01.2010-30.06.2013  
**Partner(s)** Centre national de la recherche scientifique, Universitaet Potsdam, Institut de Recherche en Informatique et en Automatique, Czech Technical University in Prague
- Name** **PASCAL2** (Pattern Analysis, Statistical Modelling and Computational Learning 2)  
**Funding** FP7 - NoE  
**Coordinator** University of Southampton  
**Duration** 01.03.2008-28.02.2013  
**Partner(s)** University College London, Helsinki University of Technology, Max Planck Society for the Advancement of Sciences, Jozef Stefan Institute, Idiap Research Institute, Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V., Stichting Centrum voor Wiskunde en Informatica, Centre national de la recherche scientifique, Xerox SAS, Université Pierre et Marie Curie-Paris 6, University of Bristol, Università degli Studi di Milano, University of Manchester, Helsingin yliopisto, University of Edinburgh



9. Name **ROCKIT** (Roadmap for Conversational Interaction Technologies)  
 Funding FP7- CSA  
 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
10. Name **SSPNET** (Social Signal Processing Network)  
 Funding FP7 - NoE  
 Coordinator Idiap Research Institute  
 Duration 01.02.2009-31.01.2014  
 Partner(s) Universiteit Twente, Università di Roma Tre, University of Geneva, University of Edinburgh, Delft University of Technology, Queen's University Belfast, Institut de Recherche en Informatique et en Automatique, Imperial College London, German Research Centre for Artificial Intelligence
11. Name **TABULARASA** (Trusted Biometrics under Spoofing Attacks)  
 Funding FP7 - STREP  
 Coordinator Idiap Research Institute  
 Partner(s) University of Southampton, University of Oulu, Università Degli Studi Di Cagliari, Universidad Autónoma de Madrid, Startlab Barcelona SL, Sagem Sécurité S.A., KeyLemon SA, Eurocom, Biometry.com AG, Centre for Science, Society and Citizenship, Institute of Automation Chinese Academy of Sciences
12. Name **VANAHEIM** (Video/Audio Networked surveillance system enhancement through Human-centered adaptive Monitoring)  
 Funding FP7 - IP  
 Coordinator Multitel ASBL  
 Duration 01.02.2010-30.09.2013  
 Partner(s) Thales Italia spa, Thales Communications SA, Régie Autonome des Transports Parisiens, Ludwig Boltzmann Institute for Urban Ethology, Institut de Recherche en Informatique et en Automatique, Gruppo Torinese Trasporti
13. Name **SODA** (Person Recognition in debate and broadcast news)  
 Funding ANR - Agence Nationale de la Recherche  
 Coordinator Université du Maine  
 Duration 01.06.2011-31.05.2014  
 Partner(s) Idiap Research Institute



## Projects funded by Swiss agencies

1. Name **A-MUSE** (Adaptive Multilingual Speech Processing)  
Funding SNSF - Swiss National Science Foundation  
Coordinator Idiap Research Institute  
Duration 01.10.2012-30.09.2015  
Partner(s)
2. Name **AROLIS** (Automatic Recommendation of Lectures and Snippets)  
Funding SNSF - Swiss National Science Foundation  
Coordinator Idiap Research Institute  
Duration 01.06.2012-31.05.2014  
Partner(s) Klewel
3. Name **COMTIS** (Improving the coherence of machine translation output by modeling intersentential relations)  
Funding SNSF - Swiss National Science Foundation  
Coordinator Idiap Research Institute  
Duration 01.03.2010-31.07.2013  
Partner(s) University of Geneva
4. Name **DASH** (Object Detection with Active Sample Harvesting)  
Funding SNSF - Swiss National Science Foundation  
Coordinator Idiap Research Institute  
Duration 01.09.2012-31.08.2015  
Partner(s)
5. Name **DIMHA** (Diarizing Massive Amounts of Heterogeneous Audio)  
Funding SNSF - Swiss National Science Foundation  
Coordinator Idiap Research Institute  
Duration 01.06.2012-31.05.2014  
Partner(s) Koemei
6. Name **FLEXASR** (Flexible Grapheme-Based Automatic Speech Recognition)  
Funding SNSF - Swiss National Science Foundation  
Coordinator Idiap Research Institute  
Duration 01.03.2010-30.04.2013  
Partner(s)
7. Name **FLEXASR** (Flexible Grapheme-Based Automatic Speech Recognition)  
Funding SNSF - Swiss National Science Foundation  
Coordinator Idiap Research Institute  
Duration 01.05.2013-30.04.2014  
Partner(s)
8. Name **HAI-2010** (Human activity and interactivity modeling)  
Funding SNSF - Swiss National Science Foundation  
Coordinator Idiap Research Institute  
Duration 01.10.2010-30.09.2013  
Partner(s)
9. Name **ICS-2** (Interactive Cognitive Systems, Indoor Scene Recognition for Intelligent Systems)  
Funding SNSF - Swiss National Science Foundation  
Coordinator Idiap Research Institute  
Duration 01.04.2013-31.12.2013  
Partner(s)
10. Name **IM2-3** (Interactive Multimodal Information Management Phase 3)  
Funding SNSF - Swiss National Science Foundation  
Coordinator Idiap Research Institute  
Duration 01.01.2010-31.12.2013  
Partner(s)
11. Name **LOBI** (Low Complexity Binary Features for Robust-to-Noise Speaker Recognition)  
Funding SNSF - Swiss National Science Foundation  
Coordinator Idiap Research Institute  
Duration 01.06.2012-31.05.2014  
Partner(s) KeyLemon SA



12. Name **MAAYA** (Multimedia Analysis and Access for Documentation and Decipherment of Maya Epigraphy)  
 Funding SNSF - Swiss National Science Foundation  
 Coordinator Idiap Research Institute  
 Duration 01.06.2013-31.05.2016  
 Partner(s) University of Geneva, University of Bonn
13. Name **MCM-FF** (Multimodal Computational Modeling of Nonverbal Social Behavior in Face to Face Interaction)  
 Funding SNSF - Swiss National Science Foundation  
 Coordinator Idiap Research Institute  
 Duration 01.11.2011-31.10.2014  
 Partner(s)
14. Name **MODERN** (Modeling discourse entities and relations for coherent machine translation)  
 Funding SNSF - Swiss National Science Foundation  
 Coordinator Idiap Research Institute  
 Duration 01.08.2013-31.07.2016  
 Partner(s) Universiteit Utrecht, Universität Zürich, University of Geneva
15. Name **NINAPRO** (Non-Invasive Adaptive Hand Prosthetics)  
 Funding SNSF - Swiss National Science Foundation  
 Coordinator Idiap Research Institute  
 Duration 01.01.2011-30.06.2014  
 Partner(s) HES-SO Valais, German Aerospace Research Center
16. Name **PROMOVAR** (Probabilistic Motifs for Video Action Recognition)  
 Funding SNSF - Swiss National Science Foundation  
 Coordinator Idiap Research Institute  
 Duration 01.02.2012-31.07.2013  
 Partner(s)
17. Name **RODI** (Role based speaker diarization)  
 Funding SNSF - Swiss National Science Foundation  
 Coordinator Idiap Research Institute  
 Duration 01.11.2011-31.10.2014  
 Partner(s)
18. Name **SIVI** (Situating Vision to Perceive Object Shape and Affordances)  
 Funding SNSF - Swiss National Science Foundation  
 Coordinator University of Bonn  
 Duration 25.10.2011-30.09.2014  
 Partner(s) Technische Universität Wien, RWTH Aachen, Idiap Research Institute
19. Name **SIWIS** (Spoken Interaction with Interpretation in Switzerland)  
 Funding SNSF - Swiss National Science Foundation  
 Coordinator Idiap Research Institute  
 Duration 01.12.2012-30.11.2015  
 Partner(s) University of Geneva, University of Edinburgh, Eidgenössische Technische Hochschule Zürich
20. Name **SONVB** (Sensing and Analysing Organizational Nonverbal Behavior)  
 Funding SNSF - Swiss National Science Foundation  
 Coordinator Idiap Research Institute  
 Duration 01.06.2010-30.06.2014  
 Partner(s) Université de Neuchâtel, The Trustees of Dartmouth College Corp.
21. Name **TRACOME** (Robust face tracking, feature extraction and multimodal fusion for audio-visual speech recognition)  
 Funding SNSF - Swiss National Science Foundation  
 Coordinator École Polytechnique Fédérale de Lausanne  
 Duration 01.01.2011-31.03.2014  
 Partner(s) Idiap Research Institute
22. Name **UBM** (Understanding Brain Morphogenesis)  
 Funding SNSF - Swiss National Science Foundation



- Coordinator University of Basel  
Duration 01.01.2011-31.12.2013  
Partner(s) Idiap Research Institute, École Polytechnique Fédérale de Lausanne
23. Name **VELASH** (Very Large Sets of Heuristics for Scene Interpretation)  
Funding SNSF - Swiss National Science Foundation  
Coordinator Idiap Research Institute  
Duration 01.09.2012-31.08.2013  
Partner(s)
24. 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)
25. Name **DEEPSTD** (Universal Spoken Term Detection with Deep Learning) Funding Hasler Foundation  
Coordinator Idiap Research Institute  
Duration 01.12.2011-30.11.2014  
Partner(s)
26. Name **IMAGECLEF** (The Robot Vision Task @ ImageCLEF: Towards Web-Robotics)  
Funding Hasler Foundation  
Coordinator Idiap Research Institute  
Duration 01.01.2013-28.02.2014  
Partner(s)
27. 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)
28. Name **SESAME** (SEarching Swiss Audio MEMories)  
Funding Hasler Foundation  
Coordinator Idiap Research Institute  
Duration 01.01.2011-28.02.2015  
Partner(s)
29. Name **SODS** (Semantically Self-Organized Distributed Web Search)  
Funding Hasler Foundation  
Coordinator Idiap Research Institute  
Duration 01.01.2012-31.12.2014  
Partner(s)
30. Name **UBSL** (User-Based Similarity Learning for Interactive Image Retrieval)  
Funding Hasler Foundation  
Coordinator Idiap Research Institute  
Duration 01.08.2012-31.03.2013  
Partner(s)
31. Name **MEDIAPARL** (MediaParl) Funding Etat du Valais  
Coordinator Idiap Research Institute  
Duration 01.01.2012-31.12.2013  
Partner(s)
32. Name **VALAIS-2015** (Valais\*Wallis Digital)  
Funding Etat du Valais  
Coordinator Idiap Research Institute  
Duration 01.04.2013-31.12.2015  
Partner(s) Société Coopérative Migros Valais, Médiathèque Valais et Archives de l'Etat, Ecole professionnelle des arts contemporains





## Projects funded by industrial partners

1. Name **ARIS** (Automatic French and German Speech Recognition System for Efficient Indexing and Search)  
 Funding Fondation The Ark  
 Coordinator Idiap Research Institute  
 Duration 01.04.2013-31.01.2014  
 Partner(s) Koemei, Klewel
2. Name **COMMIC2** (Système cognitif pour une commande intelligente de chauffage, phase2)  
 Funding Fondation The Ark  
 Coordinator Centre de recherches energetiques et municipales  
 Duration 01.09.2012-30.06.2013  
 Partner(s) Idiap Research Institute, HES-SO Valais, ARLO Communications
3. Name **LAMI** (Large scale multi-energy information platform)  
 Funding Fondation The Ark  
 Coordinator Centre de recherches energetiques et municipales  
 Duration 01.08.2013-28.02.2014  
 Partner(s) ESR - Energie Sion Region, Sinergy, Idiap Research Institute, HES-SO Valais
4. Name **SMARTCAP** (Smart Caps Orientation)  
 Funding Fondation The Ark  
 Coordinator AISA Automation Industrielle SA  
 Duration 01.05.2013-21.02.2014  
 Partner(s) Idiap Research Institute
5. Name **ASLEEP** (Adapting the Static Luggage Detection module for the Protect Rail demonstrator)  
 Funding IND - Industrial Project  
 Coordinator Idiap Research Institute  
 Duration 15.11.2012-31.05.2013  
 Partner(s)
6. Name **DAUM2012** (Domain Adaptation Using Sub-Space Models)  
 Funding IND - Industrial Project  
 Coordinator Idiap Research Institute  
 Duration 15.10.2012-14.10.2013  
 Partner(s)
7. Name **NEC** (NEC collaboration)  
 Funding IND - Industrial Project  
 Coordinator Idiap Research Institute  
 Duration 01.05.2012-30.04.2014  
 Partner(s)
8. Name **NTT** (NISHA - NTT Idiap Social Behaviour Analysis Initiative)  
 Funding IND - Industrial Project  
 Coordinator Idiap Research Institute  
 Duration 15.11.2012-30.09.2014  
 Partner(s)
9. Name **CLEAR** (Online Cloud-based Platform for Efficient and Robust Face Recognition Servi)  
 Funding CTI  
 Coordinator Idiap Research Institute  
 Duration 01.06.2012-30.11.2013  
 Partner(s) KeyLemon
10. Name **DOMOCARE** (DomoCare - A new Home Care Preventive Protocol)  
 Funding CTI  
 Coordinator Idiap Research Institute  
 Duration 01.10.2013-30.09.2014  
 Partner(s) DomoSafety
11. Name **FEDARS** (Feature Extraction from Deep learning Architectures for face Recognition Systems)  
 Funding CTI  
 Coordinator Idiap Research Institute  
 Duration 01.08.2013-31.01.2014  
 Partner(s) KeyLemon SA



12. 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
13. Name **PANDA** (Perceptual Background Noise Analysis for the Newest Generation of Telecommunication Systems)  
Funding CTI  
Coordinator Idiap Research Institute  
Duration 01.07.2012-31.12.2013  
Partner(s) SwissQual AG
14. Name **RECOD2013** (low bit-rate speech coding)  
Funding ArmaSuisse  
Coordinator Idiap Research Institute  
Duration 01.03.2013-31.12.2013  
Partner(s)



## Projects awarded in 2013 and starting in 2014

- 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** **REMUS** (REMUS: Re-ranking Multiple Search Results for Just-in-Time Document Recommendation)  
**Funding** Hasler Foundation  
**Coordinator** Idiap Research Institute  
**Duration** 01.01.2014-31.10.2014  
**Partner(s)**
- Name** **SCOREL2** (Automatic scoring and adaptive pedagogy for oral language learning)  
**Funding** CTI  
**Coordinator** Idiap Research Institute  
**Duration** 01.04.2014-30.11.2015  
**Partner(s)** SpeedLingua
- Name** **SIIP** (Speaker Identification Integrated Project)  
**Funding** FP7 - IP  
**Coordinator** Verint System Ltd  
**Duration** 01.05.2014-30.04.2018  
**Partner(s)** Loquendo SpA, The International Criminal Police Organization (Interpol), Inov Inesc inovacao - Instituto De Novas Tecnologias, Idiap Research Institute, International Biometric Group (UK) Limited, green Fusion Limited (data Fusion International), Cassidan SAS; University of Warwick, Snthema S.R.L, Singularlogic Anonymos Etairia Pliroforiakon Systematon & Efarmogon Pliroforkis, Sail Labs Technology AG, Rijksuniversiteit Gronigen, Police Service of Northern Ireland, Ministério da Justiça, Lisboa, ok2go, Laboratorio Di Scienze Della Cittadinanza
- Name** **SUVA** (Intégration de la transcription vocale dans le dossier patient informatisé CRR)  
**Funding** Fondation The Ark  
**Coordinator** Idiap Research Institute  
**Duration** 01.03.2014-31.03.2014  
**Partner(s)** Clinique romande de réadaptation, Cimark
- Name** **UBIMPRESSED** (UBImpressed: Ubiquitous First Impressions and Ubiquitous Awareness)  
**Funding** SNSF - Swiss National Science Foundation  
**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 - Swiss National Science Foundation  
**Coordinator** Idiap Research Institute  
**Duration** 01.01.2014-31.12.2016  
**Partner(s)** Eidgenössische Technische Hochschule Zürich, École Polytechnique Fédérale de Lausanne
- Name** **YOUTH@NIGHT** (A multi-disciplinary multi-method study of young people's outgoing and drinking behaviors)  
**Funding** SNSF - Swiss National Science Foundation  
**Coordinator** Sucht Schweiz - Research Institute  
**Duration** 01.01.2014-31.12.2016  
**Partner(s)** Universität Zürich, Idiap Research Institute



# LIST OF PUBLICATIONS – 2013

## 1. Book

1. Hervé Bourlard and Andrei Popescu-Belis. *Interactive Multimodal Information Management*. EPFL Press, Lausanne, 2013.
2. Bonnie Webber, Andrei Popescu-Belis, Katja Markert, and Jorg Tiedemann. *Proceedings of the ACL Workshop on Discourse in Machine Translation (DiscoMT 2013)*. Association for Computational Linguistics, August 2013.

## 2. Book Chapters

1. Barbara Caputo. Medical image annotation. In Hervé Bourlard and Andrei Popescu-Belis, editors, *Interactive Multimodal Information Management*. EPFL Press, 2013.
2. Remi Emonet and Jean-Marc Odobez. Unsupervised methods for activity analysis and detection of abnormal events. In J. Y. Dufour, editor, *Intelligent Video Surveillance Systems (ISTE)*. Wiley, February 2013.
3. Francois Fleuret, Horesh Ben Shitrit, and Pascal Fua. Re-identification for improved people tracking. In *Person Re-Identification*. Springer, 2013.
4. Mathew Magimai-Doss. Speech processing. In Hervé Bourlard and Andrei Popescu-Belis, editors, *Interactive Multimodal Information Management*, chapter 15, pages 221–245. EPFL Press, 2013.
5. Fabian Nater, Tatiana Tommasi, Luc Van Gool, and Barbara Caputo. Learning to learn new models of human activities in indoor settings. In Hervé Bourlard and Andrei Popescu-Belis, editors, *Interactive Multimodal Information Management*. EPFL Press, 2013.
6. Andrei Popescu-Belis and Hervé Bourlard. Interactive multimodal information management: Shaping the vision. In Hervé Bourlard and Andrei Popescu-Belis, editors, *Interactive Multimodal Information Management*, pages 1–17. EPFL Press, Lausanne, 2013.

## 3. Articles in Scientific Journals

1. Marcel Alcoverro, Xavier Suau, Adolfo Lopez-Mendez, Josep R. Morros, Javier Ruiz-Hidalgo, Albert Gil, and Josep R. Casas. Gesture control interface for immersive panoramic displays. *Multimedia Tools and Applications*, 1380-7501:1–27, July 2013.
2. André Anjos, Murali Mohan Chakka, and Sébastien Marcel. Motion-based counter-measures to photo attacks in face recognition. *Institution of Engineering and Technology Journal on Biometrics*, July 2013.
3. Afsaneh Asaei, Hervé Bourlard, Bhiksha Raj, Mohammad J. Taghizadeh, and Volkan Cevher. Binary sparse coding for optimally speech reconstruction. *IEEE Transactions on Audio, Speech and Language Processing*, 2013.
4. Afsaneh Asaei, Hervé Bourlard, Mohammad J. Taghizadeh, and Volkan Cevher. Computational methods for spatio-spectral speech recovery via sparse modeling. *Journal of Signal Processing Systems*, 2013.
5. Afsaneh Asaei, Mohammad Golbabaee, Hervé Bourlard, and Volkan Cevher. Structured sparsity models for reverberant speech separation. *IEEE Transaction on Audio, Speech and Language Processing*, 2013.
6. Alice Aubert, Romain Tavenard, Remi Emonet, A. de Lavenne, Simon Malinowski, Thomas Guyet, René Quiniou, Jean-Marc Odobez, Philippe Merot, and Chantal Gascuel. Clustering flood events from water quality time-series using latent dirichlet allocation model. *Water Resources Research*, 2013. Online published version before inclusion in an issue.
7. Horesh Ben Shitrit, Jerome Berclaz, Francois Fleuret, and Pascal Fua. Multi-commodity network flow for tracking multiple people. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013.
8. Joan-Isaac Biel, Daniel Gatica-Perez, John Dines, and Vagia Tsminiaki. Hi youtube! personality impressions and verbal content in social video. *15th ACM International Conference on Multimodal Interaction, Sydney, Australia, ACM, 2013*, 2013.
9. Antoine Bordes, Léon Bottou, Ronan Collobert, Dan Roth, Jason Weston, and Luke Zettlemoyer. Introduction to the special issue on learning semantics. *Machine Learning*, June 2013.



10. Bruno Cartoni, Sandrine Zufferey, and Thomas Meyer. Annotating the meaning of discourse connectives by looking at their translation: The translation-spotting technique. *Dialogue and Discourse*, 4(2):65–86, April 2013.
11. Bruno Cartoni, Sandrine Zufferey, and Thomas Meyer. Using the europarl corpus for cross-linguistic research. *Belgian Journal of Linguistics*, (27):23–42, December 2013.
12. Trinh-Minh-Tri Do and Daniel Gatica-Perez. The places of our lives: Visiting patterns and automatic labeling from longitudinal smartphone data. *IEEE Transactions on Mobile Computing*, 2013.
13. Trinh-Minh-Tri Do and Daniel Gatica-Perez. Where and what: Using smartphones to predict next locations and applications in daily life. *Pervasive and Mobile Computing*, May 2013.
14. Stefan Duffner and Jean-Marc Odobez. A track creation and deletion framework for long-term online multi-face tracking. *IEEE Transactions on Image Processing*, March 2013.
15. Laurent El Shafey, Chris McCool, Roy Wallace, and Sébastien Marcel. A scalable formulation of probabilistic linear discriminant analysis: Applied to face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(7):1788–1794, July 2013.
16. Remi Emonet, Jagannadan Varadarajan, and Jean-Marc Odobez. Temporal analysis of motif mixtures using dirichlet processes. *IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI)*, May 2013.
17. Philip N. Garner, Milos Cernak, and Petr Motlicek. A simple continuous pitch estimation algorithm. *IEEE Signal Processing Letters*, 20(1):102–105, January 2013.
18. David Imseng, Hervé Bourlard, John Dines, Philip N. Garner, and Mathew Magimai-Doss. Applying multi- and cross-lingual stochastic phone space transformations to non-native speech recognition. *IEEE Transactions on Audio, Speech, and Language Processing*, 2013.
19. David Imseng, Petr Motlicek, Hervé Bourlard, and Philip N. Garner. Using out-of-language data to improve an under-resourced speech recognizer. *Speech Communication*, 2013.
20. Elie Khoury, Laurent El Shafey, Chris McCool, Manuel Günther, and Sébastien Marcel. Bi-modal biometric authentication on mobile phones in challenging conditions. *Image and Vision Computing*, October 2013.
21. S. R. Krishnan, Mathew Magimai-Doss, and C. S. Seelamantula. A savitzky-golay filtering perspective of dynamic feature computation. *IEEE Signal Processing Letters*, 20(3):281 – 284, March 2013.
22. J. K. Laurila, Daniel Gatica-Perez, Jan Blom, Olivier Bornet, Trinh-Minh-Tri Do, O. Dousse, Julien Eberle, and Markus Miettinen. From big smartphone data to worldwide research: The mobile data challenge. *Pervasive and Mobile Computing*, 2013.
23. Riwal Lefort and Francois Fleuret. treekl: A distance between high dimension empirical distributions. *Pattern Recognition Letters*, 34(2):140–145, 2013.
24. Patrick Marmoroli, M. Carmona, Xavier Falourd, Hervé Lissek, and Jean-Marc Odobez. Observation of vehicle axles through pass-by noise: A strategy of microphone array design. *IEEE Trans. on Intelligent Transportation Systems*, March 2013.
25. Chris McCool, Roy Wallace, Mitchell McLaren, Laurent El Shafey, and Sébastien Marcel. Session variability modelling for face authentication. *IET Biometrics*, 2(3):117–129, September 2013.
26. Michael McCoy, Volkan Cevher, Quoc Tran Dinh, Afsaneh Asaei, and Luca Baldassarre. Convexity in source separation: Models, geometry, and algorithms. *IEEE Signal Processing Magazine, Special Issue on Source Separation and Applications*, 2013.
27. Petr Motlicek, Stefan Duffner, Danil Korchagin, Hervé Bourlard, Carl Scheffler, Jean-Marc Odobez, Giovanni Del Galdo, Markus Kallinger, and Oliver Thiergart. Real-time audio-visual analysis for multiperson videoconferencing. *Advances in Multimedia*, 2013:21, August 2013. Hindawi Publishing Corporation, Article ID 175745.
28. Lakshmi Saheer, Junichi Yamagishi, Philip N. Garner, and John Dines. Combining vocal tract length normalization with hierarchical linear transformations. *IEEE Journal of Selected Topics in Signal Processing - Special Issue on Statistical Parametric Speech Synthesis*, 2013.
29. Björn Schuller, Stefan Steidl, Anton Batliner, Elmar Nöth, Alessandro Vinciarelli, Felix Burkhardt, Felix Weninger, Florian Eyben, Tobias Bocklet, Gelareh Mohammadi, and Benjamin Weiss. A survey on perceived speaker traits: Personality, likability, pathology and the first challenge. *Computer Speech and Language*, 2013.
30. Jagannadan Varadarajan, Remi Emonet, and Jean-Marc Odobez. A sequential topic model for mining recurrent activities from long term video logs. *International Journal of Computer Vision*, 103(1):100–126, May 2013.



31. Alessandro Vinciarelli and Gelareh Mohammadi. A survey of personality computing. *IEEE Transaction on Affective Computing*, 2013.
32. Majid Yazdani and Andrei Popescu-Belis. Computing text semantic relatedness using the contents and links of a hyper-text encyclopedia. *Artificial Intelligence Journal*, 194:176202, January 2013.

#### 4. Articles in Conference Proceedings

1. Aniruddha Adiga, Mathew Magimai-Doss, and Chandra Sekhar Seelamantula. Gammatone wavelet cepstral coefficients for robust speech recognition. In *Proceedings of IEEE TENCON*, October 2013.
2. Oya Aran and Daniel Gatica-Perez. Cross-domain personality prediction: From video blogs to small group meetings. In *15th ACM International Conference on Multimodal Interaction*, 2013.
3. Oya Aran and Daniel Gatica-Perez. One of a kind: Inferring personality impressions in meetings. In *15th ACM International Conference on Multimodal Interaction*, 2013.
4. Afsaneh Asaei, Mohammad Golbabaee, Hervé Boursard, and Volkan Cevher. Room acoustic modeling exploiting joint sparsity and lowrank structures. In *Signal Processing with Adaptive Sparse Structured Representations SPARS*, 2013.
5. Afsaneh Asaei, Bhiksha Raj, Hervé Boursard, and Volkan Cevher. A multipath sparse beamforming method. In *Signal Processing with Adaptive Sparse Structured Representations SPARS*, 2013.
6. Alice Aubert, Romain Tavenard, Simon Malinowski, Thomas Guyet, René Quiniou, Jean-Marc Odobez, Remi Emonet, and Chantal Gascuel. Discovering temporal patterns in water quality time series, focusing on floods with the Ida method. In *European Geosciences Union*, 2013.
7. Chidansh A. Bhatt, Nikolaos Pappas, Maryam Habibi, and Andrei Popescu-Belis. I diap at mediaeval 2013: Search and hyperlinking task. In *MediaEval 2013 Workshop*, CEUR Workshop Proceedings. CEURWS.org, October 2013.
8. Chidansh A. Bhatt, Andrei Popescu-Belis, Maryam Habibi, Sandy Ingram, Stefano Masneri, Fergus McInnes, Nikolaos Pappas, and Oliver Schreer. Multi-factor segmentation for topic visualization and recommendation: the must-vis system. In *Proceedings of the 21st ACM International Conference on Multimedia*. ACM, 2013.
9. Hervé Boursard, Marc Ferras, Nikolaos Pappas, Andrei Popescu-Belis, Steve Renals, Fergus McInnes, Peter Bell, Sandy Ingram, and Maël Guillemot. Processing and linking audio events in large multimedia archives: The eu inevent project. In *Workshop on Speech, Language and Audio in Multimedia*, July 2013.
10. Milos Cernak, Petr Motlicek, and Philip N. Garner. On the (un) importance of the contextual factors in hmm-based speech synthesis. In *Proceedings of the IEEE Intl. Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 8140 – 8143, May 2013.
11. Milos Cernak, Xingyu Na, and Philip N. Garner. Syllable-based pitch encoding for low bit rate speech coding with recognition/synthesis architecture. In *Proc. of Interspeech 2013*, August 2013.
12. Ivana Chingovska, André Anjos, and Sébastien Marcel. Anti-spoofing in action: joint operation with a verification system. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Workshop on Biometrics*, June 2013.
13. Ivana Chingovska, Jinwei Yang, Zhen Lei, Dong Yi, Stan Z.Li, Olga Kahm, Naser Damer, Christian Glaser, Arjan Kuijper, Alexander Nouak, Jukka Komulainen, Tiago de Freitas Pereira, Shubham Gupta, Shubham Bansal, Shubham Khandelwal, Ayush Rai, Tarun Krishna, Dushyant Goyal, Muhammad-Adeel Waris, Honglei Zhang, Iftikhar Ahmad, Serkan Kiranyaz, Moncef Gabbouj, Roberto Tronci, Maurizio Pili, Nicola Sirena, Fabio Roli, Javier Galbally, Julian Fierrez, Allan Pinto, Helio Pedrini, William Robson Schwartz, Anderson Rocha, André Anjos, and Sébastien Marcel. The 2nd competition on counter measures to 2d face spoofing attacks. In *International Conference of Biometrics 2013*, June 2013.
14. Tiago de Freitas Pereira, André Anjos, José Mario De Martino, and Sébastien Marcel. Can face anti-spoofing countermeasures work in a real world scenario? In *International Conference on Biometrics*, June 2013.
15. Trinh-Minh-Tri Do, Kyriaki Kalimeri, Bruno Lepri, Fabio Pianesi, and Daniel Gatica-Perez. Inferring social activities with mobile sensor networks. In *15th ACM International Conference on Multimodal Interaction*, December 2013.
16. Charles Dubout and Francois Fleuret. Accelerated training of linear object detectors. In *CVPR 2013 Workshop on Structured Prediction*, 2013.
17. Charles Dubout and Francois Fleuret. Deformable part models with individual part scaling. In *British Machine Vision Conference*, 2013.



18. Nesli Erdogmus and Sébastien Marcel. Spoofing attacks to 2d face recognition systems with 3d masks. In *International Conference of the Biometrics Special Interest Group*, September 2013.
19. Nesli Erdogmus and Sébastien Marcel. Spoofing in 2d face recognition with 3d masks and antispoofing with kinect. In *Biometrics: Theory, Applications and Systems*, September 2013.
20. Marco Fornoni, Barbara Caputo, and Francesco Orabona. Multiclass latent locally linear support vector machines. In Cheng Soon Ong and Tu-Bao Ho, editors, *JMLR W&CP, Volume 29: Asian Conference on Machine Learning*, pages 229–244, 2013.
21. Kenneth Alberto Funes Mora. 3d head pose and gaze tracking and their application to diverse multimodal tasks. In *Doctoral consortium of the 15th ACM International Conference on Multimodal Interaction*, December 2013.
22. Kenneth Alberto Funes Mora, Laurent Son Nguyen, Daniel Gatica-Perez, and Jean-Marc Odobez. A semi-automated system for accurate gaze coding in natural dyadic interactions. In *15th ACM International Conference on Multimodal Interaction*. ACM, December 2013.
23. Kenneth Alberto Funes Mora and Jean-Marc Odobez. Person independent 3d gaze estimation from remote rgb-d cameras. In *International Conference on Image Processing*. IEEE, September 2013.
24. Arjan Gijsberts and Barbara Caputo. Exploiting accelerometers to improve movement classification for prosthetics. In *International Conference on Rehabilitation Robotics*, 2013.
25. Baran ozcu, Afsaneh Asaei, and Volkan Cevher. Manifold sparse beamforming. In *IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing*, 2013.
26. Manuel Günther, Artur Costa-Pazo, Changxing Ding, Elhocine Boutellaa, Giovanni Chiachia, Honglei Zhang, Marcus de Assis Angeloni, Vitomir Struc, Elie Khoury, Esteban Vazquez-Fernandez, Dacheng Tao, Messaoud Bengherabi, David Cox, Serkan Kiranyaz, Tiago de Freitas Pereira, Jerneja Zganec-Gros, Enrique Argones-Rua, Nicolas Pinto, Moncef Gabbouj, Flavio Simões, Simon Dobrisek, Daniel Gonzalez-Jiménez, Anderson Rocha, Mario Uliani Neto, Nikola Pavesic, Alexandre Falcao, Ricardo Violato, and Sébastien Marcel. The 2013 face recognition evaluation in mobile environment. In *The 6th IAPR International Conference on Biometrics*, June 2013.
27. Maryam Habibi and Andrei Popescu-Belis. Diverse keyword extraction from conversations. In *Proceedings of the ACL 2013 (51th Annual Meeting of the Association for Computational Linguistics), Short Papers*, 2013.
28. Najeh Hajlaoui. Are act's scores increasing with better translation quality? In *Are ACT's scores increasing with better translation quality?*, page 6, July 2013.
29. Najeh Hajlaoui and Andrei Popescu-Belis. Assessing the accuracy of discourse connective translations: Validation of an automatic metric. In *14th International Conference on Intelligent Text Processing and Computational Linguistics*, page 12. University of the Aegean, Springer, March 2013.
30. Alexandre Heili and Jean-Marc Odobez. Parameter estimation and contextual adaptation for a multi-object tracking crf model. In *IEEE Workshop on Performance Evaluation of Tracking and Surveillance*, 2013.
31. David Imseng and Hervé Bourlard. Speaker adaptive kullback-leibler divergence based hidden markov models. In *Proceedings IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013.
32. David Imseng, Petr Motlicek, Philip N. Garner, and Hervé Bourlard. Impact of deep mlp architecture on different acoustic modeling techniques for under-resourced speech recognition. In *Proceedings of the IEEE workshop on Automatic Speech Recognition and Understanding*, December 2013.
33. Dinesh Babu Jayagopi and Jean-Marc Odobez. Given that, should I respond? contextual addressee estimation in multi-party human-robot interactions. In *Proceedings of Human Robot Interaction (HRI) Conference*, 2013.
34. Dinesh Babu Jayagopi, Samira Sheikhi, David Klotz, Johannes Wienke, Jean-Marc Odobez, Sebastian Wrede, Vasil Khalidov, Laurent Son Nguyen, Britta Wrede, and Daniel Gatica-Perez. The vernissage corpus: a conversational human-robot-interaction dataset. In *Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction*, 2013.
35. Vasil Khalidov, Florence Forbes, and Radu Horaud. Alignment of binocular-binaural data using a moving audio-visual target. In *Proc. of IEEE Int. Workshop on Multimedia Signal Processing (MMSP)*, 2013.
36. Elie Khoury, Paul Gay, and Jean-Marc Odobez. Fusing matching and biometric similarity measures for face diarization in video. In *Proceedings of the 3rd ACM conference on International conference on multimedia retrieval*, pages 97–104. ACM, April 2013.
37. Elie Khoury, Manuel Günther, Laurent El Shafey, and Sébastien Marcel. On the improvements of uni-modal and bi-modal fusions of speaker and face recognition for mobile biometrics. In *Biometric Technologies in Forensic Science*, October 2013.



38. Elie Khoury, Bostjan Vesnicer, Javier Franco-Pedroso, Ricardo Violato, Zenelabidine Boulkenafet, Luis-Miguel Mazaira Fernandez, Mireia Diez, Justina Kosmala, Houssemeddine Khemiri, Tomas Cipr, Rahim Saedi, Manuel Günther, Jerneja Zganec-Gros, Ruben Zazo Candil, Flavio Simões, Messaoud Bengherabi, Augustin Alvarez Marquina, Mikel Penagarikano, Alberto Abad, Mehdi Boulayemen, Petr Schwarz, David Van Leeuwen, Javier Gonzalez-DomiVilda, Amparo Varona, Dijana Petrovska-Delacretaz, Pavel Matejka, Joaquin Gonzalez-Rodriguez, Tiago de Freitas Pereira, Farid Harizi, Luis Javier Rodriguez-Fuentes, Laurent El Shafey, Marcus Angeloni, German Bordel, Gérard Chollet, and Sébastien Marcel. The 2013 speaker recognition evaluation in mobile environment. In *The 6th IAPR International Conference on Biometrics*, June 2013.
39. Jukka Komulainen, Abdenour Hadid, Matti Pietikainen, André Anjos, and Sébastien Marcel. Complementary countermeasures for detecting scenic face spoofing attacks. In *International Conference on Biometrics*, June 2013.
40. Ilya Kuzborskij and Francesco Orabona. Stability and hypothesis transfer learning. In *International Conference on Machine Learning*, June 2013.
41. Ilya Kuzborskij, Francesco Orabona, and Barbara Caputo. From  $n$  to  $n+1$ : Multiclass transfer incremental learning. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, June 2013.
42. Leonidas Lefakis and Francois Fleuret. Reservoir boosting : Between online and offline ensemble learning. In *Proceedings of the international conference on Neural Information Processing Systems*, 2013.
43. Eric Malmi, Trinh-Minh-Tri Do, and Daniel Gatica-Perez. From foursquare to my square: Learning check-in behavior from multiple sources. In *The 7th International AAAI Conference on Weblogs and Social Media*, July 2013.
44. Alvaro Marcos-Ramiro, Daniel Pizarro-Perez, Marta Marron-Romera, Laurent Son Nguyen, and Daniel Gatica-Perez. Body communicative cue extraction for conversational analysis. In *Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition*, April 2013.
45. Jesus Martinez-Gomez, Ismael Garcia-Varea, Miguel Cazorla, and Barbara Caputo. Overview of the imageclef 2013 robot vision task. In *Working Notes, CLEF 2013*, 2013.
46. Thomas Meyer, Cristina Grisot, and Andrei Popescu-Belis. Detecting narrativity to improve english to french translation of simple past verbs. In *Proceedings of the 1st DiscoMT Workshop at ACL 2013 (51st Annual Meeting of the Association for Computational Linguistics)*, pages 33–42, June 2013.
47. Thomas Meyer and Lucie Polakova. Machine translation with many manually labeled discourse connectives. In *Proceedings of the 1st DiscoMT Workshop at ACL 2013 (51st Annual Meeting of the Association for Computational Linguistics)*, pages 43–50, June 2013.
48. Thomas Meyer and Bonnie Webber. Implication of discourse connectives in (machine) translation. In *Proceedings of the 1st DiscoMT Workshop at ACL 2013 (51st Annual Meeting of the Association for Computational Linguistics)*, pages 19–26, June 2013.
49. Gelareh Mohammadi, Sunghyun Park, Kenji Sagae, Alessandro Vinciarelli, and Louis-Philippe Morency. Who is persuasive? the role of perceived personality and communication modality in social multimedia. In *International Conference on Multimodal Interaction*, 2013.
50. Petr Motlicek, Philip N. Garner, Namhoo Kim, and Jeongmi Cho. Accent adaptation using subspace gaussian mixture models. In *The 38th International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 7170–7174. IEEE, May 2013.
51. Petr Motlicek, David Imseng, and Philip N. Garner. Crosslingual tandem-sgmm: Exploiting out-of-language data for acoustic model and feature level adaptation. In *Proceedings of the 14th Annual Conference of the International Speech Communication Association (Interspeech 2013)*, pages 510–514. ISCA - International Speech Communication Association, ISCA, August 2013.
52. Petr Motlicek, Daniel Povey, and Martin Karafiat. Feature and score level combination of subspace gaussians in lvcsr task. In *The 38th International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 7604–7608. IEEE, May 2013.
53. Laurent Son Nguyen, Alvaro Marcos-Ramiro, Marta Marron-Romera, and Daniel Gatica-Perez. Multimodal analysis of body communication cues in employment interviews. In *15th ACM International Conference on Multimodal Interaction Proceedings*, 2013.
54. Youssef Oualil, Mathew Magimai-Doss, Friedrich Faubel, and Dietrich Klakow. A probabilistic framework for multiple speaker localization. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, May 2013.
55. Dimitri Palaz, Ronan Collobert, and Mathew Magimai-Doss. Estimating phoneme class conditional probabilities from raw speech signal using convolutional neural networks. In *Proceedings of Interspeech*, August 2013.





56. Nikolaos Pappas, Georgios Katsimpras, and Efstathios Stamatatos. Distinguishing the popularity between topics: A system for up-to-date opinion retrieval and mining in the web. In *14th International Conference on Intelligent Text Processing and Computational Linguistics*. LNCS, ACM, 2013.
57. Nikolaos Pappas and Andrei Popescu-Belis. Combining content with user preferences for ted lecture recommendation. In *11th International Workshop on Content Based Multimedia Indexing*. IEEE, 2013.
58. Nikolaos Pappas and Andrei Popescu-Belis. Sentiment analysis of user comments for one-class collaborative filtering over ted talks. In *36th ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 2013.
59. Ramya Rasipuram, Peter Bell, and Mathew Magimai-Doss. Grapheme and multilingual posterior features for under-resourced speech recognition: A study on scottish gaelic. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013.
60. Ramya Rasipuram and Mathew Magimai-Doss. Improving grapheme-based asr by probabilistic lexical modeling approach. In *Proceedings of Interspeech*, 2013.
61. Ramya Rasipuram, Marzieh Razavi, and Mathew Magimai-Doss. Probabilistic lexical modeling and unsupervised training for zero-resourced asr. In *Proceedings of the IEEE workshop on Automatic Speech Recognition and Understanding*, December 2013.
62. Edgar Roman-Rangel, Jean-Marc Odobez, and Daniel Gatica-Perez. Evaluating shape descriptors for detection of maya hieroglyphs. In *Proc. Mexican Conf. on Pattern Recognition*, June 2013.
63. Mickael Rouvier, Gregor Dupuy, Paul Gay, Elie Khoury, Teva Merlin, and Sylvain Meignier. An open-source state-of-the-art toolbox for broadcast news diarization. In *INTERSPEECH*, August 2013.
64. Rahim Saedi, Kong Aik Lee, Tomi Kinnunen, Tawfik Hasan, Benoit Fauve, Pierre-Michel Bousquet, Elie Khoury, Pablo Luis Sordo Martinez, Jia Min Karen Kua, Changhuai You, Hanwu Sun, Anthony Larcher, Padmanabhan Rajan, Ville Hautamaki, Cemal Hanilci, Billy Braithwaite, Gonzalez-Hautamaki Rosa, Seyed Omid Sadjadi, Gang Liu, Hynek Boril, Navid Shokouhi, Driss Matrouf, Laurent El Shafey, Pejman Mowlaei, Julien Epps, Tharmarajah Thiruvaran, David Van Leeuwen, Bin Ma, Haizhou Li, John Hansen, Jean-François Bonastre, Sébastien Marcel, John Mason, and Eliathamby Ambikairajah. I4u submission to nist sre 2012: a large-scale collaborative effort for noise-robust speaker verification. In *INTERSPEECH*, August 2013.
65. Lakshmi Saheer and Milos Cernak. Automatic staging of audio with emotions. In *International Conference on Affective Computing and Intelligent Interaction*, 2013.
66. Lakshmi Saheer and Blaise Potard. Understanding factors in emotion perception. In *ISCA Speech Synthesis Workshop*, 2013.
67. Albert Ali Salah, Hayley Hung, Oya Aran, and Hatice Gunes. Creative applications of human behavior understanding. hbu 2013: 1-14. In *Human Behavior Understanding*, pages 1–14, 2013.
68. Dairazalia Sanchez-Cortes, Joan-Isaac Biel, Shiro Kumano, Junji Yamato, Kazuhiro Otsuka, and Daniel Gatica-Perez. Inferring mood in ubiquitous conversational video. In *12th International Conference on Mobile and Ubiquitous Multimedia*. ACM Press, December 2013.
69. Darshan Santani and Daniel Gatica-Perez. Revisiting the generality of the rank-based human mobility model. In *Proceedings of 3rd International Workshop on Pervasive Urban Applications (PURBA)*, 2013.
70. Darshan Santani and Daniel Gatica-Perez. Speaking swiss: Languages and venues in foursquare. In *Proceedings of the 21st ACM International Conference on Multimedia*, 2013.
71. A. Sapru and Hervé Bourlard. Automatic social role recognition in professional meetings using conditional random fields. In *Proceedings of Interspeech*, 2013.
72. A. Sapru and Hervé Bourlard. Investigating the impact of language style and vocal expression on social roles of participants in professional meetings. In *Affective Computing and Intelligent Interaction*, page 6, September 2013.
73. Samira Sheikhi, Dinesh Babu Jayagopi, Vasil Khalidov, and Jean-Marc Odobez. Context aware addressee estimation for human robot interaction. In *Proceedings of the 6th Workshop on Eye Gaze in Intelligent Human Machine Interaction: Gaze in Multimodal Interaction*, 2013.
74. Samira Sheikhi, Vasil Khalidov, David Klotz, Britta Wrede, and Jean-Marc Odobez. Leveraging the robot dialog state for visual focus of attention recognition. In *Proceedings of the 15th ACM on International Conference on Multimodal Interaction*, 2013.
75. Gyorgy Szaszak and Andras Beke. Using phonological phrase segmentation to improve automatic keyword spotting for the highly agglutinating hungarian language. In *Proc. of Interspeech*, 2013.



76. Gyorgy Szaszak and Philip N. Garner. Evaluating intra- and crosslingual adaptation for non-native speech recognition in a bilingual environment. In *Proceedings of the 4th IEEE International Conference on Cognitive Infocommunications*, pages 357–361. IEEE, December 2013.
77. Raphael Sznitman, Carlos Becker, Francois Fleuret, and Pascal Fua. Fast object detection with entropy-driven evaluation. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, 2013.
78. Mohammad J. Taghizadeh, Reza Parhizkar, Philip N. Garner, and Hervé Bourlard. Euclidean distance matrix completion for ad-hoc microphone array calibration. In *Proceedings IEEE International Conference On Digital Signal Processing*, July 2013.
79. Romain Tavenard, Remi Emonet, and Jean-Marc Odobez. Time-sensitive topic models for action recognition in videos. In *IEEE International Conference on Image Processing*, 2013.
80. Raphael Ullmann, Hervé Bourlard, Jens Berger, and Anna Llagostera Casanovas. Noise intrusiveness factors in speech telecommunications. In *Proceedings of the AIA-DAGA 2013 International Conference on Acoustics*, pages 436–439, March 2013.
81. Majid Yazdani, Ronan Collobert, and Andrei Popescu-Belis. Learning to rank on network data. In *Mining and Learning with Graphs*, 2013.
82. Majid Yazdani and Andrei Popescu-Belis. Computing text semantic relatedness using the contents and links of a hyper-text encyclopedia. In *International Joint Conference on artificial intelligence*, 2013.
83. Sree Harsha Yella and Hervé Bourlard. Improved overlap speech diarization of meeting recordings using long-term conversational features. In *ICASSP*, 2013.
84. Sree Harsha Yella and Hervé Bourlard. Improved overlap speech diarization of meeting recordings using long-term conversational features. In *ICASSP*, 2013.







Centre du Parc, rue Marconi 19, PO Box 592, CH-1920 Martigny  
T +41 27 721 77 11 F +41 27 721 77 12 [info@idiap.ch](mailto:info@idiap.ch) [www.idiap.ch](http://www.idiap.ch)

