



A N N U A L

R E P O R T

2 0 1 2



Publication details

Production and design: Céline Aymon Fournier, Relations publiques, Idiap

Drafting: Le fin mot... Communication, Fully

Translation: Dave Brooks

Graphic design: Atelier Grand, Sierre

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

Printing: calligraphy.ch, Sierre

Print run: 1,800 copies

CONTENTS

Messages

“Idiap contributes to strengthening the economy of Valais” Olivier Dumas, President of the Foundation Council of Idiap	2
“We are more than ever aligned to society’s current concerns” Hervé Boulard, Director of Idiap	3

Research

Profile	5
Selected Research Activities	
Mobile Data Challenge: The Smartphone for Quality of Life	9
Mediaparl: Parliament As If You Were There	10
Armasuisse Project: To Communicate Securely	12
Ninapro Project: New Generation of Artificial Hand	13

Network

Industrial Partners: The Trust of Giants	15
International Create Challenge: From Idea to Start-up in Three Weeks	16
Start-up News: Growth and New Commitments	18

Faces

Lakshmi Saheer: Idiap’s First female Indian PhD	21
The Project Management Team: The Research Managers	22
Antoine Dorsaz, Financial Assistant : Passion Pays	24
Joining and Leaving	25
Distinctions	26
Theses Completed	27

Finances

Operating Accounts	29
Sources of Funds / Costs / Comments	30
Balance Sheet	31

Organization

Organizational Chart	33
Employees	34
Foundation Council	36
Advisory Board	38
Main Partners	39

Scientific Inserts

Idiap Research Areas: Human and Media Computing	II
Scientific Progress Report	IV
Main projects in progress	XVIII
Major publications / Conferences	XXVI

A WORD FROM THE PRESIDENT

“IDIAP CONTRIBUTES TO STRENGTHENING THE ECONOMY OF VALAIS”



Olivier Dumas, President of the Foundation Council, Idiap

Twenty years, and then... One could be forgiven for thinking that, after the festivities of Idiap's 20th anniversary in 2011, and the satisfaction of seeing the “little institute” grow large while retaining its creativity and conviviality, a slight loss of impetus would occur. Not at all.

Mediaparl, a partnership with the regional TV channel Canal 9 and the Valais Parliamentary Service, was one of the major results of 2012. The project placed Idiap's world-class competencies in automatic speech recognition at the service of the canton. Two years of research, supported by the Loterie Romande and the Valais legislature, enabled the institute to develop its expertise in handling French and German, and to provide Valais's parliament with a tool to archive and index videos of its sessions; a platform which several other cantonal parliaments, and even the federal government in Bern, now wish to adopt (see pages 10-11).

By further developing its mission of technology transfer, particularly through its collaboration with The Ark Foundation's site IdeArk, Idiap also contributed to strengthening the economy of Valais. Proof of this is the growth of its principal start-ups which all, in 2012, expanded their ranks, confirming their dynamism; a quality crowned by numerous distinctions (see pages 18-19.)

We should also note the changes to Idiap's Board of Trustees. I would like to take this opportunity to thank the outgoing members for their invaluable contributions, and to welcome Messrs. Berclaz, Furrer, and Montserrat, representing—respectively—the HES-SO Valais Wallis and future EPFL pole Valais Wallis, Euresearch, and venturelab / venture kick. We also welcome two new representatives of EPFL, Mrs. Adrienne Corboud Fumagalli and Professor Philippe Gillet—respectively vice president for innovation & technology transfer and vice president for academic affairs. Let us wager that these newcomers will offer our institute new perspectives, and that their vision will enable us, tomorrow, to further intensify our collaboration with other actors on the regional, national, and international scientific stage (see pages 36-37).

Finally, 2012 also witnessed, in January, the signing of the agreement between the cantonal executive and EPFL to create, in Valais, a center active in the fields of energy, chemistry, and health, by the creation of eleven research chairs, and educational programs at master's level. This project should be implemented in 2015 and, given Idiap's strong history of collaboration with EPFL, our institute should have a key role to play. “EPFL Valais Wallis” will, without a doubt, by establishing close links between research institutions and The Ark Foundation, make possible the expansion of training opportunities in Valais, and the strengthening of research activities, but also offer *Valaisans* and *Valaisannes* alike new opportunities in their own canton.

A handwritten signature in blue ink, appearing to be 'O. Dumas', written in a cursive style.

A MESSAGE FROM THE DIRECTOR

“WE ARE MORE THAN EVER ALIGNED TO SOCIETY’S CURRENT CONCERNS”



Prof. Hervé Boulard, Director, Idiap

Since its inception in 1991, Idiap has been guided by the spirit of its founder, the Italian industrialist Dalle Molle: Putting the machine at the service of humanity. The idea may seem passé. Yet at a time when information technology pervades every part of our daily lives, such a quest has never made more sense. And Idiap has not forgotten, despite its extraordinary growth, a move into spacious new premises at the Centre du Parc, and a strengthening of its position on the world stage. Quite the contrary in fact.

In 2012, our institute chose to further anchor its research in current realities by diversifying and enhancing its core competencies in application domains aligned to our society’s present concerns.

Thus, no fewer than ten application domains now benefit from the expertise of our top-level scientists (see page 7). Energy, health, safety, and human-human and human-machine interaction are just some of the areas in which Idiap conducts, with great dynamism, national and European research projects, affirming with more clarity and visibility than ever before the role it intends to play in societal development. An endeavor which appears to be bearing fruit: the magazine Bilan having, in June 2012, judged two of our scientists to be among the three hundred most influential people in Switzerland. Sébastien Marcel, one of our senior researchers, was recognized for his international biometrics activities, establishing Idiap as a key player in this field in Europe. I was also mentioned for having made Idiap “one of the leading global players in man-machine interfaces, with promising start-ups such as Koemei and Klewel” (see pages 18-19).

Idiap seeks and Idiap finds; a fact which pleases me greatly. I share the acclaim for this positive track record with the three senior scientists who, along with me, comprise the institute’s Scientific College, and with our new Deputy Director, François Foglia. Enterprising and independent, they do an extraordinary job. The numbers also speak for themselves, by contributing to forty-two projects in 2012, our institute set a new record.

These research operations deserved better support, and this is now provided by the new Project Management team (see pages 22-23). Three new staff members now help researchers carry out their administrative duties, that prove to be increasingly taxing and with financial flow that need to be clearly identified.

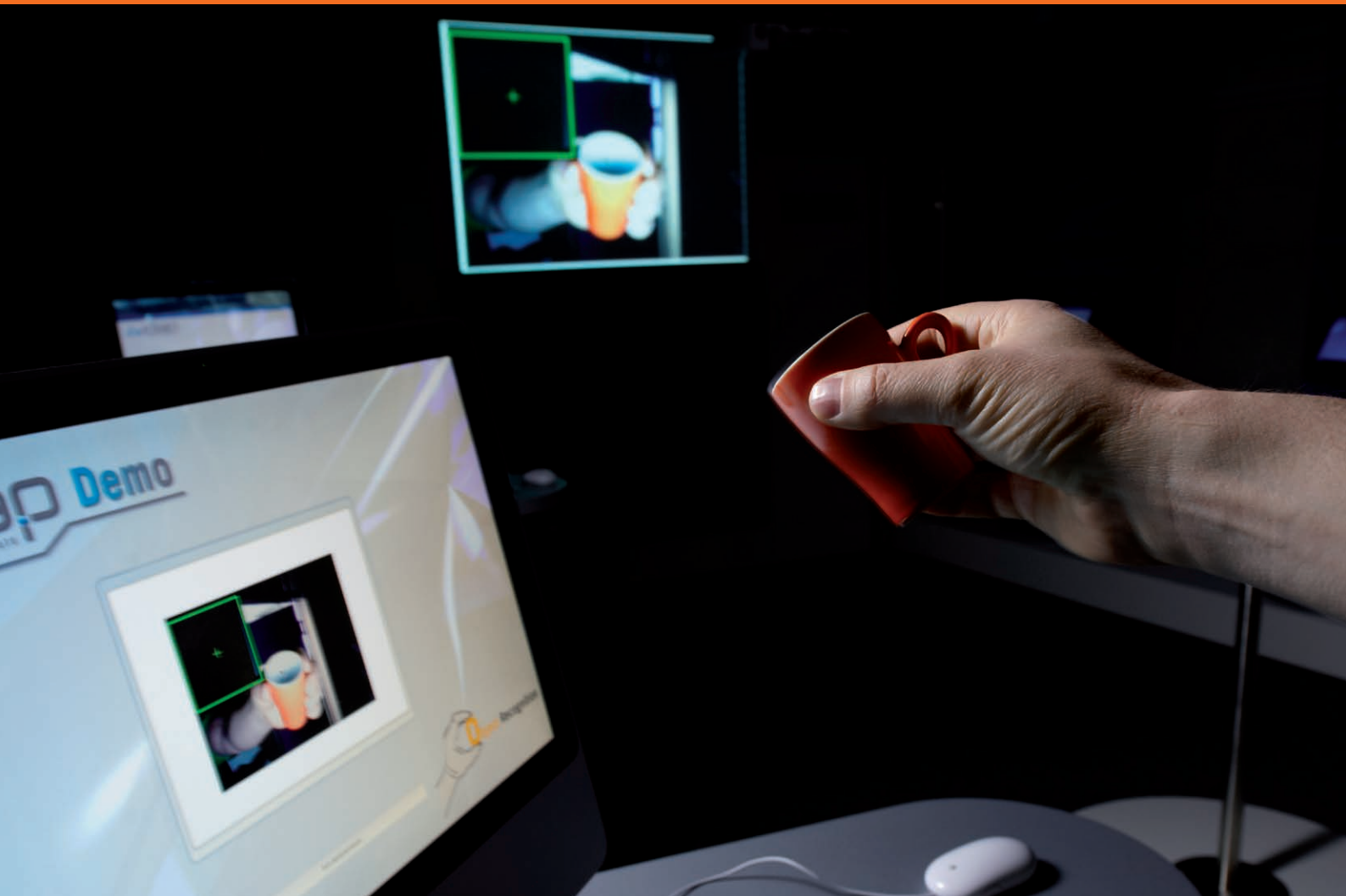
I would like to take this opportunity to wish a warm welcome and future success to the initiative “EPFL Valais Wallis”. It provides the opportunity for all institutions associated with EPFL since its inception, Idiap included, to develop even further. Our institute will certainly have a key role to play, a task which will doubtless be on our agenda throughout 2013.

Finally, I congratulate all Idiap’s employees, more than one hundred in 2012, on their commitment. Researchers, PhD students, developers, administrators; all contributed to the institute’s success. I wish them, I wish us all, a bold and successful 2013, and all in the spirit of collaboration and the high standards of quality which have always inspired Idiap.

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



RESEARCH



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

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

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

A Profusion of Technological Products

The beginning of the 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 budget – CHF 10 million – 60 percent is financed by competitively awarded research projects, and 40 percent from public funds (see "Distribution of sources of financing", page 30.)

Team / As of 2012, 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 2,600 m² 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 are research assistants.

■ 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.

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

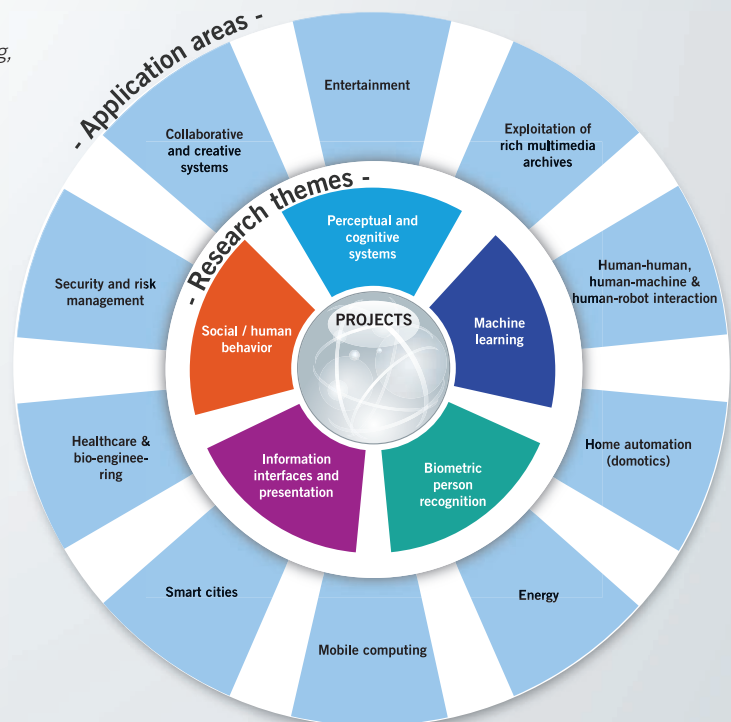
(multimedia information systems, personalization and contextualization)

4 Biometric person recognition

(speaker recognition, face recognition)

5 Machine learning

(statistical modeling, neural networks, and mathematical models)



Idiap in Figures (année 2012)

Human Resources

- 1 professor
- 3 senior scientists
- 9 permanent senior researchers
- 18 postdocs
- 34 research assistants
- 16 system engineers and development engineers
- 11 interns and visitors (average per year)
- 10 administrative staff
- 7 doctorates awarded
- 42 posts in IdeArk start-ups
- 31 nationalities represented

Scientific Activities

- National Center for Competence in Research IM2 (Interactive Multimodal Information Management) since 2001
- Participation in forty-two research programs
- Project management in six consortia
- Participation in the economic development strategy of the Canton of Valais through the program – The Ark and, in particular, the company IdeArk
- Two hundred and forty 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, Ildiap 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.

Speech & Audio Processing

Prof. Hervé Bourlard, Dr. John Dines, Dr. Phil Garner, Dr. Mathew Magimai-Doss, Dr. Petr Motlicek

Speech processing has been one of the mainstays of Idiap's research portfolio for many years. Today it is still the largest group within the institute, and Idiap continues to be recognised as a leading proponent in the field. The expertise of the group encompasses statistical automatic speech recognition (based on hidden Markov models, or hybrid systems exploiting connectionist approaches), text-to-speech, and generic audio processing (covering sound source localization, microphone arrays, speaker diarization, audio indexing, very low bit-rate speech coding, and perceptual background noise analysis for telecommunication systems).

Computer Vision & Learning

Dr. François Fleuret

The goal of our group is the development of new statistical learning techniques mainly for computer vision, with a particular interest in their computational properties. Our application domains include object detection and scene analysis, the tracking of people and biological structures, and image recognition 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 everyday 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, and urban data-mining using smartphones and location-based social networks.

Perception & Activity Understanding

Dr. Jean-Marc Odobez

The research group on the understanding of perception and activity conducts research analyses of human activities from multimodal data. This includes investigating the fundamental tasks of scene analysis such as detection, segmentation and tracking of people, their representation, and the characterization of their condition, as well as the modeling of sequential data and their interpretation in the form of gestures, activities, behavior, or social relationships, through the design of sound algorithms which exploit and extend models and methods of computer vision, machine learning, and multimodal data-fusion. Surveillance, traffic analysis, analysis of behavior, human-robot interfaces, and multimedia content analysis are the main application domains.

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 the ability to autonomously detect knowledge gaps and fill them autonomously with open-ended learning strategies. The focus of this work is on designing algorithms that are principled, computationally efficient, and that provide robust performance in very realistic settings while at the same time providing theoretical guarantees on expected behavior. The group is funded by national and international funding agencies.

Applied Machine Learning

Dr. Ronan Collobert

This group is interested in computer algorithms which can "learn" a behavior in order to achieve a given task of interest, in contrast to algorithms with behavior constrained by hand-crafted 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), as well as on unsupervised learning techniques which can leverage inherent semantics from large-scale, structured data.

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 (presentation attack detection), and emerging biometric modes (EEG and vascular). The group is geared toward reproducible research and technology transfer, using its own signal-processing and machine-learning toolbox.

Natural Language Processing

Dr. Andrei Popescu-Belis

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

THE SMARTPHONE FOR QUALITY OF LIFE

In 2012, Idiap and Nokia Research Center organized the Mobile Data Challenge, a call to the global research community to better understand human and social behavior in mobile phone use.

From 2009 to 2011, Idiap led a major project with Nokia. Entitled LS-CONTEXT, the project's primary aim was to establish a database of information sourced from mobile phones; data more difficult than usual to assemble as they affect the private sphere (see box). In 2012, this adventure reached another important milestone with the Mobile Data Challenge (MDC), a worldwide research initiative.

Daniel Gatica-Perez, senior researcher and head of Idiap's Social Computing group is the co-initiator of this vast project. "The cell phone is an excellent tool for the study of social behavior: It's always with us, we use it to call or send text messages, but also to stay connected to our virtual networks, inform ourselves, or organize our leisure or work time."

One Hundred and Eight Research Teams Worldwide

To give such wealth of data the top-quality analysis it deserves, Idiap and Nokia Research Lausanne organized a large-scale

research project. In early January 2012, the data were sent to over five hundred individual participants, in the form of labeled and numbered (watermarked) copies. In spring 2012, one hundred and eight teams of scientists from around the world, including some from the cream of academic and research institutions (Carnegie Mellon University, University of Illinois at Urbana-Champaign, CERN, EPFL, ETHZ, and TU Berlin) submitted their results. By formal agreement, all had of course previously committed to processing the data in a completely ethical manner, solely for research purposes.

Each team then chose one of two options: Dedicated Track or Open Track. Those who chose Dedicated Track would aim to predict cell-phone users' mobility, to estimate their demographic profile, or to determine in which type of place users were located (school, restaurant, park, cinema, etc.). For example: Bluetooth devices equidistant and motionless, on a Saturday at 8.00 pm... the user could be in a movie theater.

"The data collected will one day enable us to make life easier for cell-phone users", explains Gatica-Perez. Take, for example, the case of Mr. X, who always goes to the train station after work. His phone could understand that the train's departure time is approaching, and tell Mr. X it's time to leave the office, or that he shouldn't forget his keys.

The second option provided a different kind of challenge to the creativity of the scientists involved. The teams who chose Open Track were free to propose their own tasks; for example, to study the relationship between mobility and weather conditions. This option produced a significant number of new findings. Winners were invited to the Mobile Data Challenge Workshop in June 2012 in Newcastle, UK.



An Exceptional Database

For between twelve and eighteen months, nearly two hundred volunteers from the French-speaking part of Switzerland used a smartphone that recorded their movements, speed, and proximity to other phones, but also information about – among other things – device usage, frequency of calls, and camera or MP3 player use.

"So many people involved over such a long time, and so many types of data collected – it's unique for science", explains Gatica-Perez. "All this allowed us to open up huge fields of research." Despite this potential, the ethical dimension was, of course, never overlooked. "We don't know the content of calls or text messages, only in what circumstances these features were used."



PARLIAMENT AS IF YOU WERE THERE!

With the project Mediaparl, Idiap is enabling the Canton of Valais's Members of Parliament (MPs), and the general public, to find the content of any parliamentary motion, resolution, or debate, on video – a pioneering tool which is also attracting the interest of the federal government in Bern.

“We can't talk about the Mediaparl project without talking about the concept of a paperless parliament. The two are interconnected”, explains immediately Alexandre Nanchen, an Idiap development engineer since 2008 and the project's manager. In 2010, based particularly on the observation that each MP uses almost thirty-three kilograms of paper per year, Jean-Albert Ferrez, then vice president of the Valais legislature and deputy director of Idiap, launched the idea of a parliament equipped with tablet computers. At the same time, the legislature's document production since 1851 – a total of 213,000 pages – was digitized to allow easier access to archive materials.

A System Capable of Recognizing French and German

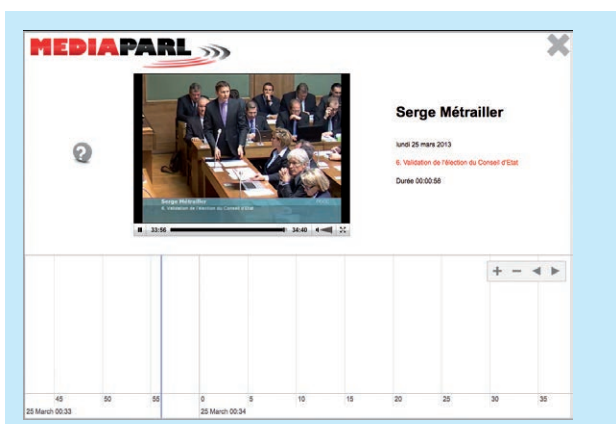
Once this first step was achieved, it was only natural to raise the question of television archives and of the best way to use them, Canal 9 having recorded all legislative sessions since 2007. “Idiap is one of the best research institutes in the world in the field of automatic speech recognition”, explains Nanchen, “so we wanted to set up a project to index all these audiovisual archives.” Just one snag: all automatic speech recognition engines developed at Idiap up to that point were in English, and here it was a question of designing a tool that works with both French and German.

Ferrez and Yann Rodriguez, Idiap's head of technology transfer, designed the project. Supported by the Loterie Romande and the parliament of Valais, Mediaparl began in January 2012 and should conclude early in 2014. The project envisages the parallel development of two tools. One will be an indexing tool that automatically extracts all the words in a stream of speech, indexes them, and then enables users to find any video sequence with a simple keyword search, just as you would with a text. The second will be an automatic transcription system: “Today, the Parliamentary Service needs six weeks to get the contents of a legislative session down on paper. Such a system would save valuable time, even if a ‘human’ rereading will still be necessary.”

A Sample of MPs' Speech

At the heart of both tools is a software capable of identifying a stream of speech. Nanchen explains: “To develop a system that's efficient from the outset, it's necessary to provide it with representative examples of the type of data that it will later have to handle. So, to begin with, we matched the words actually spoken by MPs with the text of these speeches by using recordings and transcripts of previous legislative sessions; a total of twenty hours for each language.” This made possible the “calibration” of the system; doubly important since Valais's parliament has its peculiarities. “Accents are very different from one region to the next, which doesn't make speech recognition by computer any easier.” What's more, in the Valais legislature, 10 percent of discussion takes place in German, and there is nothing to warn the system that the language being spoken has changed. “The next step was the integration of concepts of sentence structure and phonetics, enabling the system to learn French and German.” Of particular importance were the particularities of political syntax and vocabulary.

A few months before the prototype's launch, Nanchen is keen to underline that the project truly was a team effort. “Phil Garner, Idiap senior researcher; David Imseng, research assistant; and in particular the development engineers created the French and German speech recognition systems, and the indexing tools. Even students worked during the summer on this project.” Finally, Mediaparl is a partnership between Idiap and the Valais Parliamentary Service, particularly Daniel Petitjean, the service's IT manager, and the regional television channel Canal 9.



Via a simple search form, Mediaparl will enable MPs and members of the public to search all parliamentary debates that have taken place since 2009. A media player will allow users to navigate within the sequence in question. Idiap is one of the premier institutes in the world in the domain of automatic speech recognition,

The Parliamentary Service, which employs around ten people, provides legal and scientific support to Valais's MPs and carries out a range of administrative duties. Mediaparl is enabling it to further improve these services.

Interview with Claude Bumann, head of the Parliamentary Service.

Why did you want to participate in the Mediaparl project?

Valais's parliament has always played a pioneering role compared to other cantonal parliaments. We were the first to be televised, our website aims to be a powerful tool, and our team is constantly striving to improve the services it offers to MPs. Mediaparl fits perfectly into this framework.

It all started with the idea of a paperless parliament, launched under the chairmanship of Paul-André Roux. In 2010, when he was vice president of parliament and deputy director of Idiap, Jean-Albert Ferrez sped the project up. For the Parliamentary Service's annual outing, he invited us to Idiap and showed us work then in progress in the field of speech recognition, suggesting the possibility of a parliamentary application. When I saw the system, entirely in English and still in need of improvement, I told myself it would take another ten years before it would be operational. Then the Loterie Romande and the canton began to support Mediaparl, and when we returned to Idiap two years later, the progress made was impressive.

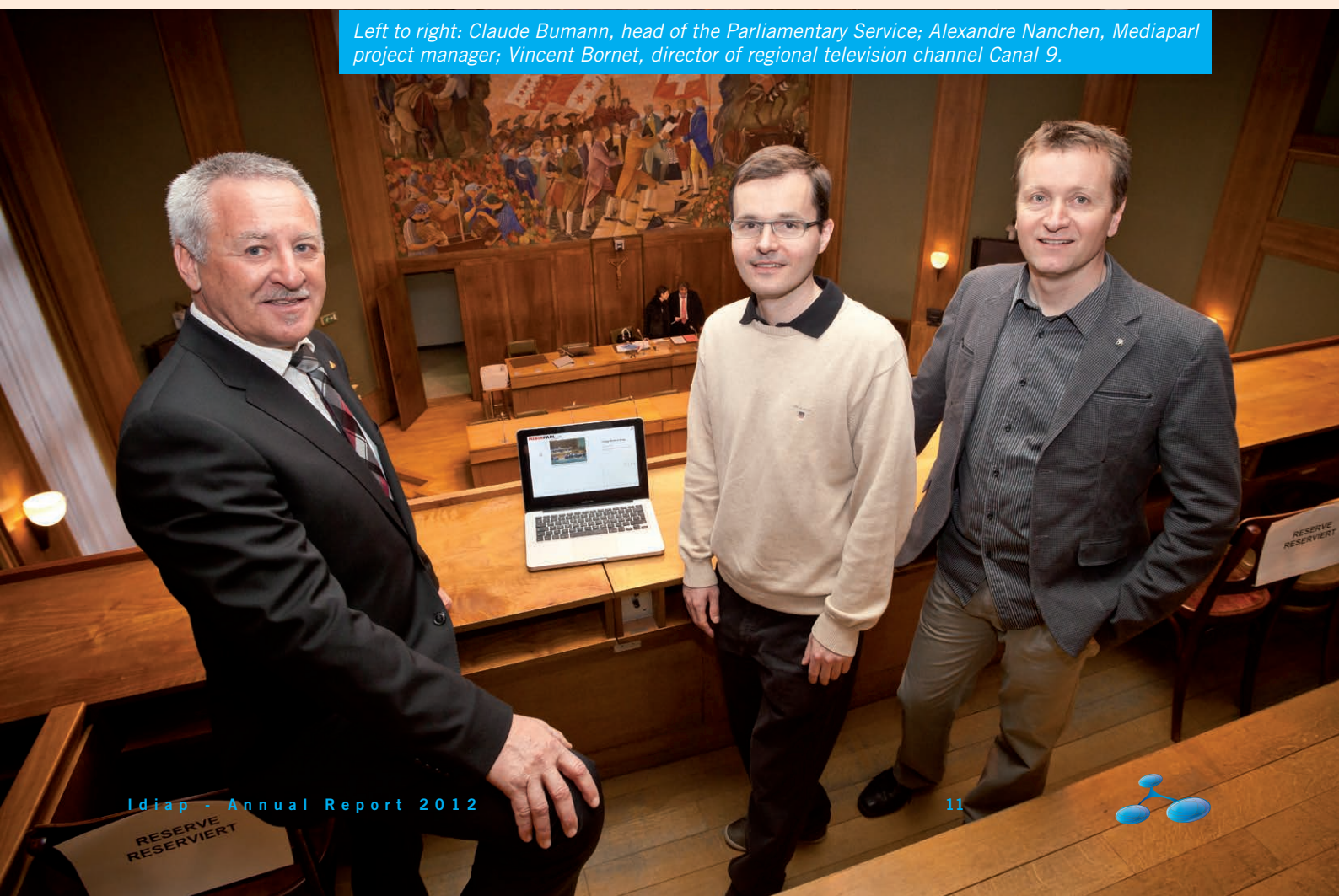
Mediaparl should become operational in spring 2013. What are the benefits?

Today, even if we have an internet search engine, it's still tedious to find a particular statement made during any given session. First you need to know the date of the session, the exact date, then watch the Canal 9 archives of the whole session. With Mediaparl, you'll be able to quickly and easily find any given statement or speech, just with a keyword. It's a real revolution. Incidentally, when the press started to speak about the project, our peers, including the Canton of Vaud's parliament and the Federal parliament, began to express an interest. If use of the tool becomes widespread, Mediaparl will make possible a quite precious exchange of ideas and information between cantons.

The second phase of Mediaparl will transcribe legislative sessions. Will the machine replace man?

I don't think so. Françoise Manni (who currently performs data entry) will save time, and this time can then be reinvested to improve our services for MPs. We still have lots of ideas about how to do this, so there's clear room for improvement.

Left to right: Claude Bumann, head of the Parliamentary Service; Alexandre Nanchen, Mediaparl project manager; Vincent Bornet, director of regional television channel Canal 9.



ARMASUISSE PROJECT

TO COMMUNICATE SECURELY

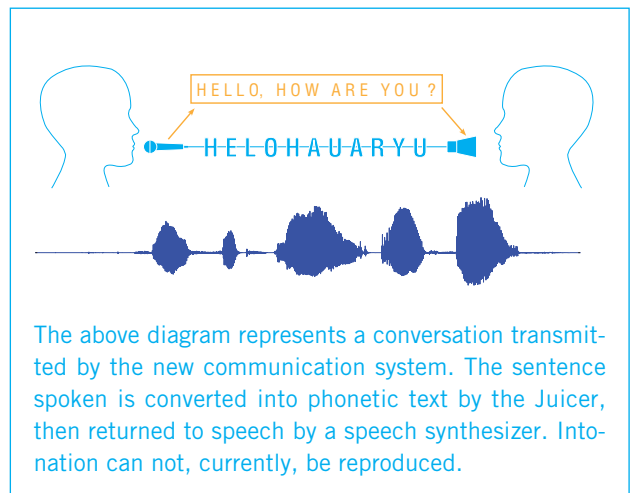
Idiap has been working for two years on a research project on behalf of armasuisse. Objective: to facilitate and protect “sensitive” communications.

Around two years ago, armasuisse* contacted Idiap. Its Networks and Protection department was looking for innovative solutions to facilitate and protect sensitive communications in adverse environments. Voice synthesis, one of Idiap’s core functions, rapidly became a priority feature of the mission entrusted to Philip Garner, senior researcher. Garner works on the project team in collaboration with the researcher Petr Motlicek, and development engineer Milos Cernak.

Reduce Throughput by Sending Text Rather Than Sound

A practical example will help illustrate what is at stake. In critical situations, very long distance communication takes place using high-frequency (HF) radio. “To ensure the robustness and reliability of exchanges”, Garner explains. “We need to minimize bandwidth, reserving a significant proportion of it for the repetition of information, encoded or not; hence the need to greatly reduce the amount of data to be transmitted. In short, we need to ensure that even if interference or other transmission problems occur, the content is still sufficiently clear for the recipient to understand it.”

This is where voice recognition comes in. Rather than transmitting actual speech, it transmits text. This transformation takes place thanks to the Juicer, a device developed at Idiap. Upon reception, a speech synthesizer, the kind of device used by the British physicist, Stephen Hawking renders the transmitted text as speech. The system can operate, either between military radios or via satellite. “For the moment, this technique can reproduce neither the emotions conveyed by the human voice nor its intensity”, says Garner, “but we’re working on it.”



As for throughput, it falls, thanks to this technique, from several thousand bits per second (the throughput of a normal conversation), to around 100-200 bits per second, and communication is thus both further facilitated and secured. “This low throughput rate makes communication possible in a hostile environment, despite ambient noise interference or attempts at jamming.”

Particularities of the Swiss Situation

Armasuisse, naturally, remains discreet regarding the engagement scenarios of this future communication system. Garner comments, “We understand this approach perfectly, and it hasn’t hindered our research in any way.”

Last feature of the project: the device will, given the linguistic particularities of Switzerland, operate in multiple languages. While English remains the reference language of research and database work, at Idiap as at any other research institute of global standing, “in the last few years, we’ve been trying to change this, and use French, German, and German variants such as ‘Walliserdeutsch’ too. It is a nice challenge, and here in Martigny we’re in an ideal position to meet it.”

References

With a master’s degree in Electrical Engineering from the University of Southampton, Philip N. Garner joined Idiap in April 2007. The 43-year-old Briton, specialist in speech processing, had previously worked in industry, notably for the firm Canon in Surrey (UK) and in Tokyo. In 2012, he was awarded his doctorate by the University of East Anglia (UK).



* Armasuisse is a center of excellence of the Swiss Confederation. It is responsible, particularly, for the acquisition of technologically complex systems and equipment related to the security domain.

NEW GENERATION OF ARTIFICIAL HANDS

By combining advances in mechatronics and micro-electricity with Idiap algorithms, the project Ninapro is developing a new prosthetic hand, more adroit and easier to master.



For the data collection phase, Idiap worked with a "CyberGlove", a glove equipped with sensors that recorded the daily movements of the nondisabled participants.



The CyberGlove was developed to record and reproduce the daily movements of the hand.

One can only imagine the daily difficulties faced by someone who has to learn to live without one of their hands. Washing, dressing, driving, cooking, every action requires external assistance and new learning. In most cases, the remaining hand needs to compensate for the loss of the amputated hand, as current surface-electromyography prosthetics (controlled by an electrical signal from the muscles of the stump) are mainly limited to the opening and closing of the prosthetic hand, with very little control. Mastering such actions takes months of effort and often discouraging, repetitive training.

To capture the movements of a nondisabled hand, electrodes are placed on the muscles concerned. By wearing the CyberGlove, testers provide data which will later allow the prosthetic hand alone perform the task in question. ✓

How then can artificial hands be improved? Faced with this question from a colleague from the German Aerospace Center DLR's Institute of Robotics and Mechatronics, Idiap senior researcher Barbara Caputo decided to take up the challenge: "I immediately understood that our research in the field of intelligent household robots was applicable here. I was also motivated by the human stakes of the project, the prospect of a quick and fun rehabilitation phase for patients, giving them more autonomy into the bargain." The challenge: to provide a prosthesis that has all the advantages, but none of the disadvantages, of an implant.



Mme Barbara Caputo

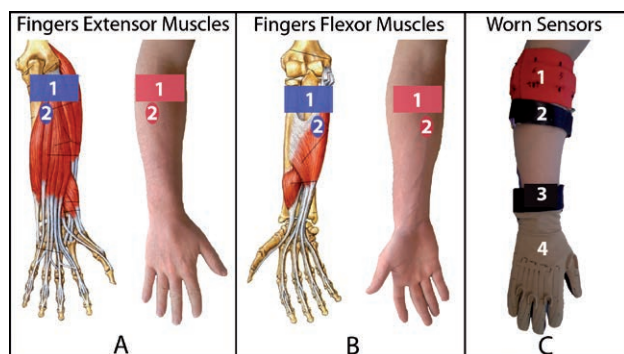
50 Nondisabled Volunteers Model Their Movements

Launched in 2011, NinaPro ("Non Invasive Adaptive Hand Prosthetics") builds on recent advances in mechatronics and micro-electronics, and Idiap's know-how in visual recognition, to increase the dexterity of a prosthetic hand.

First stage: collect and analyze data. Around fifty nondisabled participants took part, helping to identify most frequently made movements. "The more models we have, the greater our chances of success, because the range of actions depends on a person's physical characteristics (age, size, strength, etc.)," says Caputo, who hopes to involve 1,000 volunteers in the project. Tests with amputees will follow. The project will then seek funding for a further three years of research from 2013 to 2016.

Attracting Financing with Convincing Demonstrations

To raise Ninapro's profile, Idiap and its partners (DLR and HES-SO) are preparing a presentation video and several interactive demonstrations. Caputo expands on this: "The eventual capabilities of the prostheses will depend on the needs of amputees themselves, but also on the prosthesis industry and its level of financial engagement. We have the technology to produce prostheses that are either tailor-made or low-cost."



NETWORK



THE TRUST OF GIANTS

For several years now, Idiap has enjoyed close relations with Asian and European industrial giants in the fields of mobile telephony and security – a calling card for the institute’s technology transfer activities. Overview.

THALES

Video Surveillance of Public Places

In 2011, the French group Thales merged two of its subsidiaries, creating Thales Communications & Security. The new company should enable Thales to consolidate its position as the European leader in information systems and secure communications for the global defense, security, and ground transportation markets.

Based in Gennevilliers (Hauts-de-Seine), Thales Communications & Security employs 7,000 people at nine sites in France and invests 20 percent of its turnover in research and development. In 2012, the company – Idiap’s industrial partner in a European project launched two years earlier – gave Idiap a research mandate in the field of video surveillance and detection of left baggage. At a time when the number of surveillance cameras in public spaces is increasing, without necessarily a parallel increase in the human resources needed to view the images captured, Idiap’s expertise in automatic image recognition is invaluable.

NOKIA

Connecting People

Working Together for the Telephony of Tomorrow

For Nokia and Idiap, 2012 was a special year. Their collaboration reached an important milestone with their joint organization of the Mobile Data Challenge (see page 9), a worldwide research initiative.

Since 2009, Nokia and Idiap have been working together on a project to collect data from two hundred volunteers in western Switzerland equipped with smartphones that record their behavior. Now complete, this database – its contents rendered anonymous and watermarked to ensure its protection and traceability – is a fabulous research tool, and Idiap has been given exclusive rights for its distribution to the scientific community.



Social Interaction in Teamwork

The NTT (Nippon Telegraph Telecom) Communication Science Laboratories are working with Idiap on the analysis of behavior, especially non-verbal interactions, in the context of teamwork (the quality of which is of great importance to industry). Using probabilistic methods, the project NISHA aims to identify these behaviors, analyze their properties, and use this knowledge to characterize the key aspects of interpersonal relationships.



Idiap – the Only Swiss Institution Selected

Number one in cell-phone sales in Europe, each year Samsung launches calls for projects from the worldwide research community. In 2011, as the only Swiss institution selected by the South Korean group, Idiap was granted USD 100,000 in funds and given the task of developing a new method of voice recognition. This collaboration continued in 2012, Idiap receiving a new research mandate in the field of automatic multilingual voice recognition – a genuine acknowledgment of the institute’s competencies, and a source of pride for Idiap.



Help Machines to Understand the Meaning of a Text

The Japanese group NEC (Nippon Electronic Company), with two laboratories in the United States, is currently funding a PhD at Idiap. The project goal – an extremely complex task that begins on a small, simple scale – is to help machines understand the structure and, by implication, the meaning of a text.



INTERNATIONAL CREATE CHALLENGE



FROM IDEA TO START-UP IN THREE WEEKS

In 2012, Idiap launched a new initiative – the International Create Challenge (ICC). Its supportive environment offers a unique opportunity for young researchers to become entrepreneurs. With new companies being created and research projects being negotiated, the ICC's first edition was a huge success.

For a first shot at a new competition, it was a bull's-eye: seventy submissions; twenty-eight researchers selected and then assembled into teams; twelve projects presented, and six prize winners; several startups created and numerous research projects negotiated. The first edition of the ICC, which took place from 1-21 September, 2012, surpassed our expectations.

The ICC is an initiative of the Idiap Research Institute, the National Center of Competence in Research (NCCR) IM2 and IdeArk. The competition offers young researchers a unique opportunity coupled with a challenge: to go from idea to start-up in three weeks. François Foglia, deputy director of Idiap and the ICC's creator explains the idea that led to the competition's launch: "Researchers are often overwhelmed by their workload or their scientific publications, to such an extent that they don't take the time to develop their ideas for commercial products. With this initiative, we create the right environment for them to step, for a moment, out of their daily lives, and devote themselves solely to their project, in the hope of producing a prototype at the end of the event."

Three Weeks of Immersion, Under Specialist Supervision

The right environment? A three-week, all-expenses-paid stay at Idiap, in the Centre du Parc, Martigny. An immersion program during which participants receive expert advice on starting a business, and are oriented in the areas of fund-raising, project management, business models, and intellectual property, all while discovering the art of presenting their project in ninety seconds.

Through the International Create Challenge, Idiap aims to promote technology transfer and the creation of new start-ups and, with them, new jobs. "This is part of our research institute's mission. The event also enables us to give a certain visibility to technology in Valais."

Prize Money of 10,000 Swiss Francs and Three Years of Free Access to the Incubator "The Ark"

Scientific help is provided too, since competitors have two whole days to exchange ideas and results with around fifty researchers who work at the heart of the National Center for Competence in Research IM2. In addition to these ideal conditions, the winner is given three years of free access to the incubator The Ark, on the IdeArk site in Martigny (The Ark Prize) and a sum of CHF 10,000 (the IdeArk Prize).

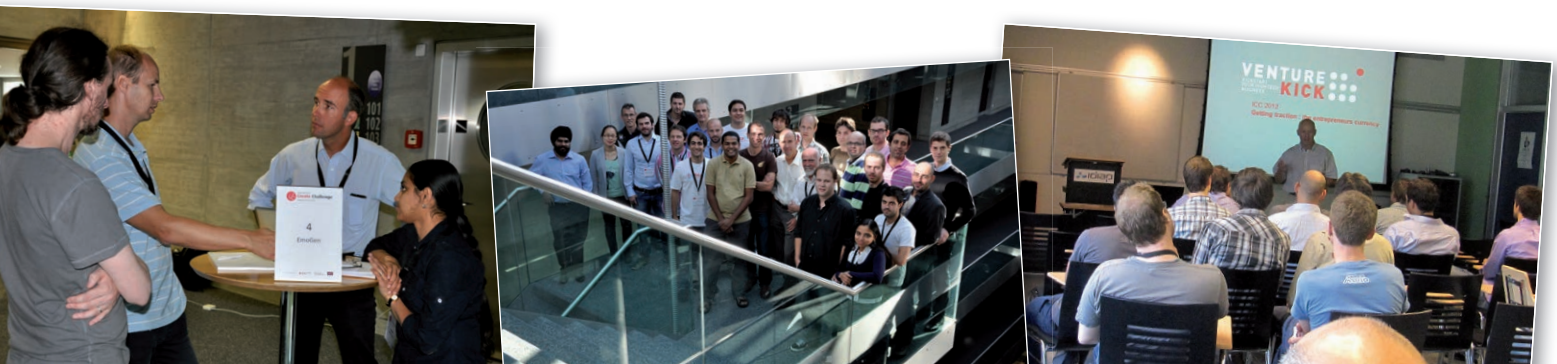
All in all, this has proved to be an extremely attractive prospect. Launched in early 2012, the call attracted no less than seventy people working in Idiap's research domains. "First we selected fifty-five candidates, then put them in touch with each other via the social network LinkedIn so they could form project teams", Foglia explains. "In the end, we selected twelve teams, comprising twenty-eight people."

A Top-Level International Competition

Successful candidates came from Idiap, EPFL, ETH Zurich, and the universities of Geneva and Fribourg, but also from other institutions in Greece, Romania, and Scotland. "We wanted an international competition, and that's what we got", says Foglia, also ad interim director of IdeArk and president of the research center created by the NCCR IM2.

ICC 2012's jury rewarded no fewer than six projects, with—at the forefront—VocaBoca, an application to facilitate language learning via smartphone or tablet computer. The competition also led to the creation of new start-ups. "It's very satisfying, as was the overall quality of the projects presented. That said, the success of the first edition isn't measured by the number of companies created alone. It also, for example, allowed new networks and research projects to be created."

All the more reason to repeat the experience, as will be the case in 2013 with a second edition planned for the period September 21 to October 11.



LEARN A LANGUAGE WITH YOUR SMARTPHONE

Interview With Alberto Armijo Daccarett Whose Application VocaBoca Won the ICC 2012.

While working on their application, Alberto Armijo Daccarett, his brother Elias, and Javier Sierra noticed an internet ad for the International Create Challenge 2012. Aware of what Idiap's expertise could bring to their project, the trio (from Colombia and Spain), comprising two financial analysts and a developer, signed up. A great move as it turned out, since their application, VocaBoca, won first prize.

VocaBoca is an application for smartphone and tablet computer allowing new words to be memorized in a fun way, and enabling users to create their own illustrated sound dictionaries. Users only need to photograph an object to obtain, in just a few seconds, its description, its definition, its translation, and its pronunciation. Or the user can enter an as-yet unknown word, in writing or orally, and receive a definition and photo of the object in question.



Left, Alberto Daccarett Armijo, right, Javier Sierra; winners – with Elias M. Daccarett Armijo (not pictured) – of ICC 2012.

“Idiap’s competition was fundamental to the development of our application; it saved us time”, explains a delighted Daccarett Armijo, before describing further his impressions.

What motivated you to participate in the ICC?

We liked the concept a lot. My teammates and I don’t live in the same country—Javier and I are in Geneva, Elías lives in Spain—so we often have to work by telephone. The ICC offered us the opportunity to focus on the project for three weeks, without distractions, without friends, without family. We could work late, until midnight, even on Saturdays and Sundays.

How was the atmosphere during those three weeks?

Awesome! Even if the teams were in competition with each other, they still formed a little community. In fact, we’ve kept in touch with some of the other participants. During the competition, we exchanged ideas and helped each other. We were surrounded by investors, coaches, and researchers, whose advice was very useful. VocaBoca is only our second application; we still have a lot to learn.

Where is VocaBoca now?

Users find the application attractive, but few are willing to buy. The most urgent tasks for us are to develop a business model and to improve our application by asking what kind of features people would be willing to pay for. This is probably the most difficult thing for us. For the moment we use generic web tools, but in the future we hope to use Idiap technologies, particularly for visual- and speech recognition. Thousands of new applications appear every week, and disappear after a few days; it’s a very competitive market. Once we’ve established the business model, the incubator The Ark will play an important role in the execution phase.

If you were asked to sum up the experience?

In terms of professional knowledge, networking, and media exposure, it helps; in fact it helps a lot. I mentioned the second edition of the ICC on my Facebook wall, and I see that there are already people interested. I really advise them to participate, because it’s a great experience from which they would learn a lot.

www.vocaboca.com



GROWTH AND NEW COMMITMENTS

Born from Idiap technology, at the heart of the incubator The Ark at IdeArk, the start-ups KeyLemon, Klewel, and Koemei all expanded in 2012, strengthening their ranks. Interviews with the new recruits.

Klewel

www.klewel.com

Since 2007, Klewel has provided innovative solutions to record, archive, and disseminate presentations and conferences on the web. Its main tool, Triskel, is a turnkey webcasting solution comprising an audiovisual recording station that fits in a suitcase. This is linked to a web platform which allows users to reference, edit, and publish audio and video content of a speaker's presentation, and the speaker's support materials (presentation slides, etc.) The presentation can be viewed on the client's website, and viewers can search for specific content using keywords. In 2008, the solution was awarded the European Label of Excellence at the CeBit international trade fair, and in 2010 was a finalist for the International Association of Congress Centres' (AIPC) Innovation Award.

Klewel currently provides custom recording technology for major conference organizers such as Nestlé, BCV, EPFL, UNICEF, and EHL. In 2013, the company plans to integrate speech recognition and a public conference-portal into its solutions.



The newcomer: Dr. Sandy Ingram

Twenty-nine years old – Lebanese – living in Martigny

(Valais, Switzerland) – Research & Development Engineer

“Involved in both a European and a federal research project, Klewel needed additional resources in R & D. I’ve been working, since December 2012, on the development of a web portal that allows the integration of different technologies provided by six European partners including IBM Research and the University of Edinburgh. I think that, in the future, Klewel’s solutions will offer an access point to a pool of recorded conference-presentations, edited and published intuitively and almost automatically.”



www.keylemon.com

Created in 2008 by an Idiap research assistant, an economist, and a professor of entrepreneurship, KeyLemon develops facial-recognition software which allows users to open a session on their computer by presenting their face to the screen. From 2011 on, the tool also works with smartphones.

Over 3 million people already use the application, and KeyLemon has 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 agreement with Fujitsu, and the start-up's technology will be pre-installed on an entire range of the Japanese giant's computers. The year also saw voice recognition integrated into their tool, and a project to deploy it on a cloud computing platform.



The newcomer: Romain Cherix

Twenty-four years old – Swiss – living in Bex (Vaud, Switzerland) – Software Development Engineer

"I've worked at KeyLemon since October 1, 2012. The company needed additional resources to develop a new cloud computing platform. I take care of web development and build demonstration applications for our customers. I think that, in the future, KeyLemon's technology will be installed on most of the tools we use. You save so much time, and it's so much more secure."

KOEMEI

www.koemei.com

By 2015, ninety percent of web content will be in video format. There is, however, no reference system for this kind of content. Koemei offers a voice transcription solution that distinguishes between different speakers and allows search engines access. Koemei's solution is a favorite of higher education institutions and has been adopted, in particular, 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. In 2012, the company opened a branch in San Francisco (US), and was also selected to participate in both TechCrunch's Disrupt New York, and in Demo Spring, a prestigious event in the heart of Silicon Valley that aims to promote new technologies in the field of ICT, to which 800 companies had submitted bids, only 50 being selected.



The newcomer: Sébastien Dupont

Thirty years old – Belgian – living in Daillon (Valais, Switzerland) – Lead Developer

"In order to make available the indexing and subtitling features made possible by transcription technology, Koemei built a web platform. My role, with the help of my team, is to develop this service: make it robust, easy to use, and able to adapt to changes in demand. Eventually, Koemei will offer a solution to improve access to all audiovisual media."



FACES



LAKSHMI SAHEER, IDIAP'S FIRST FEMALE INDIAN PHD

“MY COLLEAGUES HERE ARE AMAZING”

Since her arrival in Switzerland five years ago, Lakshmi Saheer has been awarded her PhD, has started a family, and is now preparing to launch her own start-up. Encounter with a young woman who loves scaling the heights.

You were born in India thirty-two years ago. Why did you choose to continue your studies in Switzerland?

When I was studying for my master's degree at the Indian Institute of Technology, Madras, I knew EPFL and Idiap by reputation. The Idiap-EPFL Graduate School seemed the ideal place for my doctorate. I have always been interested in speech processing, and Idiap is one of the best institutes in the world in that field. Courses at EPFL, research at Idiap—it was the perfect mix! [Editor's note: Saheer was awarded her PhD in April 2013]. And before coming, I worked for a year in Sweden for Sony Ericsson and realized that I felt good in Europe.

What surprised you most when you arrived here?

I'd never been to Switzerland. I'd heard that it was a beautiful country, but I had no idea how beautiful. I had never seen such a magnificent countryside. At first, I was a bit surprised by the length of the working week, much longer than in Sweden. But now I think this explains the high levels of productivity and efficiency that you find here.

What do you remember about your first day at Idiap?

I didn't expect the institute to be quite so big. There was a very warm welcome for me, perhaps because I was the first female Indian researcher to join. When I walked into my office, I was instantly carried away by the view: this magnificent panorama of mountains. My husband and I went hiking, to the Pierre Avoi, and to the Panossière Hut where we spent the night. I have fond memories of that and I look forward to returning as a family. That looks like a strong possibility; my daughter, 16 months old, already likes walking a lot.

How does it feel to be a woman in the world of research?

There are fewer women than men, so maybe there are fewer friends to make. But my colleagues here are amazing—I've never been discriminated against, or been put under pressure. That said, when I received Google's Anita Borg Memorial Scholarship [reserved for women students] in Zurich in 2011, I still found that community of women extremely stimulating.

And you're not only a woman, but a mother too...

Indeed. My husband gives me tremendous support, but I must confess that it's the family-friendly culture of Switzerland in general, and Idiap in particular, which has allowed me to complete my PhD and dare to start my own business. My colleagues and acquaintances also help me a lot.

Where do you see your future: in Europe, in India, elsewhere in the world?

In the long term, I want to return to India to be with my family and friends. I want to give something back to the society I come from, and I think it's important that my daughter gets to know India. But for the next few years we want to stay in Switzerland. It's an ideal place for us.

Finally, just to test how well you're integrated here: Do you know raclette?

Yes, of course, and fondue too! In Switzerland, as a vegetarian who loves cheese, deciding what to eat is not a problem.



Emotion in Voice Messages

Lakshmi Saheer is preparing to launch her own start-up. Her project, entitled Emogen (for Emotion Generation), allows emotion and expression to be added to any type of voice message, synthesized messages in particular. This technology could greatly enrich the speech of a person deprived of the ability to talk, and forced to express themselves via a speech synthesizer.



THE PROJECT MANAGEMENT TEAM

THE RESEARCH MANAGERS

Forms, contacts, deadlines, contracts, budgets... nothing is too much for them. The Project Management team supports researchers in their administrative duties, and in their search for partnerships and funding.

In 2012, Idiap managed a record forty-two scientific projects. This growth is expected to continue: By the end of January 2013, no fewer than fifteen project proposals had already been submitted. Responding to calls for tenders, coordinating projects with partners, managing mandates received—these tasks are becoming increasingly important and contribute directly to Idiap's success. This is why the Project Management team was strengthened in 2012, and three team members now assist research teams with their administrative workload. Léonore Miauton, Christophe Ecoeur, and Martina Guetl make up a team with highly complementary skills, ensuring the submission, launch, and monitoring of every project. Highly tuned to the importance of commas, cents, guidelines, and deadlines, nothing escapes them. "We're involved in all stages of a project; it's fascinating!" enthuses Christophe Ecoeur.

Report Writing, Distribution of Funds, ...

When Idiap responds, for example, to a call for tenders from the European Union, it must comply with a host of criteria. Researchers often face the challenge of only having a matter of weeks to prepare their documentation and bring together the elements required to develop a dossier of up to two hundred pages. Some mandates require the involvement of other partners in Europe, with whom it is necessary to negotiate

and sign contracts. Once funds are granted, the project management itself begins. "The European administration is very meticulous. We have to produce an annual report, redistribute funds among partners, and justify our budgets to the cent", explains Léonore Miauton. "We also have to adapt ourselves to researchers and their culture, to their IT tools and their working methods. Here, administration doesn't mean routine. Each project is a unique adventure, with its actors, its rules, and its own challenges."

Participation in Regional Events

With its intimate knowledge of current projects, the new team is well positioned to promote, and render accessible to a wider public, the research institute's activities. Hence, the team is responsible for managing Idiap's participation in various regional events, including organizing technology demonstrations in the framework of the event "Oser tous les métiers" and running science workshops at the festival "Hérisson sous gazon". The Project Management team also organizes the ICC (International Create Challenge [see pages 16-17]), a program that provides entrepreneurs and researchers worldwide with a unique opportunity to develop their product, and even launch a start-up.



Three project managers at the service of research:
Léonore Miauton, Christophe Ecoeur, and Martina Guetl.



Christophe Ecoeur, the economist
“I didn’t hesitate for a second.”

When Idiap’s management asked this accounting department assistant to join the Project Management team, the response shot out: “I’m on my way!” Christophe Ecoeur does not believe in coincidence.

The opportunity presented itself right at the moment he was looking for a new professional challenge. “I was already tracking project budgets, but now I have more responsibilities and contact with people. It’s exactly what I was dreaming of.”

Ecoeur knows Idiap well. He arrived in 2010 after completing a bachelor’s degree in Business Economics. He particularly appreciates the great atmosphere in the office: “We complement each other perfectly; working together is a pleasure.”



Martina Guetl, the literatus
“I discovered a new world.”

Nothing predestined Martina Guetl to work in a scientific environment. This 26-year-old Austrian came to Valais for love, after studying humanities and public administration in Lausanne.

Her mastery of languages and communication skills convinced Idiap’s management, and she was hired in November 2012. “I was worried about immersing myself in such a specialized world. But in fact, I like it a lot. I learn plenty of new things and I appreciate the multinational side of the institute.”

Her editing and translating skills have proven to be essential in the management of projects.



Léonore Miauton, the scientist
“We have a lot of freedom.”

A physicist educated at EPFL and ETHZ, Léonore Miauton knows the workings of scientific projects intimately. She worked in Management Intellectual Property at The Johns Hopkins University in Baltimore (US) for five years before founding her own IT company in Singapore. Her last position at EPFL involved leading large projects for National Centers of Competence in Research (NCCR).

She joined Idiap in September 2012, happy to embark on a new adventure in the service of researchers. “It’s an organization on a human scale, and management encourages us to use our own ideas and make our own proposals. It’s a great opportunity.” Miauton takes advantage of her contact with researchers to keep abreast of all the latest developments, and her experience in the field is invaluable.



ANTOINE DORSAZ, FINANCIAL ASSISTANT

PASSION PAYS

Former elite athlete Antoine Dorsaz now divides his time between the HES-SO Valais, where he studies economics, and Idiap. Encounter with a young perfectionist who is interested in everything.

On his right wrist is a discrete tattoo of the Olympic rings, bearing witness to his past as an elite athlete. At twenty-four years of age, Antoine Dorsaz has already “lived several lives”. Yesterday on ice rinks around the world, today as a financial assistant at Idiap: a part-time position which allows him to pursue his studies in economics at the HES-SO Valais.

Dorsaz’s daily routine has changed, but the passion is still there. “I want to be good at everything I do.” The first field in which this native of Valais expressed his thirst for excellence was figure skating. “I started at ten, which is quite late, so I had to work very hard to catch up. I loved it.”

His Dream: Working in an International Company

Figure skating, the demanding discipline that led Dorsaz, and his partner Anaïs Morand, to the Vancouver Olympics in 2010, is a great “university of life”. “You have to be a perfectionist. You have to be persistent. You have to be not afraid to fall, and fall again. Between a successful jump and fall, it’s a thin line.”

After placing fifteenth at the Vancouver games, Dorsaz put an end to his sporting career. “I’d lived my dream. I no longer had the motivation to train daily at the highest level.” Above all, he wants to feel connected to “real life.” “The competition, the travel... it’s a great life, but you live in something of a bubble.”

“Interested in everything”, Dorsaz wanted to combine study and work. A lover of travel and languages – he speaks French, English, German, and Russian – he returned to his native Valais with the dream of working for an international company.

“Researchers? Cerebral, but approachable”

Upon learning that Idiap was looking for a part-time financial assistant, Dorsaz applied immediately: “I knew the ins-

titute by reputation. It’s a tremendous asset to the Canton of Valais.” Today, he is in his element. “I work with people from all cultures, and I love talking with them. The researchers are both cerebral, and relaxed and approachable; it’s very impressive. I have great respect for what they do. As for my work, I

have the good fortune to be in a nice, well-organized, efficient team. Nothing drags, and I like that.”

Thanks of his part-time work at Idiap, Dorsaz has time to study at the HES-SO Valais in Sierre, where he is working toward a bachelor’s degree in Business Economics, devouring books on the economy. “It’s interesting, especially in the current context, to understand the issues, the mistakes made, ...”

As if his days were not quite full enough, Dorsaz still finds time to teach young skaters.

“Passing something on is rewarding, and it allows me to stay connected to the world of skating.” A world for which Dorsaz feels, he admits, a hint of nostalgia.



Sporting Achievements

Principal placings of Antoine Dorsaz & Anaïs Morand:

2009-2010 season

- Vancouver Olympics: 15th
- World Championships, Turin: 13th
- European Championships, Tallinn: 8th

2008-2009 season

- World Championships, Los Angeles: 14th
- Junior World Championships, Sofia: 10th
- European Championships, Helsinki: 12th

JOINING AND LEAVING

In 2012, eighteen members of staff left and twenty-one joined the team. The new arrivals include two program managers, one financial assistant, seven postdoctoral researchers, and eleven research assistants.

JOINING US IN 2012

First name, last name, position, country of origin, place of residence

Chindansh Bhatt, Postdoc, India	Joël Legrand, Research Assistant, France
Olivier Canévet, Research Assistant, France	Adolfo Lopez Mendez, Postdoc, Spain
Antoine Dorsaz, Financial Assistant, Switzerland, Fully	Léonore Miauton, Senior Program Manager, Switzerland, Chexbres
Nesli Erdogan, Postdoc, Turkey	Youssef Oualil, Research Assistant, Morocco
Paul Gay, Research Assistant, France	Nikolaos Pappas, Research Assistant, Greece
Martina Guetl, Junior Program Manager, Austria	Novi Patricia, Research Assistant, Indonesia
Manuel Günther, Postdoc, Germany	Darshan Santani, Research Assistant, India
Pierre-Edouard Honnet, Research Assistant, France	Gyorgy Szaszak, Postdoc, Hungary
Ilja Kuzborskij, Research Assistant, Lithuania	Romain Tavenard, Postdoc, France
Alexandros Lazaridis, Postdoc, Greece	Raphaël Ullmann, Research Assistant, Switzerland, Lausanne
Rémi Lebret, Research Assistant, France	

MOVING ON IN 2012

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

Constantin-Cosmin Atanasoaei, research assistant, Romania, 2008, Pix4D, Ecublens
Venkatesh Bala Subburaman, research assistant, India, 2007, Multitel Mons, Belgium
Cheng Chen, postdoc, China, 2010, ISTB (University of Bern), Berne
Valérie Devanthéry, program manager, Switzerland, 2008
Stefan Duffner, postdoc, Germany, 2008
Jean-Albert Ferrez, deputy director, Switzerland, 2001, Energie de Sion Région, Sion
Pierre Ferrez, program manager, Switzerland, 2004, HES-SO Valais, Sion
Danil Korchagin, postdoc, Russia, 2008
Gwenolé Lecorvé, postdoc, France, 2011, Université de Rennes, Rennes
Hui Liang, research assistant, China, 2008, ETHZ, Zurich
Christopher McCool, postdoc, Australia, 2008, NICTA, Brisbane, Australia
Youssef Oualil, research assistant, Morocco, 2012, Universität des Saarlandes, Saarbrücken, Germany
Edgar Francisco Roman Rangel, research assistant, Mexico, 2008, University of Geneva, Geneva
Serena Soldo, research assistant, Italy, 2009
Gokul Thattaguppa Chittaranjan, research assistant, India, 2010
Fabio Valente, researcher, Italy, 2005
Jagannadan Varadarajan, research assistant, India, 2008, Illinois at Singapore Pte LTD, Singapore
Roy Geoffrey Wallace, postdoc, Australia, 2010, Zap Technology, Brisbane, Australia



DISTINCTIONS

DISTINCTIONS

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 **David Imseng** for his record of outstanding publications and for the excellence of his research into the topic **“multilingual speech recognition”**. The Research Prize went to **Lakshmi Saheer** for her outstanding scientific article entitled **“Vocal Tract Length Normalization for Statistical Parametric Speech Synthesis”**.



David Imseng



Lakshmi Saheer

External

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

Daniel Gatica-Perez and Kate Farrahi

Best Paper Award, IEEE Int. Symposium on Wearable Computers, June 2012

“Extracting Mobile Behavioral Patterns with the Distant N-Gram Topic Model”

Kenneth Funes

Best Student Paper Award, CVPR workshop on Gesture Recognition, July 2012

“Gaze Estimation from Multimodal Kinect Data”

Anindya Roy

European Biometrics Industry Award 2012, September 2012

“A Fast Parts-Based Approach to Speaker Verification Using Boosted Slice Classifiers”

Dinesh Babu Jayagopi and Dairazalia Sanchez-Cortes

Outstanding Paper Award at ICMI 2012, October 2012

“Linking Speaking and Looking Behavior Patterns with Group Composition, Perception, and Performance”

Daniel Gatica-Perez

Paper Award, ACM Int. Conf. on Multimodal Interaction, October 2012

“Linking Speaking and Looking Behavior Patterns with Group Composition, Perception, and Performance”

Eric Malmi, Trinh-Minh-Tri Do and Daniel Gatica-Perez

Best Paper Award nomination at MUM 2012, December 2012

“Checking In or Checked In: Comparing Large-Scale Manual and Automatic Location Disclosure Patterns”

THESES COMPLETED

Seven students completed their thesis in 2012: Venkatesh Bala Subburaman, Constantin-Cosmin Atanasoaei, Jagannadan Varadarajan, Hui Liang, Tatiana Tommasi, Edgar Francisco Roman-Rangel, and Lakshmi Babu Saheer.

- ***Alternative Search Techniques for Face Detection Using Location Estimation and Binary Features***
Venkatesh Bala Subburaman, February 14, 2012
Thesis directors: Prof. Hervé Bourlard and Dr. Sébastien Marcel.
Members of the thesis committee: Dr. Hazim Kemal Ekenel, Prof. Abdenour Hadid, and Prof. Jean-Philippe Thiran.
- ***Boosting Multivariate Look-up Tables for Face Processing***
Constantin-Cosmin Atanasoaei, April 27, 2012
Thesis directors: Prof. Hervé Bourlard and Dr. Sébastien Marcel.
Members of the thesis committee: Prof. Josef Kittler, Prof. Timothy Cootes, and Dr. Vincent Lepetit.
- ***Sequential Topic Models for Mining Recurrent Activities and their Relationships: Application to Long Term Video Recordings***
Jagannadan Varadarajan, October 30, 2012
Thesis directors: Prof. Hervé Bourlard and Dr. John Dines.
Members of the thesis committee: Prof. Keiichi Tokuda, Prof. Tanja Schultz, and Dr. Jean-Marc Vesin.
- ***Data-Driven Enhancement of State Mapping-Based Cross-Lingual Speaker Adaptation***
Hui Liang, 30 octobre 2012
Directeurs de thèse: Prof. Hervé Bourlard, Dr. John Dines
Membres du jury: Prof. Keiichi Tokuda, Prof. Tanja Schultz, Dr. Jean-Marc Vesin
- ***Learning to Learn by Exploiting Prior Knowledge***
Tatiana Tommasi, November 7, 2012
Thesis directors: Prof. Hervé Bourlard and Dr. Barbara Caputo.
Members of the thesis committee: Prof. Jean-Philippe Thiran, Prof. Vittorio Ferrari, and Prof. Jim Little.
- ***Statistical Shape Descriptors for Ancient Maya Hieroglyphs Analysis***
Edgar Francisco Roman-Rangel, November 28, 2012
Thesis directors: Dr. Daniel Gatica-Perez and Dr. Jean-Marc Odobez.
Members of the thesis committee: Prof. Jean-Philippe Thiran, Prof. Stéphane Marchand-Maillet, and Dr. Changhu Wang.
- ***Unified Framework of Feature Based Adaptation for Statistical Speech Synthesis and Recognition***
Lakshmi Babu Saheer, November 28, 2012
Thesis director: Prof. Hervé Bourlard.
Members of the thesis committee: Prof. Alan W. Black, Dr. Ralf Schlüter, and Prof. Jean-Philippe Thiran.



FINANCES



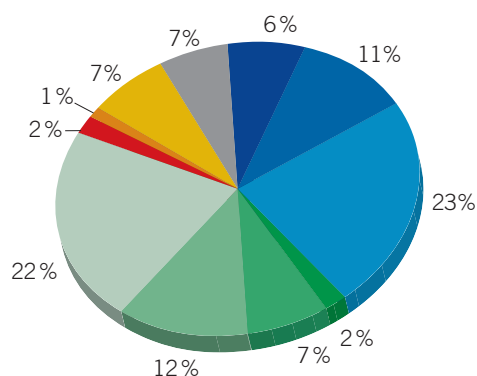
OPERATING ACCOUNTS

(Swiss francs)	2011	2012	%
INCOME			
City of Martigny	650,000	650,000	6.29%
Canton of Valais	900,000	1,120,000	10.83%
Swiss Confederation	2,357,000	2,393,000	23.14%
Loterie Romande	-	250,000	2.42%
NCCR IM2 projects	533,900	764,071	7.39%
Swiss National Science Foundation Projects	1,072,445	1,203,153	11.64%
European Commission Projects	2,390,855	2,219,526	21.47%
CTI Projects	105,135	251,168	2.43%
EPFL Contribution	72,000	72,000	0.70%
Industrial Financing	943,656	746,307	7.22%
Other income / extraordinary income	903,443	669,450	6.47%
TOTAL INCOME	9,928,434	10,338,675	100.00%
EXPENSES			
Personnel expenses	7,520,332	7,608,490	73.59%
Education and travel	473,803	528,042	5.11%
Third party expenses	119,968	92,469	0.89%
Computer equipment and maintenance	237,787	249,146	2.41%
Administrative costs	159,423	178,872	1.73%
Promotion and communication / 20 th anniversary (2011)	241,641	77,007	0.75%
Rent	872,967	889,306	8.60%
Depreciation	290,622	281,777	2.72%
Exceptional expenses	40,000	421,340	4.08%
TOTAL EXPENSES	9,956,543	10,326,449	99.88%
OPERATING PROFIT / LOSS	-28,109	12,226	0.12%



SOURCES OF FUNDS / 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
- European Commission projects
- CTI projects
- EPFL contribution
- Industrial financing
- Other income / Extraordinary income

Management comments

In 2012, Idiap maintained its momentum and its financial statement shows a profit of CHF 12,226, and a nice increase of 4.13% in volume compared to the previous year. The mythical 10-million volume barrier was passed for the first time since the institute's creation in 1991. The proportion of funding from public sources remains within its strategically defined range, at 40.3%. The Canton of Valais increased its contribution significantly to CHF 1,120,000. Compared to 2011, total public subsidies grew by 6.55%. All existing provisions (CHF 301,000) were released. This consideration is reflected in the exceptional charges.

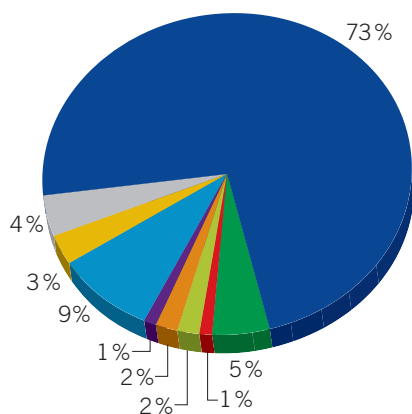
Swiss Confederation, Canton, and Municipality Subsidies

(In thousands of Swiss francs)

YEAR	2010	2011	2012	2013*
Confederation	1,795	2,357	2,393	2,428
Canton	900	900	1,120	1,220
Municipality	600	650	650	700

* Budget

Distribution of Costs



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

BALANCE SHEET

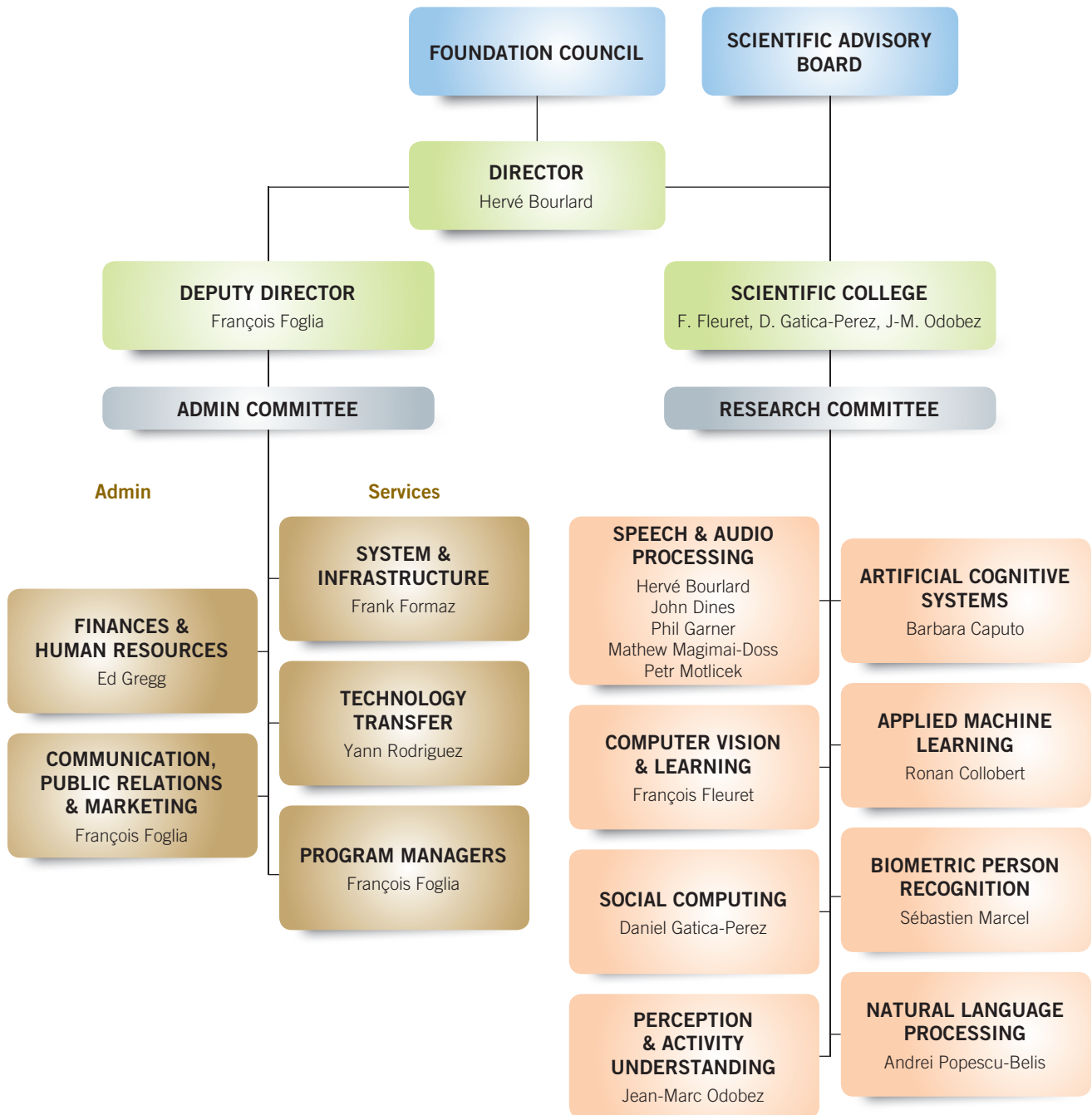
(Swiss francs)

	31.12.2011	31.12.2012
ASSETS		
Cash	2,443,047.24	2,104,427.07
Accounts Receivables	499,887.55	287,382.35
Accrued income and other	3,016,530.83	2,536,046.69
TOTAL CURRENT ASSETS	5,959,465.62	4,927,856.11
Equipment	523,402.89	504,960.99
Financial assets	10,000.00	10,000.00
TOTAL NON-CURRENT ASSETS	533,402.89	514,960.99
TOTAL ASSETS	6,492,868.51	5,442,817.10
LIABILITIES		
Accounts payable	461,687.65	365,838.69
Accrued expense	4,583,667.75	3,918,238.84
Provisions	301,000.00	0.00
TOTAL FOREIGN FUNDS	5,346,355.40	4,284,077.53
Share capital	40,000.00	40,000.00
Reserve	1,000,000.00	1,000,000.00
Retained earnings	134,622.17	106,513.11
Net income	-28,109.06	12,226.46
TOTAL OWN FUNDS	1,146,513.11	1,158,739.57
TOTAL LIABILITIES	6,492,868.51	5,442,817.10



ORGANIZATION







EMPLOYEES

Scientists

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

Oya Aran Karakus, Postdoc, Turkey, 2009
Afsaneh Asaei, Research Assistante, Iran, 2008
Chindansh Bhatt, Postdoc, India, 2012
Joan Isaac Biel, Research Assistant, Spain, 2008
Hervé Bourlard, Director, Switzerland, Saxon, 1996
Olivier Canévet, Research Assistant, France, 2012
Barbara Caputo, Senior Research Scientist, Italia, 2005
Ivana Chingovska, Research Assistante, Macedonia, 2011
Ronan Collobert, Research Scientist, France, 2010
John Dines, Research Scientist, Australia, 2003
Trinh-Minh-Tri Do, Postdoc, Vietnam, 2009
Charles Dubout, Research Assistant, Switzerland, Renens, 2009
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, Italia, 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
Arjan Gijsberts, Postdoc, Netherlands, 2011
Manuel Günther, Postdoc, Germany, 2012
Maryam Habibi, Research Assistant, Iran, 2011
Najeh Hajlaoui, Postdoc, France, 2011
Alexandre Heili, Research Assistant, France, 2010
Pierre-Edouard Honnet, Research Assistant, France, 2012
David Imseng, Research Assistant, Switzerland, Rarogne, 2009
Dinesh Babu Jayagopi, Research Assistant, India, 2007
Samuel Kim, Postdoc, South Korea, 2011
Ilja Kuzborskij, Research Assistante, Lithuania, 2012
Alexandros Lazaridis, Postdoc, Greece, 2012
Rémi Lebret, Research Assistant, France, 2012
Leonidas Lefakis, Research Assistant, Greece, 2010
Riwal Lefort, Postdoc, France, 2011
Joël Legrand, Research Assistant, France, 2012
Adolfo Lopez-Mendez, Postdoc, Spain, 2012
Mathew Magimai Doss, Research Scientist, India, 2007
Sébastien Marcel, Senior Research Scientist, France, 2000
Thomas Meyer, Research Assistant, Switzerland, Martigny, 2010
Gelareh Mohammadi, Research Assistante, Iran, 2009
Petr Motlicek, Research Scientist, Czech Republic, 2005
Laurent Nguyen, Research Assistant, Switzerland, Lausanne, 2011
Jean-Marc Odobez, Senior Research Scientist, France/Switzerland, Clarens, 2001
Dimitri Palaz, Research Assistant, Switzerland, Martigny, 2011
Nikolaos Pappas, Research Assistant, Greece, 2012
Novi Patricia, Research Assistante, Indonesia, 2012
Andrei Popescu-Belis, Senior Research Scientist, France / Romania, 2007
André Rabello Dos Anjos, Postdoc, Brazil, 2010
Ramya Rasipuram, Research Assistante, India, 2010
Lakshmi Saheer, Research Assistante, India, 2008
Dairazalia Sanchez-Cortes, Research Assistante, Mexico, 2009
Darshan Santani, Research Assistant, India, 2012
Ashtosh Sapru, Research Assistant, India, 2011
Samira Sheikhi, Research Assistante, Iran, 2010
Nicolae Suditu, Research Assistant, Romania, 2008
Gyorgy Szaszak, Postdoc, Hungary
Mohammad Javad Taghizadeh, Research Assistant, Iran, 2011
Romain Tavenard, Postdoc, France, 2012
Tatiana Tommasi, Research Assistante, Italia, 2008
Raphaël Ullmann, Research Assistant, Switzerland, Lausanne, 2012
Alessandro Vinciarelli, Senior Research Scientist, Italia, 1999
Majid Yazdani, Research Assistant, Iran, 2008
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 senior, Slovakia, 2011
Vasil Khalidov, Senior Dev. Engineer, Russia, 2010
Salim Kayal, Dev. Engineer, Switzerland, Vevey, 2011
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, Junior Program Manager, Switzerland, Collombey, 2010
François Foglia, Deputy Director, Switzerland, Saxon, 2006
Edward-Lee Gregg, Financial Manager, United States, 2004
Martina Guetl, Junior Program Manager, Austria, 2012
Léonore Miauton, Senior Program Manager, Switzerland, Chexbres, 2012
Sylvie Millius, Administrative assistant, Switzerland, Vétroz, 1996
Yann Rodriguez, Technology Transfer, 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, Suisse, Martigny-Combe, 2004

Internships

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

Idiap interns generally spend between three and ten months at the research institute. Some are students at l'Ecole Polytechnique Fédérale de Lausanne (EPFL), and follow their internship as part of their course work. Others are welcomed within the framework of exchange programs established within European projects in which Idiap participates.

Aniruddha Adiga, IND, Florida Tech University, USA
Philippe Ballestraz, CHE, HES-SO Valais, Sion, CH
Okal Billy, KEN, Jacobs University, Bremen, DEU
Holger Caesar, DEU, Karlsruhe Institute of Technology, DEU
Aleksandra Cerekovic, HRV, Univ. of Zagreb, HRV
Thiyagarajan Chockalingam, IND, Colorado State University, USA
Tiago de Freitas Pereira, BRA, Univ. de Campinas, Sao Paulo, BRA
Michal Drozdal, POL, Autonomous University of Barcelona, ESP
Manon Fournier, CHE, Collège de l'Abbaye, St-Maurice, CHE
Javier Galbally, ESP, Autonomous University of Madrid, ESP
Huseyn Gasimov, AZE, EPFL, Lausanne, CHE
Hu Hainan, CHN, EPFL, Lausanne, CHE
Mohsen Kaboli, IRN, KTH University, Stockholm, SWE
Yunus-Emre Kara, TUR, Bogazici University, Istanbul, TUR
Alp Kindiroglu, TUR, Bogazici University, Istanbul, TUR
Ilja Kuzborskij, LTU, Univ. of Edinburgh, GBR
Kexing Li, CHN, EPFL, Lausanne, CHE

Jukka Määttä, FIN, University of Oulu, FIN
Alvaro Marcos, ESP, Univ. of Alcala, ESP
Miranti Mandasari, IDN, Radboud University, NLD
Braid Regula Meyer, CHE, ISME, St. Gallen, CHE
Peter Naoki, CHE, EPFL, Lausanne, CHE
Pedro Henrique Oliveira Pinheiro, BRA, EPFL, Lausanne, CHE
Jérémie Rappaz, CHE, HES-SO Valais, Sion, CHE
Sandrine Revaz, CHE, Univ. de Fribourg, CHE
Bin Shen, CHN, Purdue University, Indiana, USA
Yang Sun, CHN, Radboud University, NLD
Romain Tavenard, FRA, Univ. of Rennes 1, IRISA-INRIA, FRA
Lucia Teijeiro Mosquera, ESP, Univ. of Vigo, ESP
Kelly Tiraboschi, CHE, Collège de l'Abbaye, St-Maurice, CHE
Aleksi Triastcyn, RUS, EPFL, Lausanne, CHE
Vagia Tsiminaki, GRC, EPFL, Lausanne, CHE
Na Xingyu, CHN, Beijing Institute of Technology, CHN
Eric Malmi, FIN, Aalto University, FIN

Visitors

First name, last name, home institution

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

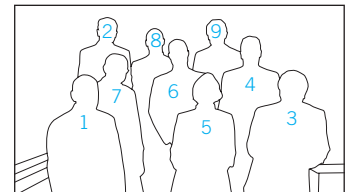
Elie El Khoury, Liban, Université de Toulouse, Toulouse, France
Filiberto Pla, Espagne, Jaume I University, Castelló de la Plana, Spain
Adrian Penate-Sanchez, Espagne, Polytechnic University of Catalonia, Barcelone, Spain



FOUNDATION COUNCIL



The Foundation Council is responsible for the economic and financial management of Idiag. It defines the institute's structure, appoints its director, and – more generally – ensures Idiag's development and defends its interests.



- 1 **Olivier Dumas**, President
Director of Electricité d'Emosson SA
- 2 **Jean-Daniel Antille**, Vice President
Manager of "Antenne Régions Valais romand"
- 3 **Prof. Philippe Gillet**, Vice President
Vice President for Academic Affairs
Ecole Polytechnique Fédérale de Lausanne (EPFL)
Marc-André Berclaz (*not pictured*)
President of the Board of Directors of HES-SO
Stefan Bumann
Head of Higher Education, state of Valais
- 4 **M. Stefan Bumann**
Head of Higher Education, State of Valais
- 5 **Adrienne Corboud Fumagalli**
Vice President for Industrial Relations
Ecole Polytechnique Fédérale de Lausanne (EPFL)
- 6 **Marc-Henri Favre**
President of the city of Martigny
- 7 **Patrick Furrer**
Vice Director of Euresearch
Jean-René Germanier (*not pictured*)
Member of Parliament
- 8 **Jordi Montserrat**
Regional Director of venturelab and venture kick
Prof. Christian Pellegrini (*not pictured*)
Honorary Professor in the Faculty of Sciences
at the University of Geneva
- 9 **Walter Steinlin**
Swisscom University Relations
President of the Commission for Technology
and Innovation (CTI)

In 2012, Idiap's Foundation Council welcomed three new members. Members who, by their professional positions, their experience, and their vision, offer Idiap new perspectives. Interviews.

- 1) What binds you to Idiap?
- 2) What is your vision of the institute and its strengths?
- 3) How do you see the Idiap of tomorrow?

“The institute has made, since its inception, a visionary choice of themes”



Marc-André Berclaz
Chairman of the Board of the HES-SO Valais Wallis, Operations Director of the EPFL Valais Wallis centre

1) The HES-SO is engaged in applied research and development and, as such, regularly collaborates with Idiap. We want to contribute to developing and strengthening this relationship further.

2) Developing a real center of excellence in Valais, without a university, wasn't a foregone conclusion. The institute has made, since its inception, a quite visionary choice of themes. It has been able to create for itself a very important place on the international scientific stage. This is evidence of great energy and a strong entrepreneurial capacity.

3) Current changes favor large structures. This is why Idiap must now develop through alliances, especially those with EPFL. It also needs to consider how to ensure its longevity through the generations to come.

“I see the Idiap of tomorrow as a European- even a global center of excellence”



Patrick Furrer
Deputy Director Euresearch and national contact point for information and communication technologies (ICT)

1) We have enjoyed a decade of strong links with Idiap, which we advise on its strategy of integration into European Union projects.

2) In the space of twenty years, the “little institute” that Idiap was when it started out has done something magical in Valais, by growing and creating innovative businesses. This success story, which—as a native of Valais—delights me, owes much to the institute's director, but also to the support of the city of Martigny, the Canton of Valais, and the Confederation.

3) The institute has unique competencies in the areas of speech and image recognition, biometrics, and conference management. I see the Idiap of tomorrow as a European- even a global center of excellence. The know-how is there; it just lacks a little more visibility.

“A fine institute which knows how to position itself”



Jordi Montserrat
Director of “venturelab” and “venture kick” for Western Switzerland

1) The promotion of research has today become a fundamental theme, and it's in this context that we've been working with Idiap for a number of years, participating in the financing and the guiding of some of its spin-offs.

2) Idiap has demonstrated its ability to manage projects of a very high level. Its research domains, such as human-computer interaction or artificial intelligence, are very promising. I like institutes of this kind: researching specific topics, well-led, and standing out at an international level.

3) The challenge of the years to come will be to remain at the forefront. In this context, collaboration with EPFL and the CTI (Commission for Technology and Innovation) is essential. It is also a question of placing an accent on promoting entrepreneurship: more spin-offs and more and more international exposure for a fine institute which knows how to position itself.



ADVISORY BOARD

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

Dr. Jordan Cohen

Independent Consultant, Spelamode
Half Moon Bay, CA, USA

Prof. Dr. Donald Geman

Professor of Mathematics, Johns Hopkins University
Baltimore, MD, USA

Dr. John Makhoul

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

Prof. Dr. Nelson Morgan

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

Dr. David Nahamoo

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

Prof. Gerhard Sagerer

Rector, Bielefeld Universität
Germany

Dr. Roelof van Zwol

Senior Research Scientist, Multimedia, Audience Sciences, Yahoo!
Research
Santa Clara, CA, USA

Prof. Dr. 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



www.swisscom.com



Swiss Power Group.

www.groupemutuel.ch



www.epfl.ch



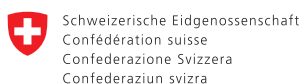
www.theark.ch



www.ideark.ch



www.snf.ch



Innovation Promotion Agency CTI

www.bbt.admin.ch/kti

HASLERSTIFTUNG

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, related, research directions, while still fostering internal multi-disciplinary collaborations, Idiap reorganized its internal structuring of its research themes along the following dimensions. Idiap has thus changed/adapted the way it presents itself and describes its current activities, to take into account the new areas of development not only towards human-computer interaction but also toward human-to-human interaction, collaboration, behavior, and innovation. Thus, after several (13) years of positioning itself under the general theme of “Multimodal human-computer interaction”, Idiap decided to officially cover a larger research domain, now referred to as “**Human and Media Computing**”.

Articulated around our current activities, “Human and Media Computing” now covers the following research themes:

- **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.

As evidenced in the present report, new major trends and developments at Idiap are now directed towards multilingual speech processing (multilingual speech recognition and synthesis), text-level machine translation and, possibly later, speech-to-speech translation.

Building on Idiap’s expertise in audio-visual processing, machine learning and robotics, there is also more and more emphasis on new activities in monitoring and autonomous robots able to build semantic spatial representations for indoor environments, updating continuously such representation taking advantage of incoming data as well as of prior experience from other robots. Furthermore, we are making significant contributions in the emerging area of advanced prostheses, leveraging on Idiap’s expertise in robotics, brain-computer interface and machine learning.

- **Social/human behavior:** Social media; Mobile media; Social interaction analysis; Social signal processing.

This area spans methods for analyzing human and social behavior 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.

Continued projects include LS-CONTEXT (a long-term collaboration on smartphone sensing and data mining with Nokia Research that resulted in two Paper Awards in 2011); SONVB (a SNSF Sinergia interdisciplinary project featured in the Digital Life issue of the official SNSF Research Magazine); and VlogSense (funded by the NCCR IM2). New projects that started in 2011 include NISHA (in collaboration with NTT Communication Science Labs, Japan) and SOBE (a SNSF Ambizione Fellowship Grant). Idiap also worked together with the University of Geneva on an NCCR proposal towards creating a new type of Social Computer.

- **Information interfaces and presentation:** 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; Countermeasures to spoofing attacks.

Biometric person recognition refers to the process of automatically recognizing a person using distinguishing behavioral patterns (gait, signature, keyboard typing, lip movement, hand-grip) or physiological traits (face, voice, iris, fingerprint, hand geometry, EEG, ECG, ear shape, body odor, body salinity, vascular).



Idiap is currently a recognized leader in that field, fully exploiting our multidisciplinary expertise in image processing, computer vision, pattern recognition and machine learning. After the end of the MOBIO EU project, this leadership Idiap–EPFL Joint Development Plan – 2011 Activity Report (PART 1) page 9/35 was confirmed in 2011 by the granting of new large EU projects, like the ICT project TABULA RASA and the SECURITY project BEAT. In 2011, Idiap and EPFL also worked together to initiate a joint NCCR proposal for a “Swiss Center for Biometric Security”.

- **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 DeepSTD project), Natural Language Processing with a particular interest in semantic analysis (SemTex 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 community; these include the AMI corpus¹ and the Juicer ASR decoder². More recently, the MediaParl database has been released³. 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.

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 EU project D-Box⁴), 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 effectively exploit acoustic and linguistic resources of a different language or multiple languages (funded through SNSF projects like ICS-2010⁵, and FlexASR⁶).

- **Cross-lingual and multi-lingual speech recognition.**
In addition to the above, and in the context of an Industry sponsored project (DAUM⁷), we are also investigating new fast acoustic model adaptation techniques (currently referred to as "subspace gaussian modeling") 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.
- **Grapheme-based ASR.**
State-of-the-art ASR systems typically use phones as the standard subword units. However, phones are not always well defined, 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 AddG2SU⁸).
- **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 the EU Marie-Curie project SCALE⁹ or SNSF projects like AMUSE¹⁰, 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 fletched 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-re-

1 <http://www.idiap.ch/mmm/corpora/ami>

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

3 <http://www.idiap.ch/dataset/mediaparl>

4 <http://www.idiap.ch/scientific-research/projects/d-box-a-generic-dialog-box-for-multilingual-conversational-applications>

5 <http://www.idiap.ch/scientific-research/projects/interactive-cognitive-systems>

6 <http://www.idiap.ch/scientific-research/projects/flexible-grapheme-based-automatic-speech-recognition>

7 <http://www.idiap.ch/scientific-research/projects/domain-adaptation-using-sub-space-models>

8 <http://www.idiap.ch/scientific-research/projects/flexible-acoustic-data-driven-grapheme-to-subword-unit-conversion>

9 <http://www.idiap.ch/scientific-research/projects/speech-communication-with-adaptive-learning>

10 <http://www.idiap.ch/scientific-research/projects/adaptive-multilingual-ASR-TTS>



sourced 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.

Speech Synthesis

- Text-to-speech synthesis (TTS).

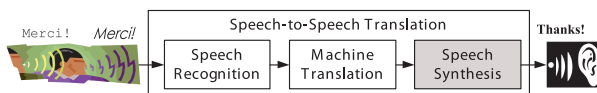
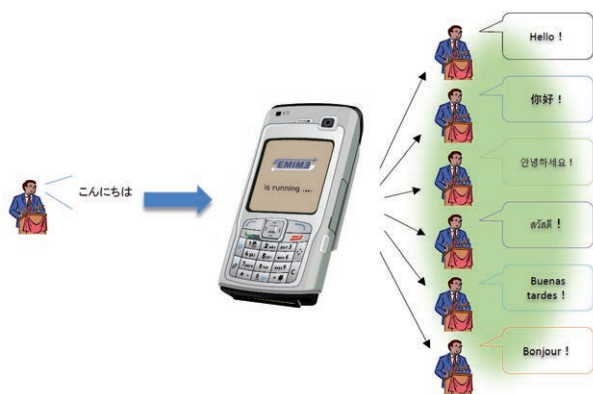


Figure 1 : Typical speech-to-speech translation system architecture developed at Idiap.

TTS is a relatively new venture for the speech group at Idiap, but two events recently marked its maturity: The first was the graduation of the first two PhD students in TTS; the second was the beginning of the SIWIS¹¹ project (Spoken Interaction With Interpretation in Switzerland). 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 EU FP7 EMIME¹² 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.

11 <http://www.idiap.ch/project/siwis>

12 <http://www.emime.org/>

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.

- 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. Funded by armasuisse¹³, 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. This work is mainly carried out in the context of the EU Marie-Curie project SCALE¹⁴.
- Microphone arrays. Idiap further performs studies on audio recorded from microphone arrays, especially those organized in “ad-hoc” topologies. Recently, we mainly focused on the automatic calibration of arbitrary microphone array geometries (using techniques like matrix completion and multi-dimensional scaling) in order to perform 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. This work is currently pursued in the context of the NCCR IM2¹⁵.

13 <http://www.ar.admin.ch/internet/armasuisse/en/home.html>

14 <http://www.idiap.ch/scientific-research/projects/speech-communication-with-adaptive-learning> and <http://www.scale.uni-saarland.de/cms/>

15 <http://www.im2.ch>



- **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 FP7 project inEvent¹⁶ and the SNSF project RODI¹⁷.

- **Audio indexing**

In the context of the Hasler Foundation project SESAME¹⁸, 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.**

Recently, 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 project¹⁹. 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 PANDA²⁰, 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 applications mostly in the processing of images and video streams.

Our current research follows two main axes. The first is the study of novel generic learning methods to cope with large dimension feature spaces 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

- **Learning in large dimension feature space.**

In the context of the EU FP7 MASH project²¹, we are investigating the learning of classifiers in large features space, ideally to cope with hundreds of families of feature extractors, each of tens of thousands of dimensions.

We have proposed two variants of AdaBoost to cope with these difficulties. The first one consists of 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, so that we can bias the sampling accordingly. The second one models the loss reduction as a function of the number of features looked at, and the number of samples used to estimate edges. This model allows to optimize the trade-off between the two.

- **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: identify good “sources” of samples (web sites, data-bases, 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. The latter is addressed currently in the context of the MASH project, in which we are developing novel techniques between batch and on-line learning by keeping a limited amount of samples in a *reservoir*, that optimizes the reduction of the loss on the full sample population.

- **High-dimension density similarity measures.**

We are investigating the clustering and classification of distributions of points in high-dimension spaces. To leverage the remarkable ability of random forests to cope with such

16 <http://www.inevent-project.eu/>

17 <http://www.idiap.ch/scientific-research/projects/role-based-speaker-diarization>

18 <http://www.idiap.ch/scientific-research/projects/searching-swiss-audio-memories>

19 <http://www.idiap.ch/project/armasuisse/recod/index.html>

20 <http://www.idiap.ch/scientific-research/projects/perceptual-background-noise-analysis-for-the-newest-generation-of-telecommunication-systems-1>

21 <http://mash-project.eu>



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.

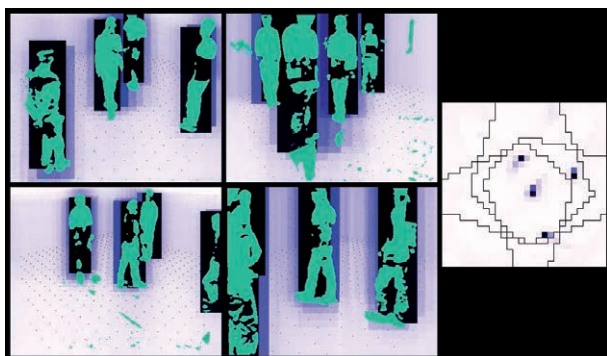


Figure 2 : Our fast version of the Deformable Part Model object detector (left) is close to one order of magnitude faster than other existing implementations. The Probabilistic Occupancy Map (right) estimates probabilities of presence on the ground, given the result of a background subtraction procedure in individual views.

Scene understanding

- **Fast object detection.**

In the SNSF project VELASH, we are investigating 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.

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. In practice, instead of dealing only with scale and location at the algorithmic level, we let the system similarly cope with any out-of-plan rotation, by making the image measurements follow the pixel displacements.

- **Multi-camera tracking.**

We have a long-lasting collaboration with the CVLab at EPFL around the design of a robust multi-camera tracking system. It is composed of the Probabilistic Occupancy Map, and a flow-based convex multi-target 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.

We used the same flow-based tracking procedure to minimize the exponential loss by optimizing the locations of training examples, alternatively with a standard AdaBoost to train an image-based model for detection. The resulting FlowBoost procedure reduces by more than an order of magnitude the required amount of labelled examples, and was used with success for pedestrian tracking in monocular sequences, and neuron tracking in microscopic videos.



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 2012 was composed of 1 head of group, 2 postdoctoral researchers, 1 SNSF Ambizione Research Fellow, 5 PhD students, and 8 visiting researchers (1 faculty-on-sabbatical, 1 postdoc, and 6 MS/PhD students). Research lines that were investigated in 2012 included analysis of ubiquitous face-to-face interaction, behavioral analysis of online conversational video, and urban data mining using smartphones and location-based social networks.

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 (job interviews and job implementation), we have investigated connections between automatically measured audio and visual nonverbal cues and stress, hireability, and performance.



Figure 3 : 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 collective classification and transfer learning approaches. This research was conducted in the framework of two SNSF projects (SONVB²²: Sensing and Analyzing Or-

²² <http://www.idiap.ch/project/sonvb>

ganizational Behavior; SOBE²³, an Ambizione Research Fellowship) and the NISHA²⁴ project (NTT-Idiap Social Behavior Analysis Initiative) funded by NTT Communication Science Laboratories, Japan. One of our works on small-group analysis was awarded an Outstanding Paper Award at the ACM Int. Conf. on Multimodal Interaction 2012. Our research in this domain has also been featured in New Scientist, ACM Tech-News, and the SNSF Horizons Research Magazine.

Online conversational video

The work in this line develops around 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 4). More specifically, we are examining 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.

As a new element, video crowdsourcing with platforms like Amazon Mechanical Turk has been used to collect interpersonal judgments; this is a powerful technique that can scale 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 Behavior in Social Media).

Urban data mining

Our work in this domain spans two lines: large-scale smartphone sensing and location-based social networks (Figure 4). For the first line, jointly with Nokia Research, we led the Mobile Data Challenge 2012, an initiative in which data from the Lausanne Data Collection Campaign was released for research purposes to the mobile computing community.

The Lausanne data is perhaps the largest mobile sensing academic effort in Europe to date, involving close to 200 volunteers in French-speaking Switzerland and smartphone data for up to 18 months. The Challenge attracted participation from several hundred researchers around the world. Regarding modeling, our work focused on automatic place recognition and probabilistic location prediction using a variety of location sensor data (GPS, Wifi, etc).

Regarding location-based social networks, supported by the

²³ <http://www.idiap.ch/scientific-research/projects/multimodal-computational-modeling-of-nonverbal-social-behavior-in-face-to-face-interaction-1>

²⁴ <http://www.idiap.ch/project/nisha/>



SNSF HAI²⁵ project (Human Activity and Interactivity), we are investigating the use of FourSquare data for two goals: the development of transfer learning techniques to boost smartphone-based location prediction models, and the characterization of phenomena at large geographic scales like cities, including characterization of urban places according to language usage, and sampling effects in location-based social networks.

Our work in this domain received three Conference Paper Awards between 2010 and 2012, and was featured in 2012 in New Scientist, PhysOrg, and Les Echos.

4. Perception and Activity Understanding

The Perception and Activity Understanding group conducts research in human activities analysis from multi-modal data. This entails the investigation of fundamental tasks like the representation, detection, segmentation and tracking of people, the characterization of their state, and the modeling of sequential data and their interpretation in forms of gestures, activities, behavior or social relations, through the design of principled algorithms extending models from computer vision, statistical learning, or multimodal signal processing. Some of the main research carried out this year are described below.

Tracking and person representation

In the context of the EU FP7 VANAHEIM project, we have continued our investigation towards enhanced behavior recognition in surveillance context. Different research directions have been investigated this year on this topic. For the human tracking task, we have proposed several approaches such as using the tracking context (crowd density) to adapt the association parameters of human detections, or learn them from intermediate tracklet results, leading to state-of-the-art results on PETS benchmark data and on the Torino metro project data. On the learning side, we have proposed a novel algorithm designed for the joint adaptation of several classifiers whose output are coupled. It was applied to the joint estimation of head and body orientations, which are coupled by physical constraints and are also related to the motion direction when people are walking. Samples results are shown in the top of Fig. 5.

Non-verbal behavior extraction

Our long-standing work on human interaction modelling has been continued in the context of several projects including the EU FP7 HUMAVIPS project, and the TRACOME and the SONVB funded by SNSF. HUMAVIPS seeks to endow humanoid robots with basic social skills necessary to deal with small

groups of people. In this project, we have improved our multi-person realtime head tracking and pose estimation algorithm to handle low quality sensors, sensor motion (when the robot nods, turns the head etc) and variable situations encountered in the Human-Robot interface (HRI) context, and are working on attention modeling by studying different (non-linear, switching) gaze models relating sequences of Visual Focus of Attention to head pose sequences. Research in the project focused on the maintaining of persons identity over time and use of stereo information for the tracking part, and on the use of dialog sensitive contextual information for attention modeling.

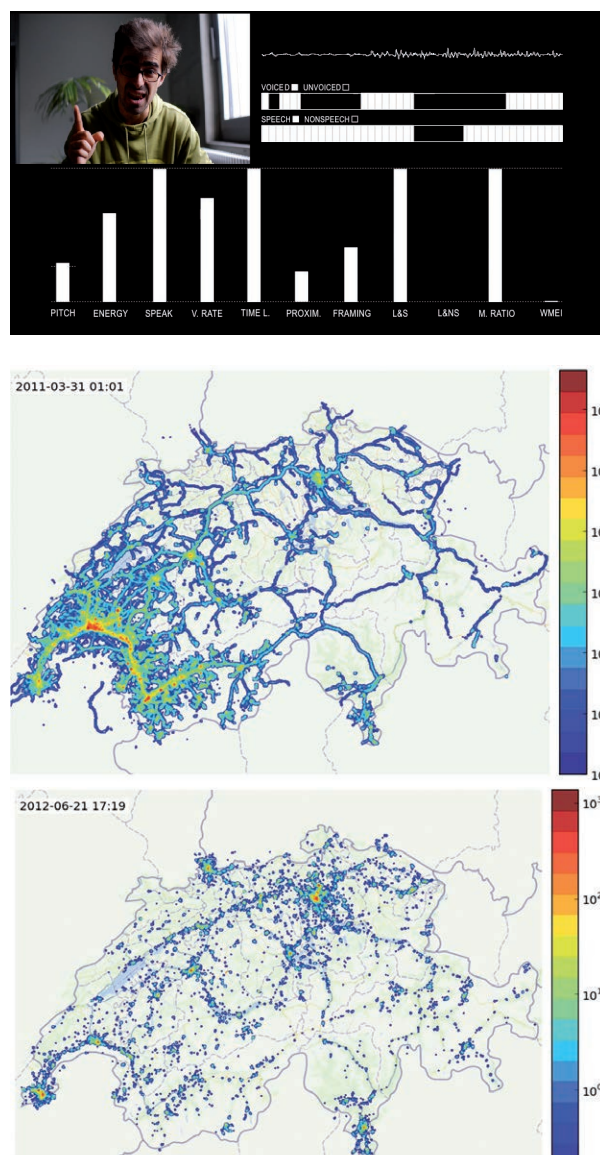


Figure 4 : Top: The VlogSense concept: Online vloggers are characterized by a variety of nonverbal behavioral features. Center: Heatmap of the Lausanne smartphone location data in Switzerland. Bottom: Heatmap of check-ins in Switzerland using six months of Foursquare data.

25 <http://www.idiap.ch/scientific-research/projects/human-activity-and-interactivity-modeling>



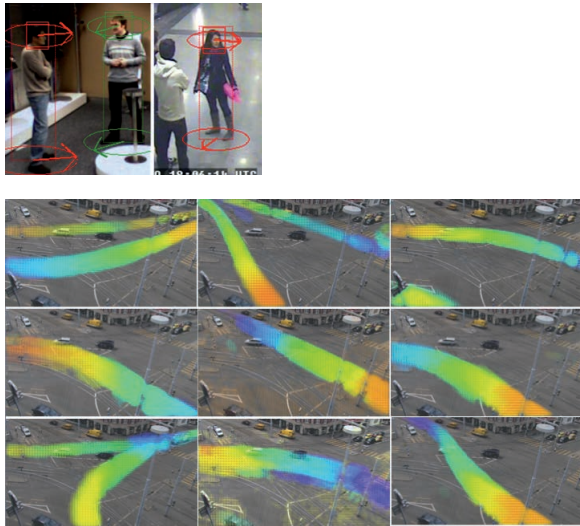


Figure 5 : Top: sample results of head and body orientation estimation. Bottom: automatic activity discovery from traffic videos. Each image shows an extracted activity motif: a sequence of observations that regularly co-occur over time. The color gradient indicates the timing since the start of the motif (from blue to red).

In SNSF project TRACOME we have investigated multimodal gaze tracking using depth sensor data. Indeed, the advent of cheap depth cameras like Kinect changed the HRI research landscape in recent years, and we are exploring the use of such device for non-verbal behavior extraction. This year, we proposed a framework to estimate gaze directions under free head motion. The depth information was exploited to accurately track the head pose using 3D deformable models, and render the face and eyes as if they were frontal, while the RGB image of the eye/pupil is used to estimate the gaze direction, as illustrated in Fig. 6. Despite the relatively low resolution of the image, errors from 10 to 15 degrees under complex head pose were achieved. The work received the best student paper award at the CVPR Kinect Gesture challenge workshop. Finally, in the context of the SONVB project that investigates new analysis approaches for automatic social inference in face-to-face organizational scenarios, we have worked with the social computing group on head gesture recognition in natural conversations, developing a multimodal nodding detector that leverages on the audio- based self-context making use of findings in psychology establishing that the dynamics of head gestures are conditioned on whether people are speaking or not.



Figure N°6 : Top: 3D rendered mesh from depth and RGB image. Left top: using the 3D tracker output, eye regions from the RGB image can be cropped and rendered as if the head pose was frontal. Bottom: sample of recognized gaze under free head movements. Green lines materialize the ground-truth gaze directions, while the red ones represent estimated gaze.

Activity analysis

In the context of the EU FP7 VANAHEIM project and SNSF HAI project, we have continued our efforts in 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 (see 5). This is a typical situation where multiple sensor record the activities of multiple objects/people, like multimodal sensors (proximity, water, light sensors, etc.) in home automation (domotics) applications. This year, we proposed a graphical model to discover, from such multi-dimensional event time series, both global states and local rules that governs the occurrence of events. It was demonstrated in traffic scenes where global states accounts for traffic signal state (phase of a traffic cycle) while local rules could account for right of way or series of street segments a car can follow. These models can be used for abnormality detection and with ethologists from 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 comparison of the the the rankings of video streams according to the 'unusualness' or interestingness of their content made by the humans with those resulting from the automatic abnormality rating validate the approach. Finally, in the context of the SNSF project PROMOVAR project funded by SNF,



we are exploring the use of these pattern discovery techniques for the classification of action from video inputs. The difficulty there lies in the exploitation of labeled information in the learning of these patterns for improving class discrimination.

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 SNSF 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 NINAPRO²⁶ 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

²⁶ <http://www.idiap.ch/project/ninapro/>



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.

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 America²⁷, 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 investi-

²⁷ <http://ml.nec-labs.com>



gate new efficient end-to-end systems based on recurrent convolutional neural networks (see Figure 7), alleviating the need of any hand-crafted feature and allowing the modeling of complex spatial dependencies with a low inference cost

- **Spoken Term Detection.**

The DeepSTD project, funded by the Swiss Hasler Stiftung²⁸, is concerned about applying new deep learning techniques to audio processing, with a particular interest in fast spoken term detection. The year 2012 was devoted to estimating phoneme class conditional probabilities directly from raw speech signal, using convolutional neural networks.



Figure 7: 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.

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 Hasler Foundation project SODS²⁹, we wish to develop a new type of search engine distributed over available web servers (in the same spirit than YaCy³⁰), 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.

28 <http://www.haslerstiftung.ch>

29 <http://www.haslerstiftung.ch>. The SODS project is part of the SmartWorld special call.

30 <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 KeyLemon³¹ 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 2012, we delivered a first version of our multi-pose face and gender detection system (see Figure 8), which runs in real time with good performance even in adverse conditions.

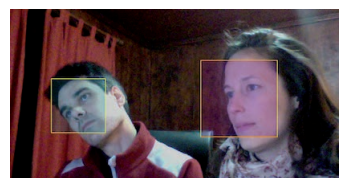


Figure 8: Our real time face detector detects multi-pose faces and labels gender (pink for female, bluish for male) in adverse conditions.

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

SENNA³² is a standalone software distributed under a non-commercial license, that we develop through our ongoing collaboration with NEC Laboratories America³³. 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 Torch³⁴ 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.

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

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

33 <http://ml.nec-labs.com>

34 <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 localization 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 9).

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 localization 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

Face and Speaker Recognition

We leveraged prior work on distribution modeling for part-based face recognition (Figure 10) using session variability modeling techniques. Session variability modeling aims to explicitly model and suppress detrimental within-class (inter-session) variation. We examined two techniques to do this, inter-session variability modeling (ISV) and joint factor analysis (JFA), which were initially developed for speaker recognition.



Figure 9: Illustration of the LBP feature map (b) for the original image (a).

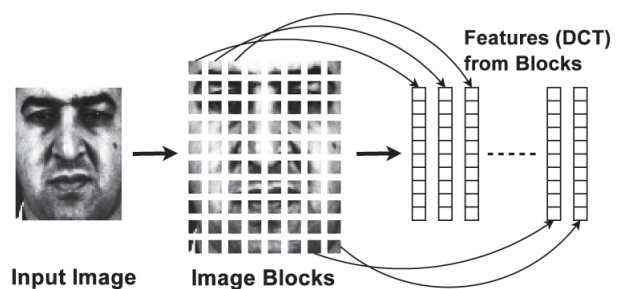


Figure 10: The concept of a parts-based approach: dividing the face into blocks and obtaining a feature vector from each block.

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 recognize identities without concern whether the sample comes from a live person or not. Despite the existence of very sophisticated biometric systems nowadays, the



task of implementing anti-spoofing schemes for them has attracted much less attention.

Recently, we advanced the state-of-the-art in anti-spoofing to printed photo attacks in face recognition by proposing public benchmarks and effective solutions based on texture modeling, motion detection or spatio-temporal pattern recognition. We also organized the first International Competition on the topic.

Reproducible Research

Over the last year the group 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 project³⁵, 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 parameterizations with minimal interaction using one or potentially many biometric traits. Here are three practical examples:

1. A teacher asks students to implement a specific illumination normalization algorithm in a pre-configured face recognition toolchain. 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 toolchain and measure its impact on the overall system performance;

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

8. Natural Language Processing

The Idiap Natural Language Processing group carries out research in semantic-level language analysis with two principal applications: machine translation (MT) and information retrieval (IR). The NLP group currently includes seven people, with additional support from engineers and interns. The overall goal of the NLP group is to improve text analysis performance by using robust semantic and pragmatic analyses in combination with state-of-the-art, low-level, data-driven techniques. These baseline techniques are, respectively, phrase-based statistical machine translation, and various content-based or collaborative filtering algorithms for computing similarity in networked data.

Machine Translation

Work on this topic, mainly within the COMTIS SNSF Sinergia project³⁶ (started in 2010 and coordinated by Idiap) focuses on a problem that is not so often targeted in the current statistical machine translation (SMT) paradigm: the translation of words that establish relationships between sentences. In collaboration with two teams in Geneva, from linguistics and computational learning, we have analyzed large collections of examples of various types of dependencies between sentences, such as rhetorical relations signaled by discourse connectives, which are problematic for current MT engines. We have proposed computational methods for the automatic disambiguation of connectives, using classifiers such as maximum entropy or decision trees based on surface features, but also on more semantic features such as synonymy or word polarity. The accuracy of our classifiers for discourse connec-

³⁵ <http://www.beat-eu.org/>

³⁶ <http://www.idiap.ch/comtis>



tives (e.g. distinguishing between temporal and causal uses of '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, without affecting the rest of the words. In a similar vein, we are dealing with pronouns and verb tenses, which also depend on dependencies across sentences, and are a hot topic in SMT research. Following an internship on pronouns in 2011, we collected data and started classification experiments on the translation of English past tenses into French, in collaboration with the University of Geneva. Data is a central issue and annotation has been carried out. An evaluation metric intended for discourse connectives has been designed, under the name of ACT, for Accuracy of Connective Translation. The metric compares the correct translation with the one generated by an SMT, in terms of identical, synonymous and incompatible translations, managing also the cases when a connective is skipped in one of the translations.

Similarity and Retrieval in Graph-based 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 AMI³⁷ and AMIDA³⁸ EU integrated projects, and the NCCR IM2³⁹. We also contribute to the recent InEvent⁴⁰ EU project, and receive funding for technology transfer in collaboration with Klewel SA.

In theoretical work, we focused on 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 from a graph node to another, and a distance measure was defined based on it. Several similarity learning models were proposed over graphs, and were applied to natural language processing tasks (such as word similarity or document clustering) and information retrieval ones, including learning-to-rank. The Automatic Content Linking Device (ACLD) serves here as a general applicative framework. The ACLD is a multimedia re-

trieval system with spoken queries (explicit or implicit) and running in real-time (e.g. to enrich a conversation) or offline (e.g. to enrich a past lecture). Several retrieval mechanisms have been explored to increase the robustness of the ACLD against ASR errors. In particular, a method based on Visiting Probability has been shown to outperform keyword-based retrieval. Evaluation metrics for this task are mainly based on binary comparisons of results by human subjects recruited using a crowdsourcing platform. Current work attempts to derive, in the ACLD setting, the optimal subset of keywords from the spoken conversation, by maintaining multiple hypotheses about the topics of the information needs of the users, in order to maximize the diversity *and* relevance of ACLD results when only a few of them can be displayed.

The NLP group currently approaches 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, to compare a number of approaches: 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, and we are now considering aspect-based sentiment analysis, as well as attempting to recommend segments rather than entire lectures.

37 <http://www.amiproject.org>

38 <http://www.amidaproject.org>

39 <http://www.im2.ch>

40 <http://www.inevent-project.eu>



MAIN PROJECTS IN PROGRESS

ACRONYM NAME, NAME

PARTNERS

EUROPEAN PROJECTS

BBFOR2

Bayesian Biometrics for Forensics

Radboud University Nijmegen
 Universidad Autonoma de Madrid
 Politecnico di Torino
 Universiteit Twente
 University of York
 Katholieke Universiteit Leuven
 Högskolan i Halmstad
 Netherlands Forensic Institute
 Agnitio Voice Biometrics
 Netherlands organisation for applied scientific research (TNO)

BEAT

Biometrics Evaluation and Testing

University of Surrey
 Universidad Autónoma de Madrid
 TÜViT
 TUBITAK Informatics and Information Security, Advanced Technologies Research Center
 Sagem Sécurité SA
 Katholieke Universiteit Leuven
 École Polytechnique Fédérale de Lausanne (EPFL)
 Commissariat à l'Energie Atomique

HUMAVIPS

Humanoids with Auditory and Visual Abilities in Populated Spaces

Institut National de Recherche en Informatique et Automatique
 The Czech Technical University
 Aldebaran Robotics
 Bielefeld University

InEvent

Accessing Dynamic Networked Multimedia Events

Radvision Ltd
 IBM Israël
 Klewel SA
 Fraunhofer Heinrich-Hertz-Institute
 University of Edinburgh

MASH

Massive Sets of Heuristics for Machine Learning

Centre National de la Recherche Scientifique
 Universität Potsdam
 Institut National de Recherche en Informatique et en Automatique
 The Czech Technical University

PASCAL2

Pattern Analysis, Statistical Modelling and Computational Learning

56 sites in the network

SCALE

Speech Communication with Adaptive Learning

Universität des Saarlandes
 University of Edinburgh
 University of Sheffield
 Radboud University Nijmegen
 RWTH Aachen
 Motorola Limited UK
 Philips
 Eurice



DURATION (MONTH/YEAR)**WEB****COORDINATOR****CONTACT**

01.10 – 12.13

<http://lands.let.ru.nl/bbfor2>

Stichting Katholieke Universiteit

Dr. Sébastien Marcel

03.12 – 02.16

www.beat-eu.org

Idiap Research Institute

Dr. Sébastien Marcel

02.10 – 01.13

<http://humavips.inrialpes.fr/>Institut de Recherche
en Informatique et
en AutomatiqueDr. Daniel Gatica-Perez
Dr. Jean-Marc Odobez

11.11 – 10.14

www.inevent-project.eu

Idiap Research Institute

Prof. Hervé Bourlard

01.10 – 12.12

<http://mash-project.eu>

Idiap Research Institute

Dr. François Fleuret

03.08 – 02.13

www.pascal-network.org

University of Southampton

Dr. François Fleuret

01.09 – 12.12

www.scale.uni-saarland.de

Universität des Saarlandes

Prof. Hervé Bourlard



ACRONYM NAME, NAME**PARTNERS****SSPnet**

Social Signal Processing Network

Imperial College of Science, Technology and Medicine, London
University of Edinburgh
University of Twente
Università Di Roma Tre
Queen's University Belfast
DFKI
INRIA
Université de Genève
Technische Universiteit Delft

TA2

Together Anywhere, Together Anytime

EURESCOM - European Institute for Research and Strategic Studies in Telecommunications GmbH
British Telecommunications plc
Alcatel-Lucent Bell NV
Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V.
Goldsmiths' College
Netherlands Organisation For Applied Scientific Research – TNO
The Interactive Institute II Aktiebolag
Stichting Centrum voor Wiskunde en Informatica
Ravensburger Spieleverlag GmbH
Philips Consumer Electronics BV
Limbic Entertainment GmbH
Joanneum Research Forschungsgesellschaft GmbH
Brno University of Technology

TA2-EEUTogether Anywhere, Together Anytime
- Enlarged European Union

Brno University of Technology
Joanneum Research
Forschungsgesellschaft mbH
Goldsmiths College, University of London

TABULA RASATrusted Biometrics
under Spoofing Attacks

University of Oulu
Universidad Autonoma de Madrid
University of Southampton
University of Cagliari
EURECOM
Chinese Academy of Sciences
Starlab Barcelona S.L.
MORPHO
KeyLemon SA
BIOMETRY.com AG
Centre for Science, Society and Citizenship
Sagem Sécurité SA

VANAHEIMVideo/Audio Networked Surveillance
System Enhancement through
Human-Centered Adaptive Monitoring

Multitel ASBL
Institut National de Recherche en Informatique et Automatique
Thales Communications France
Thales Italia
Gruppo Torinese Trasporti
Régie Autonome des Transports Parisiens
University of Vienna



DURATION (MONTH/YEAR)	WEB	COORDINATOR	CONTACT
02.09 – 01.14	www.sspnet.eu	Idiap Research Institute	Dr. Alessandro Vinciarelli
02.10 – 03.12	www.ta2-project.eu	Eurescom	Dr. Phil Garner
02.10 – 03.12	www.ta2-project.eu	Eurescom	Dr. Phil Garner
11.10 – 06.14	www.tabularasa-euproject.org	Idiap Research Institute	Dr. Sébastien Marcel
02.10 – 07.13	www.vanaheim-project.eu	Multitel ASBL	Dr. Jean-Marc Odobez



ACRONYM NAME, NAME**PARTNERS****SNSF PROJECTS****AMSP**

Auditory-motivated signal processing and applications to robust speech enhancement and recognition

A-MUSE

Adaptive Multilingual Speech Processing

AROLES

Automatic Recommendation of Lectures and Snippets

Kleweil

CODICES

Automatic Analysis of Mexican Codex Collections

COMTIS

Improving the Coherence of Machine Translation Output by Modeling Intersentential Relations

University of Geneva/Department of Linguistics
University of Geneva/Department of Computer Science

DASH

Object Detection with Active Sample Harvesting

DIMHA

Diarizing Massive Amounts of Heterogeneous Audio

Koemei SA

FlexASR

Flexible Grapheme-Based Automatic Speech Recognition

HAI-2010

Human Activity and Interactivity Modeling

ICS-2010

Interactive Cognitive Systems

NCCR IM2

Interactive Multimodal Information Management

Ecole Polytechnique Fédérale de Lausanne (EPFL)
University of Geneva
University of Fribourg
Swiss Federal Institute of Technology in Zurich (ETHZ)

LOBI

Low Complexity Binary Features for Robust-to-Noise Speaker Recognition

KeyLemon SA

MCM-FF

Multimodal Computational Modeling of Nonverbal Social Behavior in Face to Face Interaction

MULTIO8EXT

Multimodal Interaction and Multimedia Data Mining

NINAPRO

Non-Invasive Adaptive Hand Prosthetics

Institute of Robotics and Mechatronics
HES-SO Valais

PROMOVAR

Probabilistic Motifs for Video Action Recognition

RODI

Role based speaker diarization

Institute of Robotics and Mechatronics
HES-SO Valais



DURATION (MONTH/YEAR)	WEB	COORDINATOR	CONTACT
01.12 – 07.12		Indian Institute of Science	Dr. Mathew Magimai-Doss
10.12 – 09.15		Idiap Research Institute	Prof. Hervé Bourlard
06.12 – 05.14		Idiap Research Institute	Dr. Andrei Popescu-Belis
08.08 – 07.12		Idiap Research Institute	Dr. Daniel Gatica-Perez
03.10 – 02.13	www.idiap.ch/comtis	Idiap Research Institute	Dr. Andrei Popescu-Belis
09.12 – 08.15		Idiap Research Institute	Dr. François Fleuret
06.12 – 05.14		Idiap Research Institute	Dr. Fabio Valente
03.10 – 04.13		Idiap Research Institute	Dr. Mathew Magimai-Doss
10.10 – 09.12		Idiap Research Institute	Dr. Jean-Marc Odobez
10.10 – 09.12		Idiap Research Institute	Prof. Hervé Bourlard
01.02 – 12.13	www.im2.ch	Idiap Research Institute	Prof. Hervé Bourlard
06.12 – 05.14		Idiap Research Institute	Dr. Sébastien Marcel
11.11 – 10.14		Idiap Research Institute	Dr. Oya Aran Karakus
10.10 – 09.12		Idiap Research Institute	Prof. Hervé Bourlard
01.11 – 12.13	www.idiap.ch/project/ninapro	Idiap Research Institute	Dr. Barbara Caputo
02.12 – 07.13		Idiap Research Institute	Dr. Jean-Marc Odobez
11.11 – 10.14		Idiap Research Institute	Dr. Fabio Valente



ACRONYM NAME, NAME**PARTNERS****SIWIS**

Spoken Interaction with Interpretation in Switzerland

University of Geneva
University of Edinburgh
Eidgenössische Technische Hochschule Zürich

SONVB

Sensing and Analysing Organizational Nonverbal Behavior

University of Neuchâtel
Dartmouth College

TRACOME

Robust Face Tracking, Feature Extraction and Multimodal Fusion for Audio-Visual Speech Recognition

UBM

Understanding Brain Morphogenesis

Ecole Polytechnique Fédérale de Lausanne (EPFL)
University of Basel

VELASH

Very Large Sets of Heuristics for Scene Interpretation

VISION@HOME

Situated Vision to Perceive Object Shape and Affordances

SNSF PROJECTS (INDO-SUISSE)**CCPP**

Cross Cultural Personality Perception

University of Geneva
International Institute of Information Technology, India

HASLER FOUNDATION**ADDG2SU**

Flexible Acoustic Data-Driven Grapheme to Subword Unit Conversion

CLAS3

Cross-Lingual Adaptation for Text to Speech Synthesis

DEEPSTD

Universal Spoken Term Detection with Deep Learning

SESAME

SEarching Swiss Audio MEMories

SODS

Semantically Self-Organized Distributed Web Search

UBSL

User-Based Similarity Learning for Interactive Image Retrieval

V-FAST

Vocal-tract based Fast Adaptation for Speech Technology

ARMASUISSE**RECOD 2012**

Very low bit-rate speech coding

In addition to the above projects a number of industrial projects (CTI, The Ark) and grants are ongoing at Idiap.



DURATION (MONTH/YEAR)	WEB	COORDINATOR	CONTACT
12.12 – 11.15		Idiap Research Institute	Dr. Phil Garner
06.10 – 05.13	www.idiap.ch/project/sonvb	Idiap Research Institute	Dr. Daniel Gatica-Perez
01.11 – 12.13		Ecole Polytechnique Fédérale de Lausanne (EPFL)	Dr. Jean-Marc Odobez
01.11 – 12.13		University of Basel	Dr. François Fleuret
09.09 – 08.12		Idiap Research Institute	Dr. François Fleuret
10.11 – 09.14		Idiap Research Institute	Dr. Barbara Caputo
05.09 – 04.12	www.idiap.ch/project/ccpp	Idiap Research Institute	Dr. Alessandro Vinciarelli
01.13 – 12.15		Idiap Research Institute	Dr. Mathew Magimai-Doss
11.11 – 08.12		Idiap Research Institute	Dr. John Dines
12.11 – 11.14		Idiap Research Institute	Dr. Ronan Collobert
01.11 – 12.13		Idiap Research Institute	Dr. Fabio Valente
01.12 – 12.14		Idiap Research Institute	Dr. Ronan Collobert
08.12 – 03.13		Idiap Research Institute	Dr. François Fleuret
03.12 – 12.12		Idiap Research Institute	Dr. Phil Garner
07.12 – 10.12		Idiap Research Institute	Phil Garner



MAJOR PUBLICATIONS / CONFERENCES

This selection, from among the many publications of Idiap, illustrates the diversity of our research.

BOOKS, BOOK CHAPTERS AND JOURNAL PAPERS

Multimodal Signal Processing: Human Interactions in Meetings

Steve Renals, Hervé Bourlard, Jean Carletta and Andrei Popescu-Belis
Cambridge University Press, 2012

Together Anywhere, Together Anytime, Technologies for Intimate Interactions

Dick C. A. Bulterman, Petr Motlicek, Stefan Duffner and Danil Korchagin
Centrum Wiskunde & Informatica, 2012

Deep Learning via Semi-Supervised Embedding

Jason Weston, Frédéric Ratle, Hossein Mobahi and Ronan Collobert
in: In Neural Networks: Tricks of the Trade, Springer, 2012

Evaluation of Meeting Support Technology

Simon Tucker and Andrei Popescu-Belis
in: Multimodal Signal Processing: Human Interactions in Meetings, pages 237-252, Cambridge University Press, 2012

From Nonverbal Cues to Perception: Personality and Social Attractiveness

Alessandro Vinciarelli, Hugues Salamin, Anna Polychroniou, Gelareh Mohammadi and Antonio Origlia
in: LNCS Proceedings on Cognitive Behavioural Systems, Springer, 2012

Implementing Neural Networks Efficiently

Ronan Collobert, Koray Kavukcuoglu and Clément Farabet
in: Neural Networks: Tricks of the Trade, Springer, 2012

Multimodal Signal Processing for Meetings: an Introduction

Andrei Popescu-Belis and Jean Carletta
in: Multimodal Signal Processing: Human Interactions in Meetings, pages 1-11, Cambridge University Press, 2012

Sampling techniques for audio-visual tracking and head pose estimation

Jean-Marc Odobez and Oswald Lanz
in: Multimodal Signal Processing: Human Interactions in Meetings, pages 84-102, Cambridge University Press, 2012

Sparsity in Topic Models

Jagannadan Varadarajan, Remi Emonet and Jean-Marc Odobez
in: Practical Applications of Sparse Modeling: Biology, Signal Processing and Beyond, MIT Press, 2012

Speaker Diarization

Fabio Valente and Gerald Friedland
in: Multimodal Signal Processing: Human Interactions in Meetings, Cambridge University Press, 2012

User Requirements for Meeting Support Technology

Denis Lalanne and Andrei Popescu-Belis
in: Multimodal Signal Processing: Human Interactions in Meetings, pages 210-221, Cambridge University Press, 2012

A Fast Parts-based Approach to Speaker Verification using Boosted Slice Classifiers

Anindya Roy, Mathew Magimai.-Doss and Sébastien Marcel
in: IEEE Transactions on Information Forensics and Security, 7(1):241-254, 2012

A Method, Apparatus and Computer Program for Determining the Location of a Plurality of Speech Source

Afsaneh Asaei, Hervé Bourlard and Volkan Cevher,
in: 2012US-13/654055, 2012

A Probabilistic Approach to Mining Mobile Phone Data Sequences

Katayoun Farrahi and Daniel Gatica-Perez
in: Personal and Ubiquitous Computing, 2012

A real-time deformable detector

Karim Ali, Francois Fleuret, David Hasler and Pascal Fua
in: IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012

A Sequential Topic Model for Mining Recurrent Activities from Long Term Video Logs

Jagannadan Varadarajan, Remi Emonet and Jean-Marc Odobez
in: International Journal of Computer Vision, 2012

A Simple Continuous Pitch Estimation Algorithm

Philip N. Garner, Milos Cernak and Petr Motlicek
in: IEEE Signal Processing Letters, 2012

A Unified Structured Sparse Coding Framework for Spatio-Spectral Information Recovery

Afsaneh Asaei, Hervé Bourlard, Bhiksha Raj and Volkan Cevher
in: IEEE Transactions on Audio, Speech and Language Processing, 2012

Assessing Sparse Coding Methods for Contextual Shape Indexing of Maya Hieroglyphs

Edgar Roman-Rangel, Jean-Marc Odobez and Daniel Gatica-Perez
in: Journal of Multimedia, 7(2):179-192, 2012

Audiovisual Diarization Of People In Video Content

Elie Khoury, Christine Sénéac and Philippe Joly
in: Multimedia Tools and Applications, 2012

Automatic Attribution of Personality Traits Based on Prosodic Features

Gelareh Mohammadi and Alessandro Vinciarelli
in: IEEE Transactions on Affective Computing, 2012

Automatic Role Recognition in Multiparty Conversations: an Approach Based on Turn Organization, Prosody and Conditional Random Fields

Hugues Salamin and Alessandro Vinciarelli
in: IEEE Transactions on Multimedia, 2012



**Bridging the Gap Between Social Animal and Unsocial Machine:
A Survey of Social Signal Processing**

Alessandro Vinciarelli, Maja Pantic, Dirk Heylen, C. Pelachaud, I. Poggi,
F. D'Errico and M. Schroeder
in: IEEE Transactions on Affective Computing, 2012

**Conversation Analysis at Work: Detection of Conflict in Competitive
Discussions through Automatic Turn-Organization Analysis**

A. Pesarin, M. Cristani, V. Murino and Alessandro Vinciarelli
in: Cognitive Processing, 2012

**Cross-pollination of normalisation techniques from speaker to face
authentication using Gaussian mixture models**

Roy Wallace, Mitchell McLaren, Chris McCool and Sébastien Marcel
in: IEEE Transactions on Information Forensics and Security,
7(2):553 -- 562, 2012

Discovering Places of Interest in Everyday Life from Smartphone Data

R. Montoliu, J. Blom and Daniel Gatica-Perez
in: Multimedia Tools and Applications, 2012

**Emergent leaders through looking and speaking: from audio-visual
data to multimodal recognition**

Dairazalia Sanchez-Cortes, Oya Aran, Dinesh Babu Jayagopi,
Marianne Schmid Mast and Daniel Gatica-Perez
in: Journal on Multimodal User Interfaces, 2012

Finding Information in Multimedia Records of Meetings

Andrei Popescu-Belis, Denis Lalanne and Hervé Bourlard
in: IEEE Multimedia, 19(2):48-57, 2012

Human Interaction Discovery in Smartphone Proximity Networks

Trinh-Minh-Tri Do and Daniel Gatica-Perez
in: Personal and Ubiquitous Computing, 2012

**Improving Control of Dexterous Hand Prostheses Using Adaptive
Learning**

Tatiana Tommasi, Francesco Orabona, Claudio Castellini and
Barbara Caputo
in: IEEE TRANSACTIONS ON ROBOTICS, 2012

**Jointly Data-Driven and Phonological Knowledge-Guided Enhance-
ment of State Mapping-Based Cross-Lingual Speaker Adaptation**

Hui Liang and John Dines
in: IEEE Transactions on Audio, Speech and Language Processing,
2012

Mining Large-Scale Smartphone Data for Personality Studies

Gokul Chittaranjan, Jan Blom and Daniel Gatica-Perez
in: Personal and Ubiquitous Computing, 2012

**Multistream speaker diarization of meetings recordings
beyond MFCC and TDOA features**

Deepu Vijayasenan, Fabio Valente and Hervé Bourlard
in: Speech Communication, 54(1), 2012

**Phase AutoCorrelation (PAC) features for noise robust speech
recognition**

Shajith Ikbal, Hemant Misra, Hynek Hermansky and
Mathew Magimai.-Doss
in: Speech Communication, 54(7):867-880, 2012

**Real-time model learning using Incremental Sparse Spectrum
Gaussian Process Regression**

Arjan Gijsberts and Giorgio Metta
in: Neural Networks, 2012

Regularized Bundle Methods for Convex and Non-Convex Risks

Trinh-Minh-Tri Do and Thierry Artieres
in: Journal of Machine Learning Research, 13:3539-3583, 2012

**Structured Sparsity Models for Multiparty Speech Recovery
from Reverberant Recordings**

Afsaneh Asaei, Mohammad Golbabaee, Hervé Bourlard and Volkan
Cevher
in: IEEE Transaction on Audio, Speech and Language Processing,
2012

The ICSI RT-09 Speaker Diarization System

Gerald Friedland, Adam Janin, David Imseng, Xavier Anguera, Luke
Gottlieb, Marijn Huijbregts, Mary Tai Knox and Oriol Vinyals
in: IEEE Transactions on Audio, Speech, and Language Processing,
20(2):371--381, 2012

**The TA2 Database – A Multi-Modal Database From Home
Entertainment**

Stefan Duffner, Petr Motlicek and Danil Korchagin
in: International Journal of Computer and Electrical Engineering,
4(5):670-673, 2012

**The YouTube Lens: Crowdsourced Personality Impressions
and Audiovisual Analysis of Vlogs**

Joan-Isaac Biel and Daniel Gatica-Perez
in: IEEE Transactions on Multimedia, 2012

Transcribing meetings with the AMIDA systems

Thomas Hain, Lukas Burget, John Dines, Philip N. Garner, Frantisek
Grezl, Asmaa El Hannani, Marijn Huijbregts, Martin Karafiat,
Mike Lincoln and Vincent Wan
in: IEEE Transactions on Audio, Speech, and Language Processing,
20(2):486--498, 2012

**Vocal Tract Length Normalization for Statistical Parametric
Speech Synthesis**

Lakshmi Saheer, John Dines and Philip N. Garner
in: IEEE Transactions on Audio, Speech and Language Processing,
2012

**Wordless Sounds: Robust Speaker Diarization using
Privacy-Preserving Audio Representations**

Sree Hari Krishnan Parthasarathi, Hervé Bourlard and
Daniel Gatica-Perez
in: IEEE Transactions on Audio, Speech, and Language Processing,
2012



A Multiple Hypothesis Gaussian Mixture Filter for Acoustic Source Localization and Tracking

Youssef Oualil, Friedrich Faubel and Dietrich Klakow
in: 13th International Workshop on Acoustic Signal Enhancement, pages 233-236, 2012

A TDOA Gaussian Mixture Model for Improving Acoustic Source Tracking

Youssef Oualil, Friedrich Faubel, Mathew Magimai.-Doss and Dietrich Klakow
in: 20th European Signal Processing Conference, 2012

A tree-based distance between distributions: application to classification of neurons

Rival Lefort and Francois Fleuret
in: ICASSP 2012 : IEEE International Conference on Acoustics, Speech and Signal Processing, 2012

Acoustic Data-driven Grapheme-to-Phoneme Conversion using KL-HMM

Ramya Rasipuram and Mathew Magimai.-Doss
in: Proceedings IEEE International Conference on Acoustics, Speech and Signal Processing, 2012

An Agent-Based Focused Crawling Framework for Topic- and Genre-Related Web Document Discovery

Nikolaos Pappas, Georgios Katsimpras and Efstathios Stamatatos
in: 24th IEEE International Conference on Tools with Artificial Intelligence, Athens, Greece, IEEE, 2012

An Open Source Framework for Standardized Comparisons of Face Recognition Algorithms

Manuel Günther, Roy Wallace and Sébastien Marcel
in: Computer Vision - ECCV 2012. Workshops and Demonstrations, Idiap Research Institute, Heidelberg, pages 547-556, Springer Berlin, 2012

Annotation and Recognition of Personality Traits in Spoken Conversations from the AMI Meetings Corpus

Fabio Valente, Samuel Kim and Petr Motlicek
in: Proceedings of Interspeech 2012, 2012

Assessing the Impact of Language Style on Emergent Leadership Perception from Ubiquitous Audio

Dairazalia Sanchez-Cortes, Petr Motlicek and Daniel Gatica-Perez
in: Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia, Ulm, Germany, 2012

Automatic detection of conflict escalation in spoken conversations

Samuel Kim, Sree Harsha Yella and Fabio Valente
in: INTERSPEECH, ISCA, Portland, Oregon, USA., 2012

Automatic detection of conflicts in spoken conversations: ratings and analysis of broadcast political debates

Samuel Kim, Fabio Valente and Alessandro Vinciarelli
in: Proceedings IEEE International Conference on Acoustics, Speech and Signal Processing, Kyoto, Japan, 2012

Automatic Speaker Role Labeling in AMI Meetings: Recognition of Formal and Social Roles,

A. Sapru and Fabio Valente
in: Proceedings IEEE International Conference on Acoustics, Speech and Signal Processing, Kyoto, Japan, 2012, 2012

Baseline Multimodal Place Classifier for the 2012 Robot Vision Task, Jesus Martinez-Gomez, Ismael

Garcia-Varea and Barbara Caputo
in: Working Notes of the ImageCLEF 2012 Laboratory, 2012

Beyond Dataset Bias: Multi-task Unaligned Shared Knowledge Transfer

Tatiana Tommasi, Novi Quadrianto, Barbara Caputo and Christoph H. Lampert
in: Asian Conference on Computer Vision, 2012

Bi-Modal Authentication in Mobile Environments Using Session Variability Modelling

Petr Motlicek, Laurent El Shafey, Roy Wallace, Chris McCool and Sébastien Marcel
in: Proceedings of the 21st International Conference on Pattern Recognition, 2012

Bi-Modal Person Recognition on a Mobile Phone: using mobile phone data

Chris McCool, Sébastien Marcel, Abdenour Hadid, Matti Pietikainen, Pavel Matejka, Jan Cernocky, Norman Poh, J. Kittler, Anthony Larcher, Christophe Levy, Driss Matrouf, Jean-François Bonastre, Phil Tresadern and Timothy Cootes
in: IEEE ICME Workshop on Hot Topics in Mobile Multimedia, 2012

Bob: a free signal processing and machine learning toolbox for researchers

André Anjos, Laurent El Shafey, Roy Wallace, Manuel Günther, Chris McCool and Sébastien Marcel
in: Proceedings of the ACM Multimedia Conference, 2012

Boosting localized binary features for speech recognition

Anindya Roy, Mathew Magimai.-Doss and Sébastien Marcel
in: Symposium on Machine Learning in Speech and Language Processing (MLSLP), 2012

Boosting under-resourced speech recognizers by exploiting out of language data - Case study on Afrikaans

David Imseng, Hervé Bourlard and Philip N. Garner
in: Proceedings of the 3rd International Workshop on Spoken Languages Technologies for Under-resourced Languages, Cape Town, pages 60-67, 2012

Bridging the Past, Present and Future: Modeling Scene Activities From Event Relationships and Global Rules

Jagannadan Varadarajan, Remi Emonet and Jean-Marc Odobez
in: IEEE Conference on Computer Vision and Pattern Recognition, 2012, Providence, Rhode Island, USA, 2012

Building the NinaPro Database: a Resource for the Biorobotics Community

Manfredo Atzori, Arjan Gijsberts, Simone Heynen, Anne-Gabrielle Mittaz Hager, Olivier Deriaz, Patrick van der Smagt, Claudio Castellini, Barbara Caputo and Henning Müller
in: Proceedings of the Fourth IEEE RAS/EMBS International Conference on Biomedical Robotics and Biomechatronics, 2012

Checking In or Checked In: Comparing Large-Scale Manual and Automatic Location Disclosure Patterns

Eric Malmi, Trinh-Minh-Tri Do and Daniel Gatica-Perez
in: Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia, Ulm, Germany, 2012



Collecting data for socially intelligent surveillance and monitoring approaches: the case of conflict in competitive conversations

Alessandro Vinciarelli, Samuel Kim, Fabio Valente and Hugues Salamin
in: International Symposium on Communications, Control, and Signal Processing, 2012

Combination of Sparse Classification and Multilayer Perceptron for Noise Robust ASR

Yang Sun, Mathew Magimai.-Doss, Jort F. Gemmeke, B. Cranen, Louis ten Bosch and Lou Boves
in: Proceedings of Interspeech, 2012

Combining Acoustic Data Driven G2P and Letter-to-Sound Rules for Under Resource Lexicon Generation

Ramya Rasipuram and Mathew Magimai.-Doss
in: Proceedings of Interspeech, Portland, Oregon, 2012

Combining Cepstral Normalization and Cochlear Implant-Like Speech Processing for Microphone Array-Based Speech Recognition

Cong-Thanh Do, Mohammad J. Taghizadeh and Philip N. Garner
in: Proceedings of the IEEE Workshop on Spoken Language Technology, 2012

Combining transcription-based and acoustic-based speaker identifications for broadcast news

Elie Khoury, Antoine Laurent, Sylvain Meignier and Simon Petitrenaud
in: IEEE International Conference on Acoustics, Speech, and Signal Processing, 2012

Combining Vocal Tract Length Normalization with Hierarchical Linear Transformations

Lakshmi Saheer, Junichi Yamagishi, Philip N. Garner and John Dines
in: Proceedings in International conference on Speech and Signal processing, Kyoto, Japan, pages 4493-4496, IEEE SPS (ICASSP), 2012

Comparing different acoustic modeling techniques for multilingual boosting

David Imseng, John Dines, Petr Motlicek, Philip N. Garner and Hervé Bourlard
in: Proceedings of Interspeech, Portland, Oregon, 2012

Computational Methods For Structured Sparse Component Analysis of Convolutional Speech Mixtures

Afsaneh Asaei, Michael E. Davies, Hervé Bourlard and Volkan Cevher
in: Proceedings IEEE International Conference on Acoustics, Speech and Signal Processing, 2012

Contextual Conditional Models for Smartphone-based Human Mobility Prediction

Trinh-Minh-Tri Do and Daniel Gatica-Perez
in: Proceedings of the 14th ACM International Conference on Ubiquitous Computing, 2012

Conversion of Recurrent Neural Network Language Models to Weighted Finite State Transducers for Automatic Speech Recognition

Gwénoél Lecorvé and Petr Motlicek
in: Proceedings of Interspeech, Portland, Oregon, USA, pages to appear, 2012

Crowdsourcing Micro-Level Multimedia Annotations: The Challenges of Evaluation and Interface

Sunghyun Park, Gelareh Mohammadi, Ron Artstein and Louis-Philippe Morency
in: Proceedings of International ACM Workshop on Crowdsourcing for Multimedia, 2012

Detecting and Labeling Folk Literature in Spoken Cultural Heritage Archives using Structural and Prosodic Features

Fabio Valente and Petr Motlicek
in: IEEE Content Based Multimedia Indexing, 2012

DiarTk : An Open Source Toolkit for Research in Multistream Speaker Diarization and its Application to Meetings Recordings

Deepu Vijayaseenan and Fabio Valente
in: Proceedings of Interspeech, 2012

Discourse-level Annotation over Europarl for Machine Translation: Connectives and Pronouns

Andrei Popescu-Belis, Thomas Meyer, Jeevanthi Liyanapathirana, Bruno Cartoni and Sandrine Zufferey
in: Proceedings of the eighth international conference on Language Resources and Evaluation (LREC), pages 5, 2012

Empirical validations of multilingual annotation schemes for discourse relations

Sandrine Zufferey, Liesbeth Degand, Andrei Popescu-Belis and Ted Sanders
in: 8th Joint ACL-ISO Workshop on Interoperable Semantic Annotation, 2012

Exact Acceleration of Linear Object Detectors

Charles Dubout and Francois Fleuret
in: Proceedings of the European Conference on Computer Vision, 2012

Experiences in the Creation of an Electromyography Database to Help Hand Amputated Persons

Manfredo Atzori, Arjan Gijsberts, Simone Heynen, Anne-Gabrielle Mittaz Hager, Claudio Castellini, Barbara Caputo and Henning Müller
in: Proceedings of the 24th European Medical Informatics Conference, 2012

Extracting Directional and Comparable Corpora from a Multilingual Corpus for Translation Studies

Bruno Cartoni and Thomas Meyer
in: Proceedings of the eighth international conference on Language Resources and Evaluation (LREC), Istanbul, TR, pages 6, 2012

Extracting Informative Textual Parts from Web Pages Containing User-Generated Content

Nikolaos Pappas, Georgios Katsimpras and Efstathios Stamatatos
in: 12th International Conference on Knowledge Management and Knowledge Technologies, ACM ICPS, Graz, Austria, pages 4:1-4:8, ACM, 2012

Extracting Mobile Behavioral Patterns with the Distant N-Gram Topic Model

Katayoun Farrahi and Daniel Gatica-Perez
in: Proceedings of the IEEE International Symposium on Wearable Computers, Newcastle, 2012

Face Recognition with Disparity Corrected Gabor Phase Differences

Manuel Günther, Dennis Haufe and Rolf P. Würtz
in: Artificial Neural Networks and Machine Learning, Heidelberg, pages 411-418, Springer Berlin, 2012

Face Verification using Gabor Filtering and Adapted Gaussian Mixture Models

Laurent El Shafey, Roy Wallace and Sébastien Marcel
in: Proceedings of the 11th International Conference of the Biometrics Special Interest Group, Darmstadt, Germany, pages 397-408, GI-Edition, 2012



FaceTube: predicting personality from facial expressions of emotion in online conversational video

Joan-Isaac Biel, Lucia Teijeiro-Mosquera and Daniel Gatica-Perez
in: Proceedings International Conference on Multimodal Interfaces (ICMI-MLMI), 2012

From Speech to Personality: Mapping Voice Quality and Intonation into Personality Differences

Gelareh Mohammadi, Antonio Origlia, Maurizio Pili and Alessandro Vinciarelli
in: in Proceedings of ACM Multimedia 2012, 2012

Gaze Estimation From Multimodal Kinect Data

Kenneth Funes and Jean-Marc Odobez
in: IEEE Conference in Computer Vision and Pattern Recognition, Workshop on Gesture Recognition, Providence, RI, USA, 2012

Impact du degré de supervision sur l'adaptation à un domaine d'un modèle de langage à partir du Web

Gwénolé Lecorvé, John Dines, Thomas Hain and Petr Motlicek
in: Actes de la conférence conjointe JEP-TALN-RECITAL 2012, Grenoble, France, pages 193-200, ATALA/AFCP, 2012

Improving Acoustic Based Keyword Spotting Using LVCSR Lattices

Petr Motlicek, Fabio Valente and Igor Szoke
in: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE, Japan, pages 4413-4416, 2012

Indoor Scene Recognition using Task and Saliency-driven Feature Pooling

Marco Fornoni and Barbara Caputo
in: Proceedings of the British Machine Vision Conference, Guildford, UK, 2012

Investigating the Midline Effect for Visual Focus of Attention Recognition

Samira Sheikhi and Jean-Marc Odobez
in: Int Conf. on Multimodal Interaction (ICMI), Santa Monica, 2012

Iterative Relevance Feedback with Adaptive Exploration/Exploitation Trade-off

Nicolae Suditu and Francois Fleuret
in: Proceedings of the 21st ACM Conference on Information and Knowledge Management, pages 1323-1331, 2012

Joint Detection and Localization of Multiple Speakers using a Probabilistic Interpretation of the Steered Response Power

Youssef Oualil, Mathew Magimai.-Doss, Friedrich Faubel and Dietrich Klakow
in: Statistical and Perceptual Audition Workshop, 2012

LBP-TOP based countermeasure against face spoofing attacks

Tiago de Freitas Pereira, André Anjos, José Mario De Martino and Sébastien Marcel
in: International Workshop on Computer Vision With Local Binary Pattern Variants - ACCV, pages 12, 2012

Leveraging over prior knowledge for online learning of visual categories

Tatiana Tommasi, Francesco Orabona, Mohsen Kaboli and Barbara Caputo
in: Proceedings of the British Machine Vision Conference, 2012

Linking Speaking and Looking Behavior Patterns with Group Composition, Perception, and Performance

Dinesh Babu Jayagopi, Dairazalia Sanchez-Cortes, Kazuhiro Otsuka, Junji Yamato and Daniel Gatica-Perez
in: Proceedings of the International Conference on Multimodal Interaction (ICMI), Santa Monica, USA, 2012

Machine Translation of Labeled Discourse Connectives

Thomas Meyer, Andrei Popescu-Belis, Najeh Hajlaoui and Andrea Gesmundo
in: Proceedings of the Tenth Biennial Conference of the Association for Machine Translation in the Americas (AMTA), pages 10, 2012

Macro-Action Discovery Based on Change Point Detection and Boosting

Leonidas Lefakis and Francois Fleuret
in: International Conference on Machine Learning and Applications, 2012

MediaParl: Bilingual mixed language accented speech database

David Imseng, Hervé Boudlard, Holger Caesar, Philip N. Garner, Gwénolé Lecorvé and Alexandre Nanchen
in: Proceedings of the 2012 IEEE Workshop on Spoken Language Technology, pages 263--268, 2012

Microphone Array Beampattern Characterization for Hands-free Speech Applications

Mohammad J. Taghizadeh, Philip N. Garner and Hervé Boudlard
in: IEEE 7th Sensor Array and Multichannel Signal Processing Workshop(SAM), Hoboken, NJ, USA, pages 473-476, 2012

Modeling dominance effects on nonverbal behaviors using granger causality

Kyriaki Kalimeri, Bruno Lepri, Oya Aran, Dinesh Babu Jayagopi, Daniel Gatica-Perez and Fabio Pianesi
in: Proceedings of International Conference on Multimodal Interaction, ICMI 2012, Santa Monica, CA, 2012

Multimodal Cue Detection Engine for Orchestrated Entertainment

Danil Korchagin, Stefan Duffner, Petr Motlicek and Carl Scheffler
in: Proceedings International Conference on MultiMedia Modeling, Klagenfurt, Austria, 2012

On Speaker-Independent Personality Perception and Prediction from Speech

Polzehl Tim, Schoenenberg Katrin, Moller Sebastian, Metz Florian, Gelareh Mohammadi and Alessandro Vinciarelli
in: Proceedings of INTERSPEECH 2012, 2012

On the Challenge of Classifying 52 Hand Movements from Surface Electromyography

Ilija Kuzborskij, Arjan Gijsberts and Barbara Caputo
in: 34th Annual Conference of the IEEE Engineering in Medicine & Biology Society, 2012

On the Effectiveness of Local Binary Patterns in Face Anti-spoofing

Ivana Chingovska, André Anjos and Sébastien Marcel
in: Proceedings of the 11th International Conference of the Biometrics Special Interest Group, 2012

Overview of the ImageCLEF 2012 Robot Vision Task

Jesus Martinez-Gomez, Ismael Garcia-Varea and Barbara Caputo
in: Working Notes of the ImageCLEF 2012 Laboratory, 2012



Predicting the Conflict Level in Television Political Debates: an Approach Based on Crowdsourcing, Nonverbal Communication and Gaussian Processes

Samuel Kim, Maurizio Filippone, Fabio Valente and Alessandro Vinciarelli
in: ACM Multimedia, 2012

Reading Companion: The Technical and Social Design of an Automated Reading Tutor

Arthur Kantor, Milos Cernak, Jiri Havelka, Sean Huber, Jan Kleindienst and Doris B. Gonzalez
in: Workshop on Child, Computer and Interaction, Portland, Oregon, U.S.A., 2012

Recognizing the Visual Focus of Attention for Human Robot Interaction

Samira Sheikhi, Vasil Khalidov and Jean-Marc Odobez
in: IEEE International Conference on Intelligent Robots and Systems (IROS) - Human Behavior Understanding Workshop (IROS-HBU), 2012

Robot-to-group Interaction in a Vernissage: Architecture & Dataset for Multi-party Dialog

David Klotz, Johannes Wienke, Britta Wrede, Sebastian Wrede, Samira Sheikhi, Dinesh Babu Jayagopi, Vasil Khalidov and Jean-Marc Odobez
in: Proceedings of 5th International Conference on Cognitive Systems, 2012

Robust triphone mapping for acoustic modelling

Milos Cernak, David Imseng and Hervé Bourlard
in: Proceedings of Interspeech, Portland, Oregon, 2012

Socio-Technical Network Analysis from Wearable Interactions

Katayoun Farrahi, Remi Emonet and Alois Ferscha
in: International Symposium on Wearable Computers, 2012

Speaker Diarization and Linking of Large Corpora

Marc Ferras and Hervé Bourlard
in: Proceedings of the IEEE Workshop on Spoken Language Technology, 2012

Speaker Diarization of Meetings based on large TDOA feature vectors

Deepu Vijayasenan and Fabio Valente
in: Proceedings of International Conference on Acoustic, Speech and Signal Processing, 2012

Speaker diarization of overlapping speech based on silence distribution in meeting recordings

Sree Harsha Yella and Fabio Valente
in: INTERSPEECH, Portland, Oregon, USA, 2012

StressSense: Detecting Stress in Unconstrained Acoustic Environments using Smartphones

Hong Lu, Mashfiqui Rabbi, Gokul Chittaranjan, Denise Frauendorfer, Marianne Schmid Mast, Andrew T. Campbell, Daniel Gatica-Perez and Tanzeem Choudhury
in: Ubicomp'12, Pittsburgh, 2012

Structured Sparse Coding for Microphone Array Location Calibration

Afsaneh Asaei, Bhiksha Raj, Hervé Bourlard and Volkan Cevher
in: SAPA-SCALE Conference, The 5th ISCA workshop on statistical and perceptual audition, 2012

Structured Sparse Component Analysis of Compressive Acoustic Measurements

Afsaneh Asaei, Hervé Bourlard and Volkan Cevher
in: Proceeding of International Speech Communication Association, 2012

Sub-Band Based Log-Energy and its Dynamic Range Stretching for Robust In-Car Speech Recognition

Weifeng Li and Hervé Bourlard
in: Proceedings of the 13th Annual Conference of the International Speech Communication Association (InterSpeech), Portland, Oregon, 2012

Supervised and unsupervised Web-based language model domain adaptation

Gwénoélé Lecorvé, John Dines, Thomas Hain and Petr Motlicek
in: Proceedings of Interspeech, Portland, Oregon, USA, pages to appear, 2012

Synthetic References for Template-based ASR using Posterior Features

Serena Soldo, Mathew Magimai.-Doss and Hervé Bourlard
in: Proceedings of Interspeech, Portland, Oregon, USA, 2012

Template-based ASR using Posterior features and synthetic references: comparing different TTS systems

Serena Soldo, Mathew Magimai.-Doss and Hervé Bourlard
in: SAPA-SCALE Conference, International Speech Communication Association, 2012

The Good, the Bad, and the Angry: Analyzing Crowdsourced Impressions of Vloggers

Joan-Isaac Biel and Daniel Gatica-Perez
in: Proceedings of AAAI International Conference on Weblogs and Social Media, 2012

The I4U Submission to the 2012 NIST Speaker Recognition Evaluation

Kong Aik Lee, Rahim Saedi, Tawfik Hasan, Tomi Kinnunen, Benoit Fauve, Pierre-Michel Bousquet, Elie Khoury, Pablo Luis Sordo Martinez, Tharmarajah Thiruvaran, Changhuai You, Padmanabhan Rajan, Van Leeuwen, Seyed Omid Sadjadi, Driss Matrouf, Laurent El Shafey, John Mason, Eliathamby Ambikairajah, Hanwu Sun, Anthony Larcher, Bin Ma, Ville Hautamäki, Cernat HaniICI, Billy Braithwaite, Gonzalez-Hautamäki Rosa, Gang Liu, Hynek Boril, Navid Shokouhi, John Hansen, Jean-François Bonastre and Sébastien Marcel
in: NIST Speaker Recognition Conference, 2012

The Idiap Speaker Recognition Evaluation System at NIST SRE 2012
Elie Khoury, Laurent El Shafey and Sébastien Marcel
in: NIST Speaker Recognition Conference, NIST, Orlando, USA, 2012

The INTERSPEECH 2012 Speaker Trait Challenge

Björn Schuller, Stefan Steidl, Anton Batliner, Elmar Nöth, Alessandro Vinciarelli, Felix Burkhardt, Rob Van Son, Felix Wenginger, Florian Eyben, Tobias Bocklet, Gelareh Mohammadi and Benjamin Weiss
in: in Proceedings of INTERSPEECH, 2012

The Mobile Data Challenge: Big Data for Mobile Computing Research

J. K. Laurila, Daniel Gatica-Perez, I. Aad, Blom J., Olivier Bornet, Trinh-Minh-Tri Do, O. Dousse, J. Eberle and M. Miettinen
in: Pervasive Computing, Newcastle, 2012



**Translating English Discourse Connectives into Arabic:
a Corpus-based Analysis and an Evaluation Metric**

Najeh Hajlaoui and Andrei Popescu-Belis

in: Fourth Workshop on Computational Approaches to Arabic Script-based Languages at Proceedings of the Tenth Biennial Conference of the Association for Machine Translation in the Americas (AMTA), 2012

**Unsupervised Activity Analysis and Monitoring algorithms for
Effective Surveillance Systems**

Jean-Marc Odobez, C. Carincotte, Remi Emonet, E. Jouneau, Sofia Zaidenberg, Bertrand Raverra, Francois Bremond and Andrea Grifoni

in: European Conference on Computer Vision, 2012

**Using Crowdsourcing to Compare Document Recommendation
Strategies for Conversations**

Maryam Habibi and Andrei Popescu-Belis

in: RecSys, Recommendation Utility Evaluation (RUE 2012), 2012

**Using KL-divergence and multilingual information to improve ASR
for under-resourced languages**

David Imseng, Hervé Bouchard and Philip N. Garner

in: Proceedings IEEE International Conference on Acoustics, Speech and Signal Processing, Kyoto, pages 4869--4872, 2012

**Using Self-Context for Multimodal Detection of Head Nods
in Face-to-Face Interactions**

Laurent Son Nguyen, Jean-Marc Odobez and Daniel Gatica-Perez

in: Proceedings of the 14th ACM International Conference on Multimodal Interaction, 2012

**Using Sense-labeled Discourse Connectives for Statistical
Machine Translation**

Thomas Meyer and Andrei Popescu-Belis

in: Proceedings of the EACL2012 Workshop on Hybrid Approaches to Machine Translation (HyTra), Avignon, FR, pages 129--138, 2012

**Using Sparse Classification Outputs as Feature Observations
for Noise Robust ASR**

Yang Sun, B. Cranen, Jort F. Gemmeke, Lou Boves, Louis ten Bosch and Mathew Magimai.-Doss

in: Proceedings of Interspeech, 2012

**We are not Contortionists: Coupled Adaptive Learning for Head
and Body Orientation Estimation in Surveillance Video**

Cheng Chen and Jean-Marc Odobez

in: IEEE International Conference on Computer Vision and Pattern Recognition, 2012

*The complete list, abstracts and full texts are available on the Idiap web site at the following address:
<http://publications.idiap.ch>*





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 www.idiap.ch

