30 YEARS idiap RESEARCH INSTITUTE

SCIENTIFIC REPORT 2021

Contents

List of Figures

1 Executive summary

Statutes and areas of activities: Founded in 1991 by the State of Valais, the City of Martigny, the Ecole Polytechnique Fédérale de Lausanne (EPFL), the University of Geneva, and Swisscom, the Idiap Research Institute[1] is an independent, non-profit, *Research Foundation* devoted to advanced research, training, developments and technology transfer in the areas of *Artificial and Cognitive Intelligence*, and is among the world-level leaders in the field.

Based in Martigny (Valais, Switzerland), the Institute is particularly active in advanced signal processing, machine learning, and knowledge management, applied to multiple domains, including voice and visual recognition, multimedia data mining, human-machine interaction, robotics, language analysis, and bio-imaging. Idiap is involved in numerous research projects at local, national and international levels. As part of its core missions, Idiap is also very active in academic training (Masters and PhD students), as well as in technology transfer, through collaboration with various industries, or through direct spin-off (and its incubator IdeArk S.A.[2].

Affiliation and budget: Since its inception, Idiap has been a completely independent research institution, although academically affiliated with EPFL and the University of Geneva. Since July 2008, Idiap and EPFL are agreeing on a "Joint Development Plan", which is revisited and signed every 4 years. The last version was signed June 26, 2021, applicable until December 2024. Considered as part of the "ETH Strategic Domain" (Research Institution of National Importance[3]), Idiap is accredited and co-funded by the Federal Government, the State of Valais, and the City of Martigny, for a total averaging 40-45% of its annual budget, the remainder coming from competitive projects and industry. With an initial budget in 1991 of around 500 KCHF, the Idiap budget has been steadily growing over the last 30 years to be today well above 14.5 MCHF/year.

Projects: Idiap is active in numerous national and international projects (with an average of around 50 projects active simultaneously), while also being active in technology transfer and research contracts with industry. From 2001 to 2013, Idiap was also the Host Institution of one of the major National Centres of Competence in Research (NCCR) on "Interactive Multimodal Information Management" (IM2[4]). Project distributions and dynamics, including statistics of acceptance rates across all funding instruments, are presented in Section 5, page 47.

Staff and publications: With 14 research groups (Section 2.3.2, page 6) and a staff of approximately 150 people mainly composed of senior researchers, R&D engineers, post-docs, and PhD students, Idiap regularly generates a large amount of high-quality peer-reviewed international scientific publications per year (in addition to an equal amount of internal Research Reports), all available from a dedicated web site[5], automatically synchronised with EPFL Infoscience[6]. As reported in Section 6, page 61, it is also comforting to note here that the number of such high-quality publications keeps increasing, also following the stimulated trend of going for more high-quality journal papers (instead of the too numerous, lower quality, conference papers). In 2021, this observation is even more true. The number of publications (170) is significantly higher than usual. The ongoing health crisis since 2020, associated with the increase in personnel, probably explains this fact.

[1] www.idiap.ch
[2] www.theark.ch
[3] https://www.sbfi.admin.ch
[4] www.im2.ch
[5] publications.idiap.ch
[6] https://infoscience.epfl.ch/

Open-source software libraries: Idiap pursues an active policy towards the open-source release of high-quality software libraries, after having been cleared through a strict quality control process and clear IPR and distribution rules. Idiap's web site[7] refers to 133 open-source libraries.

Public datasets: Idiap has always been driven towards open data and researchers are encouraged to have their data published through a FAIR repository [8] to promote reproductible science. All data distribution are centrally managed and comply with SNSF guidelines on open data as well as Swiss and European regulations on data protection (FADP and GDPR). Idiap web [9] site is currently referencing 80 publicly available datasets.

Reproducible research: All our work is always driven by fully reproducible frameworks[10] and Idiap regularly releases open source codes accompanying its publications[11].

Quality indicators: To calibrate and optimise its performance, Idiap regularly screens the progress of the Institute against multiple indicators related to its key missions, i.e., research, technology transfer, training, and competitive funding. Thus, in addition to the present Research Report, our 2021 Self-Assessment Report, provides a concise and factual picture of where we stand in terms of our organisation, group structure, human resources, quality of projects and research staff, academic and professional activities, publications, teaching and technology transfer activities.

Technology transfer is also one of Idiap's key missions. In our separate 2021 Self-Assessment report, we are thus proud to report on the excellent performance and continuous increase of our technology transfer activities, patenting, collaboration with industries, also attracting startups, as well as initiating our own spin-off companies. In the present report, Section 7, page 73, we only list our portfolio of active patents.

Despite all the difficulties encountered, all of Idiap and its researchers have succeeded in continuing to successfully develop their activities in the best friendly and constructive spirit possible, always striving for excellence, collaboration, and team spirit. Special thanks should also be addressed to the covid19@idiap.ch team, as well as to the administration and IT support teams, who helped us get through this year 2020 smoothly, without any disruption in support services to our researchers.

[7] https://www.idiap.ch/en/scientific-research/resources
[8] https://www.snf.ch/en/7GhWDP8omTMLZ00O/news/news-210122-open-research-data-which-data-repositorie
[9] https://www.idiap.ch/en/scientific-research/resources
[10] See, e.g., https://gitlab.idiap.ch/bob/bob.ip.binseg
[11] See, e.g., www.idiap.ch/software/pbdlib/

2 Idiap Structure and Research Areas

2.1 Research Areas

Idiap's research activities span five broad thematic areas, all directly related to our central themes and competencies of **Signal Processing**, **Machine Learning**, and **Social Intelligence for Society**: perceptual and cognitive systems, human and social behavior, multimodal/multimedia information processing and presentation interfaces, biometrics security and privacy, and machine learning, which are summarized, with a set of descriptive keywords in Table 1 below.

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

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

Research Areas	***Keywords***
Perceptual and cognitive systems	*Speech and audio processing, computer vision, document processing, robotics, natural language processing, machine translation, computational cognitive science*
Human and social behavior	*Social media, verbal and nonverbal communication analysis, smartphone sensing, computational social science*
Information and presentation interfaces	*Multimedia information systems, user interfaces, personalization, system evaluation, mobile HCI using big data, data driven services*
Biometrics Security and Privacy	*Face recognition, speaker recognition, vein recognition, multimodal fusion, soft-biometrics, remote photoplethysmography, presentation attack detection (anti-spoofing), morphing attack detection, deepfakes detection, template protection, privacy preservation, mobile and wearable biometrics*
Machine learning	*Statistical and neural network based machine learning, continual learning, learning over vision and language, robust learning and handling of dataset biases, responsible AI*

Table 1: Overview of Idiap research areas. These research areas are currently covered by twelve research groups, listed in Section 2.3 below, and are designed to be as pro-active as possible in covering the applications areas presented in Figure 2.

2.2 Application Areas

While Idiap's research areas (Table 1) are quite generic and have a strong transverse nature, those are also key enablers of multiple applications. The key application areas targeted at Idiap are summarized in Table 2 below. All of those application areas have a strong cross-research group nature, hence the creation of "Cross Research Groups" (CRG) discussed in Section 2.3.3

Application Areas	***Keywords***
Human-human & human-machine interaction	*Voice and gesture controlled devices and robots, hand-free control, spoken language systems, translation systems, social robotics, user profiling*
Exploitation of rich multimedia archives (audio, video, text)	*Semantic indexing, knowledge graph, object detection and recognition, audio-video content filtering (summarization and recommendation), broadcast data analysis, scanned document analysis, analysis of cultural heritage media*
Collaborative and creative systems	*Remote meeting assistance, smart meeting room, video-conferencing, multimedia indexing and access, cross-lingual collaboration, interaction analysis, dynamics of negotiation*
Healthcare and bio-engineering	*Smart management of patient data, prosthesis (hearing aids, artificial body parts, voice banking), bio-systems modeling, inclusive ICT, bio-medical document and data processing.*
Entertainment	*Multilingual gaming, remote family games, togetherness*
Mobile computing	*Signal processing for mobile platforms, mobile social networks, participatory sensing*
Security and risk management	*Biometric security, deepfakes, access control, mobile biometry, multi-sensor fusion, speaker identification, video monitoring of areas/activities, natural risk modeling, intrusion detection, crowd management*
Home automation (domotics)	*Multi-sensor activity analysis, adaptation to users' behavior, efficient use of energy, home safety and security*
Energy	*Energy grids, multiple sensor and smart meter networks, large-scale sensor data integration, modeling of behaviors to anticipate demand, safer, cheaper, and cleaner energy production*
Smart processes	*Industry 4.0, smart manufacturing, predictive maintenance, fleet management, additive manufacturing, capture and management of industrial data.*
Smart cities	*Ecology, environment management, reduction in pollution, traffic and noise, better use of roads*

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

2.3 Internal Structure

2.3.1 Overview

The main research themes, introduced in Section 2.1, are currently covered by 14 Research Groups (RG), and 4 (under development) Cross-Research Groups (CRG), presented in the Section 2.3.2). The RG and CRG leaders, explicitely named, are all *Principal Investigators (PI)* of research projects.

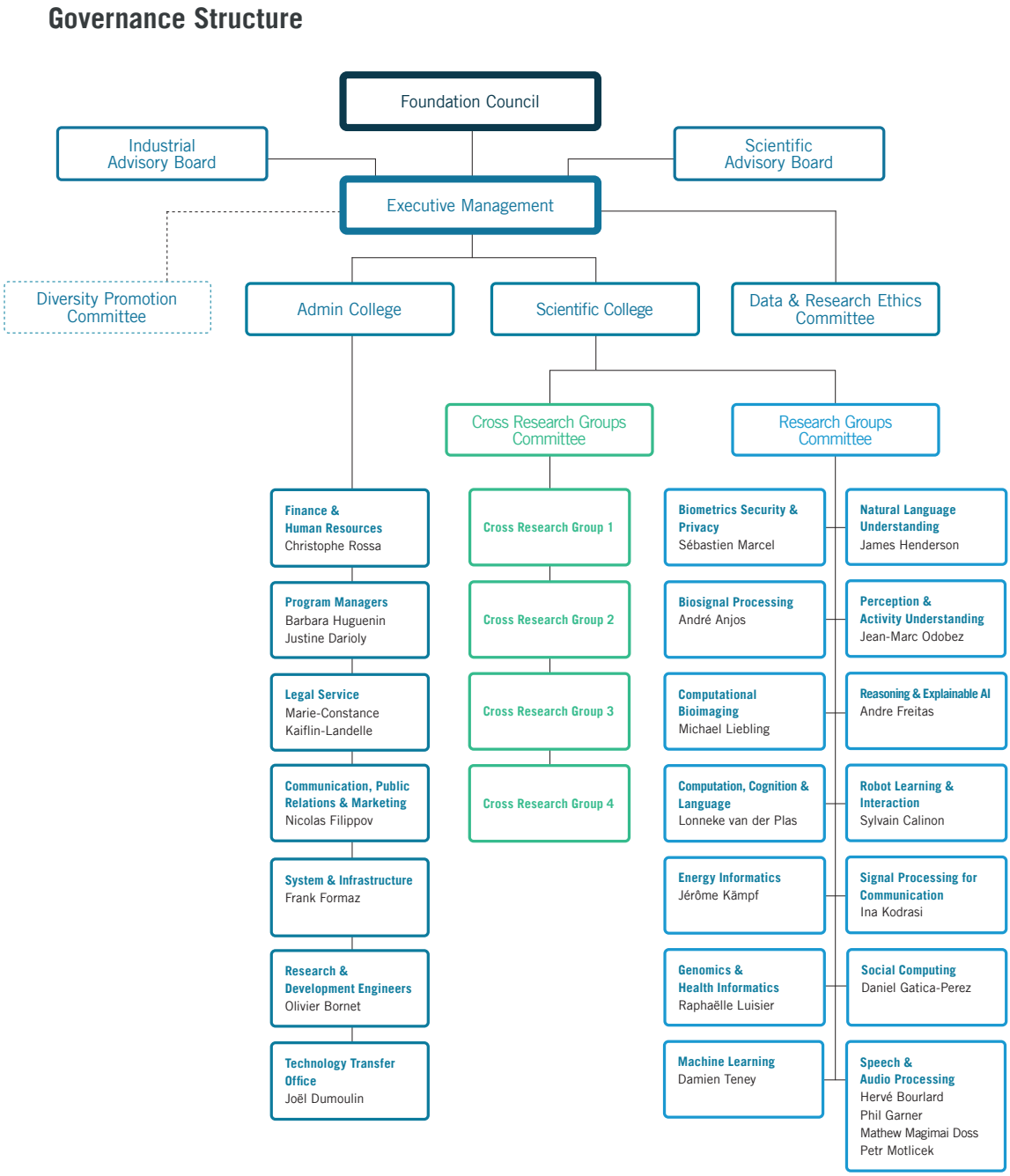


Figure 1: Idiap management and operational structure, including the 14 Research Groups on the right side, the (under-development) Cross Research Groups (CRG) at the middle, as well as the administration and services on the left side of the organigram.

2.3.2 Research Groups (RG)

In 2021, the research areas presented in Figure 1 were covered by twelve research groups. The activity of each research group (from left to right branch in Figure 1) can be summarized as follows:

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

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

   See Section 3.1, page 10, for the 2021 progress report.

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

   The social computing group studies how people and technology interface in everyday life. The group's research combines theories and methods from computing (ubiquitous computing, social media, machine learning) and the social sciences to analyze human and social behavior and design systems that support individuals and communities. Current lines of research include included mobile crowdsensing for health and cities; human-centered analysis of social media and news; and privacy-preserving machine learning.

   See Section 3.2, page 15, for the 2021 progress report.

3. **Machine Learning (Prof. François Fleuret)**

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

   See Section 3.3, page 17, for the 2021 progress report.

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

   The group aims to develop novel computational learning methods anchored in machine learning, computer vision, multimodal signal processing, or social sciences, to improve the representation and understanding of activities from real-world signals, with an emphasis on those related to humans. In particular, we investigate unsupervised learning and weakly supervised learning, user adaptation, co-training, knowledge distillation and multi-task learning, for tasks such as pose estimation, speech detection and localisation, or the recognition and analysis of non-verbal behaviors, gestures, activities or social relationships.

   See Section 3.4, page 19, for the 2021 progress report.

5. **Genomics and Health Informatics Group (Dr. Raphaëlle Luisier)**

   The Genomics and Health Informatics Group was created in 2019 to develop statistical and machine-learning methods to integrate genomic, clinical, and imaging data. The group aims to accelerate the diagnosis and improve the understanding and treatment of complex diseases such as neurodegenerative disorders in collaboration with clinical neuroscientists.

   See Section 3.5, page 21, for the 2021 progress report.

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

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

   See Section 3.6, page 23, for the 2021 progress report.

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

   The Natural Language Understanding group studies deep learning for natural language processing, focusing on learning representations of the meaning of text and attention-based models of graphs. It models summarization, abstraction (textual entailment), machine translation, knowledge extraction, syntactic structure, and lexical semantics, among other NLP problems.

   See Section 3.7, page 25, for the 2021 progress report.

8. **Computational Bioimaging (Prof. Michael Liebling)**

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

   See Section 3.8, page 27, for the 2021 progress report.

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

   The Biometrics Security and Privacy group investigates and develops novel image-processing and pattern-recognition algorithms for face recognition (2D, 3D, and near-infrared), speaker recognition, anti-spoofing (presentation attack detection), morphing attack detection, deepfakes detection and emerging biometric modes (EEG and vein).

   See Section 3.9, page 29, for the 2021 progress report.

10. **Biosignal Processing (Dr. André Anjos)**

    The Biosignal Processing group sits in the intersection between AI and Healthcare. It studies the use of Machine Learning for computer-aided diagnosis, prognosis, decision support, and risk analysis from a variety of bio-signals, including, but not limited to, imaging, clinical, demographical, and temporal data. The group has a special focus in reproducibility and the responsible use of AI, with an interest on fairness, explainability, and safety.

    See Section 3.10, page 33, for the 2021 progress report.

11. **Energy Informatics (Dr. Jérôme Kämpf)**

    The Energy Informatics Group studies the exploitation of state-of-the-art Information and Communication Technologies to tackle global warming and climate change challenges. The aim is to increase integration of renewable and distributed energy sources by making energy systems smarter, and to increase energy efficiency beyond what improvements at component level can achieve. This includes the simulation of energy transition pathways, renewable energy production, and energy storage in a changing climate.

    See Section 3.11, page 35, for the 2021 progress report.

12. **Computation, Cognition and Language (Dr. Lonneke van der Plas**

    The Computation, Cognition and Language (CCL) group was newly created in 2021. The group aims to build models of language that are inspired by human cognition and linguistic theory thereby working on the frontiers of what current models are capable of.

    Its current focus lies in cross-lingual transfer of language technology tools from one language to another language in low-resource scenarios, and modelling creative processes to create technology that supports human creativity.

    See Section 3.12, page 37, for the 2021 progress report.

13. **Reasoning and Explainable AI (Dr. André Freitas**

    Established in 2021, the Reasoning & Explainable AI (ExplAIn) group aims at developing AI models which are capable of complex, abstract and flexible inference. The group operates at the interface between neural and symbolic AI methods aiming to enable the next generation of explainable, data-efficient and safe AI models. These themes closely dialogue with the need for mitigating some of the risks for the adoption of AI models in real world scenarios.

    See Section 3.13, page 39, for the 2021 progress report.

14. **Signal Processing for Communication (Dr. Ina Kodrasi)**

    The Signal Processing for Communication group focuses on signal processing and pattern recognition approaches for acoustic communication. In particular, we develop single- and multi-channel processing and pattern recognition approaches for signal enhancement, source separation, echo suppression, pathological speech detection, and hearing support.

    See Section 3.14, page 41, for the 2021 progress report.

2.3.3 Cross Research Groups (CRG)

In Table 2, page 4, we present the main application areas, driven by *societal needs*, that can be targeted by the research currently active at Idiap.

However, over the last few years, it was also observed that it is very challenging to properly address those application areas without involving (1) multiple Research Groups (RGs), as well as (2) additional, application specific, research. Significantly contributing to those application domains, not only requires the collaboration between several research groups, but also strong multi-disciplinary focus, strong platforms (extending what we already have at Idiap), as well as interface standardization, etc.

This is a very well known challenge faced by most of the research institutions. This is why the notion of Cross Research Groups (CRG) was part of our 2021-2024 Research Program submitted to the Federal State Secretariat for Education, Research, and Innovation (SERI), which approved the opening of 4 new permanent senior positions well aligned with our application areas and CRG vision.

In 2021, we started developing this vision, investigating and testing different and complementary approaches, including:

- Outward facing: Hiring new “Group Leaders in AI applied to Industry and Societal Challenges (Cross-Research Group Leaders) F/H”, as currently advertised on our Idiap web site[12]. This search is currently still going on.
- Inward facing: Restructuring a couple of existing Research Groups (RGs), already quite well aligned with the targeted applications, and providing them with more means to achieve specific CRG goals. Internal evaluation is also currently going on.

[12] https://careers.werecruit.io/en-gb/idiap/12cd8b

2.3.4 Administration and Services

At the administration and services level, there are seven groups: finances and human resources; communication, public relations and marketing; system and infrastructure; technology transfer; development engineers; program managers; and legal advisor. The resulting admin organization is presented on the left branch of Figure 1, page 5. The main responsibilities of the seven admin groups can be summarized as follows:

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

3 Research Groups

3.1 Speech and Audio Processing

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

Group overview: Speech processing has been one of the mainstays of Idiap's research portfolio for many years, covering most of the aspects of speech processing such as multilingual automatic speech recognition (ASR), speech synthesis, speech coding (including very low bit-rate), automatic speech intelligibility evaluation, or speech processing for analysis of motor speech disorders (e.g. pathological speech). The expertise and activities of the group encompass statistical automatic speech recognition (based on hidden Markov models (HMMs), or hybrid systems exploiting deep neural networks (DNN) and new end-to-end learning architectures), text-to-speech (TTS), speaker recognition (with extensions towards text-dependent and forensics scenarios) and generic audio processing (covering sound source localization, microphone arrays, speaker diarization, audio indexing, perceptual background noise analysis for telecommunication systems) and, more recently, compressive sensing, and sparse recovery theories applied to ASR.

The Speech and Audio Processing group in 2021 was composed of 1 head of group, 3 principal investigators, 1 sabbatical academic visitor, 2 research associates, 9 postdocs, 12 PhD students, and 7 interns.

Key scientific outputs: Our primary research directions have traditionally been HMMs and DNN based approaches applied in acoustic modelling for various speech processing tasks. Techniques built around HMM and HMM-DNN architectures resulted in a unified approach used for automatic speech recognition, speech synthesis and other related classification tasks. The group is well placed to take full advantage of recent advances in new architectures of deep (i.e. end-to-end) learning, studied in particular through *PyTorch* and other open source frameworks. Advances in ASR are usually researched through *Kaldi* toolkit, now used by most of the international speech community, or its combination with other deep learning tools (particularly *Pytorch*).

In 2021, several key research contributions were achieved by the group, including: (1) multilingual automatic speech recognition, especially rapid cross-lingual adaptation, automatic speech recognition in low-resourced language conditions, with applications beyond the speech, (2) speaker recognition, through both text-independent and particularly text-dependent (i.e. for speaker verification, or authorship attribution) scenarios and information fusion for large-scale speaker identification, (3) large scale media processing, including multilingual broadcast news recognition, and information retrieval from spoken documents, (4) new compressive sensing and Sparse Recovery theories to ASR, and dualities with sparse DNN auto-encoders, (5) detection of impairments in speech signal to uncover motor speech disorders, and (5) paralinguistic speech processing with minimal prior knowledge.

Beside that, the group is also involved in the deployment of speech (including subsequent tasks) and speaker recognition algorithms for industrial applications, and is regularly involved in international evaluation campaigns.

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

Automatic speech recognition (ASR)

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

- **Robust parametrisation of ASR models**: We are investigating new features (e.g., posterior-based features) and new acoustic models (new forms of HMMs, or artificial neural networks) that are more robust to noise and acoustic environments, as well as to speaker variability (e.g., accented speech, or dialect). In the context of the recently started EC H2020 projects (ATCO2 and HAAWAII)[13], we are developing semi-supervised learning methods for rapid adaptation of speech recognition models to new (unseen) domains using unlabelled or partially labelled data.

- **Cross-lingual and multi-lingual speech recognition (including low-resource scenarios) with applications beyond speech**: From 2017, Idiap collaborates on the US IARPA SARAL project[14]. As illustrated in Figure 2, the project aims at developing cross-lingual spoken document retrieval and summarization techniques that will work for any language in the world, given minimal resources to work with. In those contexts, IDIAP focused on investigating and exploiting fast acoustic model adaptation techniques in cross-lingual and multi-lingual scenarios. Further, we collaborated with project partners on subsequent tasks processing automatically generated transcripts by NLP algorithms. The concept is also partially exploited in an industrial collaboration with Uniphore[15], a world leader in Conversational AI, to develop ASR engines for specific non-European languages.

- **Swiss languages**: We continuously improve our multilingual speech recognisers for Swiss German and Swiss French and also apply the most recent advances in speech technology employing Deep Neural Networks (DNN) and end-to-end approaches. Since 2015, we collaborate with *recapp IT AG* on a wider range of Swiss dialects towards the first commercial product that performs Swiss German (dialect) speech recognition. Idiap also participated on an recently finished Innosuisse project SM2[16] and developed a customisable technology for ASR followed by "semantic keyword and concept detection and spoken document summarization" applied to e-learning domain. From 2019 until late 2021, a collaboration with Swisscom enabled us to investigate "lexicon-free" advances in the field, which proved to be particularly suitable for dialects with no standardised orthography. This also led to work on code switching, where users change language (typically to English) during a sentence.

- **Exploiting compressive sensing and sparse recovering theories for ASR**: Through SNSF funded projects PHASER, PHASER-QUAD and SHISSM[17], Idiap is still researching new theoretical links between compressive sensing, sparse auto-encoders, and statistical/HMM-DNN approaches towards improving ASR performance and noise robustness.

Text-to-speech synthesis (TTS)

TTS has been an established venture for the speech group at Idiap for perhaps fifteen years. TTS has been central to several past projects. The group has tracked recent developments in deep learning, which continue to dominate current and anticipated research. Work on emotional speech synthesis was conducted under the NAST (neural architectures for speech technology) project. Research continued to focus on how to integrate emotional indicators based on prosody and formant position into state of the art deep

[13] https://www.atco2.org,https://www.haawaii.de
[14] https://www.idiap.ch/en/scientific-research/projects/SARAL
[15] https://www.uniphore.com
[16] https://www.idiap.ch/en/scientific-research/projects/SM2
[17] https://www.idiap.ch/en/scientific-research/projects/SHISSM

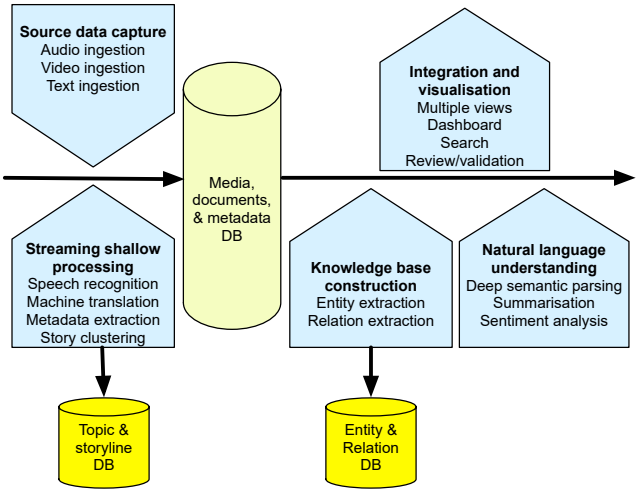


Figure 2: An overview of a typical multilingual multimedia processing stream, as used in ongoing US-DARPA SARAL project in the context of multilingual speech recognition at Idiap.

learning solutions. Results are reflected in the open-source package IdiapTTS[18], originally released in 2019.

Speaker recognition and voice biometrics

Idiap is actively carrying R&D on significantly improving capabilities of voice technologies in person identification applicable to very large scale data. Two activities were pursued in 2020: (1) combining speaker recognition with other tasks such as natural language understanding and network analysis to combat organized crime through ROXANNE H2020 project[19], (2) improving the state-of-the-art speaker technologies by integrating recent advances in machine learning (especially through the participation on NIST speaker recognition evaluations[20]), and (3) working on specific biometrics technologies applied on audio such as authorship attribution from spoken material (e.g. allowing to identify authors of spoken documents).

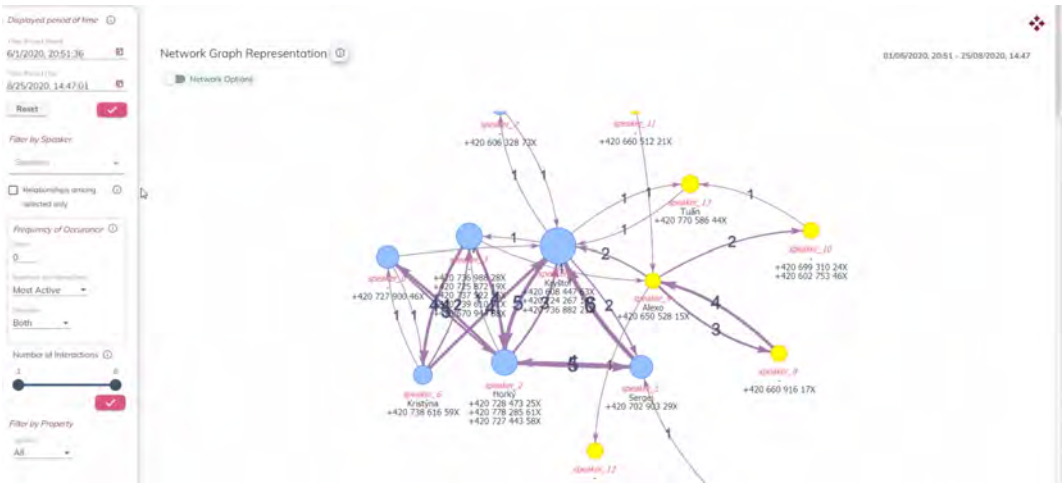


Figure 3: ROXANNE platform to process audio data and combining them with contextual information.

[18] https://github.com/idiap/IdiapTTS
[19] https://www.roxanne-euproject.org
[20] https://www.nist.gov/itl/iad/mig/speaker-recognition

Pathological speech processing

Speech and language impairments can occur due to various reasons such as, due to neurological disorders, oral cancer, hearing loss. In recent years, Idiap has been actively involved in such impaired or pathological speech processing in collaboration with clinical researchers. The SNSF Sinergia project MoSpeeDi[21] (Motor Speech Disorder) focuses on (i) developing accurate models to characterize non-impaired and impaired phonetic speech planning and motor speech programming, and ii) developing automatic techniques to detect and classify several speech impairments as well as to assess the intelligibility of patients. The EU H2020 MSCA-ITN-ETN project TAPAS[22] which is targeting three key research problems, (1) detection, (2) therapy, and (3) assisted living so that it works well for people with speech impairments and also helps in making informed clinical choices. In the recently started SNSF Bridge Discovery project EMIL[23], we are contributing towards development of paralinguistic speech processing techniques for detection and monitoring of emotion and mood in Parkinson's disease patients, and integrating them with other ambient and wearable sensor modalities for development of a new closed-loop deep brain stimulation control system.

Other research directions

- **Physiological influences**: Under the NAST project and the new NCCR Evolving Language, we are investigating both how physiological processes of perception and production can influence speech technology, and how the quite mature technology can say something about our understanding of physiology. In 2021, two distinct threads continued: in a first, stemming from the TTS work, we investigate how to combine conventional DNNs with the spiking neurons thought to be closer to the physiological function. In a second thread we investigate the latent understanding of the ear and cochlea, including the "efferent" path. The former in particular has implications for driving future ultra-low power machine learning research.

- **Sign language processing**: In the framework of recently completed SNSF Sinergia project SMILE,[24] and its follow-up SNSF Sinergia project SMILE-II[25], Idiap is developing a sign language assessment system that can assist Swiss German sign language learners in standardizing a vocabulary production test to be aligned with levels A1 and A2 of the Common European Framework of Reference for Languages. Demonstration available at vimeo.com/297803984. Our current research activities are focusing on continuous sign language processing and extension of the SMILE technology to remote online sign language assessment.

- **Sound localization and microphone array**: Idiap continues to work on distant speech processing by contributing to the Perception and Activity Understanding group through EC H2020 MuMMER project[26], focusing on audio source localization, speech detection and speaker re-identification applied in robotics.

- **Higher level semantics**: Building on past work inferring charisma from text, we embarked upon a new project with the intention of inferring storytelling from audio recordings of selection interviews. The SNSF-funded SteADI project is a collaboration with the universities of Lausanne and Neuchâtel.

[21] https://www.idiap.ch/en/scientific-research/projects/MOSPEEDI
[22] https://www.tapas-etn-eu.org
[23] https://www.idiap.ch/en/scientific-research/projects/EMIL
[24] https://www.idiap.ch/scientific-research/projects/SMILE
[25] https://www.idiap.ch/en/scientific-research/projects/SMILE-II
[26] http://mummer-project.eu

- **Joint acquisition and modeling of speech and physiological signals**: In recently funded SNSF project TIPS[27], Idiap is collaborating with CSEM[28] to develop a platform where speech and physiological signals are collected in a synchronous manner through a wearable cooperative sensor and processed to develop novel speech- and physiology-based applications. Our current research in this direction has been focusing on speech-based breathing pattern estimation using neural networks and exploitation of such pre-trained neural networks for auxiliary speech tasks such as, speech-based Covid detection.
- **Paralinguistic speech processing**: In the context of Innosuisse project CMM, SNSF Discovery project EMIL and HASLER Foundation project FLOSS, Idiap has been working on various paralinguistic speech processing tasks, namely, prediction of non-expert perceived speech fluency ratings, perceived speaker personality prediction, turn level emotion prediction and speech pseduonymization.
- **Text processing algorithms applied on automatically generated transcripts**: Many industrial applications require not only to automatically recognize the input speech with a high accuracy but also apply (or ideally combine with) subsequent modules (such as summarisation, entity recognition, retrieval) on the output of speech recognition. Among typical tasks we can count word boosting enabling to significantly increase recognition accuracy of apriori pre-defined words or entities (known to the user in advance).

Key publications

[1] **He_IEEE_2021**

[2] A. Vyas, S. Madikeri, and H. Bourlard, “Lattice-free mmi adaptation of self-supervised pretrained acoustic models,” in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Jun. 2021

[3] **Dubagunta_CSL_2022**

[4] V. S. Nallanthighal, Z. Mostaani, A. Härmä, *et al.*, “Deep learning architectures for estimating breathing signal and respiratory parameters from speech recordings,” *Neural Networks*, vol. 141, pp. 211–224, 2021

[27] https://www.idiap.ch/en/scientific-research/projects/TIPS
[28] www.csem.ch

3.2 Social Computing

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

Group overview:

The social computing group studies how people and technology interface in everyday life. The group's research combines theories and methods from computing (ubiquitous computing, social media, machine learning) and the social sciences to analyze human and social behavior and design systems that support individuals and communities.

The Social Computing group in 2021 was composed of one group head, two postdoctoral researchers, two PhD students, one scientific collaborator, and six student interns. The main research lines investigated in 2021 included mobile crowdsensing for health and cities; human-centered analysis of social media and news; and privacy-preserving machine learning.

Key scientific outputs: Publications on mobile crowdsensing to characterize youth nightlife and eating habits; human-centered social video analysis; and privacy-preserving machine learning. 16 EPFL PhD students have graduated from the group since 2002.

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

Mobile crowd-sensing for health and cities

First, in the context of the SNSF Dusk2Dawn project[29] (Characterizing Youth Nightlife Spaces, Activities, and Drinks, in collaboration with La Trobe University and the University of Zurich), we investigated the use of mobile crowdsensing to characterize urban phenomena related to nightlife. This included the study of the social context of youth during alcohol consumption at night, using smartphone sensor data (location, motion, bluetooth, and app logs) [1]. Furthermore, we investigated the methodological possibilities of analyzing audio-visual crowdsourced data to understand phenomena of interest to alcohol research, thus bringing novelty to work in public health [2]. This research was complemented with the development of privacy-preserving machine learning methods that integrate differential privacy with graph neural networks, and that could potentially be used as part of future mobile applications [3].

Second, the European H2020 WeNet project[30] is developing diversity-aware algorithms for mobile sensing to support the well-being of young adults, through a series of large-scale experiments in several European universities, as well as universities in Latin America and Asia [4]. One key motivation of this work is the advocacy for diversity in data and algorithms to improve the representation of non-western citizens. We are conducting comparative studies across countries, focused on the recognition of health-related variables like perceived food consumption levels and everyday life activities from smartphone sensing (Figure 4) [5].

Finally, regarding mobile crowdsourcing for social innovation, in the European H2020 ICARUS project[31], we are beginning to investigate human-centered approaches that can support European cities to address some of their urban security priorities. This work requires a combination of social and technological innovation approaches. This resonates with our experience using the Civique platform, which allows to collect mobile data for local causes[32]. Civique has been used in a variety of applications, ranging from supporting cities to collect information related to urban issues like street harassment, to teaching students about humanitarian technologies, to capture the experience of Swiss residents during the COVID-19 lockdown.

[29] http://www.idiap.ch/project/dusk2dawn
[30] https://www.internetofus.eu
[31] https://www.icarus-innovation.eu
[32] https://www.civique.org

Human-centered analysis of social media and news

In the context of the SNSF HealthVlogging project (in collaboration with the University of Lausanne), we investigated how human-centered, audio-visual analysis of health-related videos shared in platforms like YouTube can complement the work of health psychologists studying new practices of health promotion [6]. This line of work complements other research in our group, done in the context of the European H2020 AI4Media project [33], where we investigate human-centered approaches to understand news consumption across multiple information channels, with focus on local news sources and health-related topics.

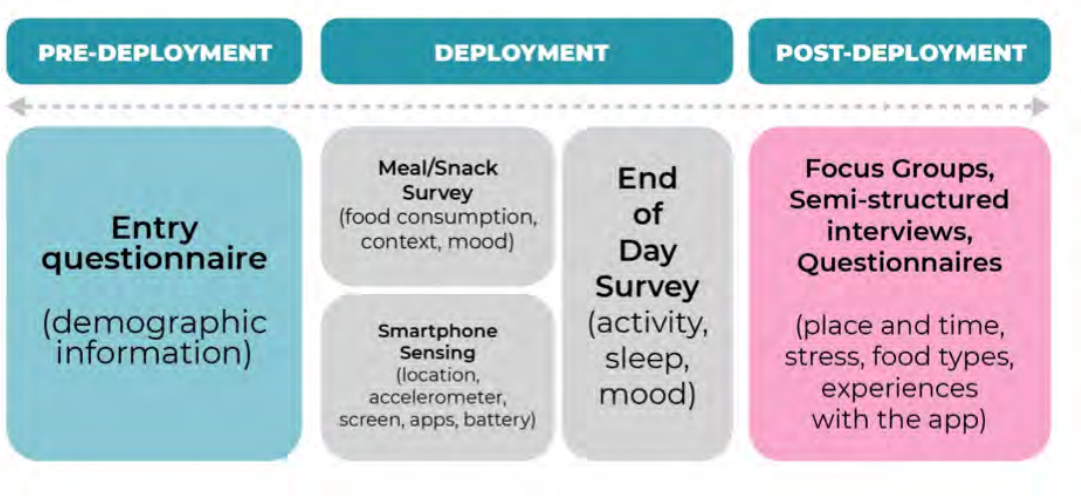


Figure 4: Smartphone-based study design to analyze food consumption (from [4])

Key publications

[1] L. Meegahapola, F. Labhart, T. T. Phan, and D. Gatica-Perez, Examining the Social Context of Alcohol Drinking in Young Adults with Smartphone Sensing, *PACM on Interactive, Mobile, Wearable, and Ubiquitous Technologies (IMWUT)*, Vol. 5, Issue 3, Article 121, Sep. 2021.

[2] F. Labhart, S. Muralidhar, B. Masse, L. Meegahapola, E. Kunstche, and D. Gatica-Perez, "Ten Seconds of My Nights: Exploring Methods to Measure Brightness, Loudness and Attendance and Their Associations with Alcohol Use from Short Videos," *PLOS ONE*, Vol. 16, No. 4, pp. 1-21, Apr. 2021.

[3] S. Sajadmanesh and D. Gatica-Perez, "Locally Private Graph Neural Networks," in *Proc. ACM Conf. on Computer and Communications Security (CCS)*, Seoul, Nov. 2021.

[4] L. Schelenz, I. Bison, M. Busso, A. de Götzen, D. Gatica-Perez, F. Giunchiglia, L. Meegahapola, and S. Ruiz-Correa, The Theory, Practice, and Ethical Challenges of Designing a Diversity-Aware Platform for Social Relations, in *Proc. AAAI/ACM Conf. on Artificial Intelligence, Ethics, and Society (AIES)*, virtual, May 2021.

[5] L. Meegahapola, S. Ruiz-Correa, V. Robledo-Valero, E. Hernandez-Huerfano, L. Alvarez-Rivera, R. Chenu-Abente, and D. Gatica-Perez, One More Bite? Inferring Food Consumption Level of College Students Using Smartphone Sensing and Self-Reports, *PACM on Interactive, Mobile, Wearable, and Ubiquitous Technologies (IMWUT)*, Vol. 5, Issue 1, Article 26, Mar. 2021.

[6] M. del Rio Carral, L. Volpato, C. Michoud, T.-T. Phan, and D. Gatica-Perez, "Professional YouTubers' Health Videos As Research Material: Formulating a Multi-Method Design in Health Psychology," *Methods in Psychology, Special Issue on Innovations in Qualitative Research*, Vol. 5, Dec. 2021.

[33] https://www.ai4media.eu

3.3 Machine Learning

Head: Dr. Damien Teney (M.Sc. University of Liège, 2009; PhD, University of Liège, 2013)

Group overview: Machine learning encompasses computer techniques that modulate their behavior according to exemplar data. It has resulted in technologies at the core of many modern every-day data-processing software and apparatus. The objective of the Machine Learning group is to develop novel machine-learning techniques of general use, applicable in particular to complex visual and textual data. The group focuses in a particular on methods able to cope with artefacts of real-world data such as noisy annotations, distribution shifts, spurious correlations, and generalization across environments. This research can be motivated by general, fundamental problems, as well as concrete industrial applications or use cases.

The machine learning group at Idiap was previously headed by Prof. Franҫois Fleuret, now at the University of Geneva. Dr. Damien Teney moved to Idiap in May 2021 and is the new head of the group. The group had been composed on average during the recent years of four PhD students, and one or two developers working on industrial applications. Dr. Teney is currently in the process of building a new team to address a new range of topics in fundamental and applied machine learning, which we outline below.

Key scientific outputs: In 2021, the machine learning group presented multiple publications at top-tier conferences in machine learning and computer vision. Our work contributed to improving the fundamental understanding of deep learning techniques. We also developed practical solutions to specific tasks such as question-answering systems over images. We also developed a state-of-the-art technique for active learning, a paradigm of great practical interest since it can reduce the costs of acquiring and labeling training data in industrial applications.

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

Robustness and generalization.

The field of machine learning has made enormous advances over the past decade. Techniques based on deep learning in particular have enabled a number of applications, but there remains however a lack of understanding of the mathematical foundations of these techniques. Our work aims to improve the understanding of the limits of applicability of deep learning, especially regarding generalization across datasets, environments, and distribution shifts. For example, data collected in varying conditions, such as images captured under different lighting conditions or with different cameras, can adversely affect the effectiveness of image recognition models. The quest for models resilient to such variations is both a practical and a fundamental one. Our recent work helped understand why deep learning techniques are prone to capture spurious correlations in training data, and how to better guide the learning process to focus on patterns that better generalize at test time.

Vision & language as a test-bed for machine learning.

We use various tasks at the intersection of computer vision and natural language processing as test-beds to evaluate our new techniques. Vision-and-language tasks constitute one of the most active and challenging application areas of machine learning. For example, we have developed models capable of automatic image captioning and visual question answering (VQA) (see Figure 5). These tasks are challenging because they require extracting and relating semantic contents from both text and images. These contributions are touching fundamental research topics in machine learning, such as multimodal embeddings, compositional generalization, while delivering techniques directly useful for the real world. Our contributions on visual question answering, for example, have been applied to medical images to help experts navigate through large repositories of medical imagery via text queries in natural language.

Figure 5: We use various tasks at the intersection of computer vision and natural language processing as a test-bed for our machine learning methods. We have developed models capable of automatic image captioning (left) as well as answering general questions about user-provided images (right). These tasks are extremely challenging since they require to extract and relate semantic contents from different modalities. Our models are entirely learned from data (e.g. hundreds of thousands of questions with their correct answer) rather than human-designed reasoning mechanisms.

Beyond statistical learning.

Most of mainstream machine learning is essentially based on learning statistical patterns, which has inherent limitations that become increasingly apparent as the tasks addressed become more and more complex. Our research investigates complementary approaches that could allow to break through these limitations. Causal reasoning offers a framework to extract models of the world beyond mere correlations. They allow drawing inferences about the effect of interventions and about the mechanisms behind the observed data. Meta learning is another line of research that attempts to "learn to learn", i.e. optimize the learning process itself. These techniques are very active and promising research areas that our group is contributing to.

As a concrete example, by framing visual question answering (see above) as a meta learning task, we learn a model capable of reasoning on-the-fly from an external source of information e.g. Wikipedia. While the traditional approach is to encapsulate all available data in a fixed model, this alternative approach allows learning reasoning mechanisms that will better generalize in the future. Such approaches promise to elevate the mainstream statistical approach to machine learning toward systems capable of robust, higher-level reasoning.

Key publications

[1] Damien Teney, Ehsan Abbasnejad and Anton van den Hengel, "Unshuffling data for improved generalization in visual question answering", in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021.

[2] Zheyuan Liu, Cristian Rodriguez-Opazo, Damien Teney and Stephen Gould, "Image Retrieval on Real-life Images with Pre-trained Vision-and-Language Models", in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021.

[3] Corentin Dancette, Remi Cadene, Damien Teney and Matthieu Cord, "Beyond question-based biases: Assessing multimodal shortcut learning in visual question answering", in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021.

[4] Damien Teney, Ehsan Abbasnejad, Simon Lucey and Anton van den Hengel, "Evading the Simplicity Bias: Training a Diverse Set of Models Discovers Solutions with Superior OOD Generalization", to appear in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2022.

[5] Armin Parvaneh, Ehsan Abbasnejad, Damien Teney, Anton van den Hengel, Javen Shi, "Active Learning by Feature Mixing", to appear in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2022.

3.4 Perception and Activity Understanding

Head: Dr. Jean-Marc Odobez (PhD, INRIA/Rennes University, France, 1994; EPFL MER)

Group overview: The group investigates novel computational learning methods from computer vision, multimodal signal processing, and machine learning for the analysis of human activity from multi-modal data. Human and human-robot interaction analysis, surveillance, traffic monitoring or multimedia content analysis are the main application domains.

Over the last 3 years, the group was composed on average of one associate researcher, two post-doctoral researchers, three PhD students, one research engineer from the development team.

Key scientific outputs: The group is known for its work on multi-object tracking, temporal motif discovery and non-verbal behavior extraction. In particular, its patended work on 3D face and gaze tracking has led to the creation of the Eyeware SA company in 2016, and the team ranked second at the ICCV 2019 Facebook Synthetic Eye Generation Challenge. In recent years, the group has investigated deep learning methods for several tasks like gesture recognition, audio-visual speaking activity modeling, gaze, audio localization and speech/non-speech detection, body landmark detection, and multimedia processing (cross-modal transfer learning, shape recognition, text localization and semantic categorization). The group also integrates its algorithms into real-time perceptual systems used in collaborative projects (HRI system for the Pepper robotic platform in the EU MuMMER project, see video), or by companies such as an anti-tailgating detection system, or our emotion recognition method for job interviews (innoswiss ADVANCE project, featured in the national TV[34]). During the period 2017-2021, the group published 15 journal papers and 35 conference papers, and filled 2 patents.

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

Human activity analysis, non-nerbal behavior (NVB) extraction, human-robot interactions

Head and body pose inference from RGB-D (color and depth) data. The team has developed a robust and accurate head pose tracking framework from RGB-D data combining the benefits of the online fitting of a 3D face morphable model with the online 3D reconstruction of the full head (Figure 6b), making head tracking a commodity for situations up to 1.5m. Research on 2D and 3D body pose estimation is also conducted (see Fig. 6a), like investigating the use of transformer models for pose predictions [1].

Gaze analytics. Due to visually unobservable gaze variabilities across people, we have investigated several methods for building user-specific models from user samples. Recently, we followed the idea of using pairs of eye images of the same person as input to predict gaze differences (differential gaze), and showed superior performance over the state-of-the-art [2]. The approach was patented by our commercial partner. We also demonstrated the possibity of designing a fully unsupervised gaze representation learning approach, allowing to leverage internet-scale data, by jointly learning a gaze representation network and gaze redirection network (see Fig. 6c) only from pairs of eye images, using as task the redirection synthesis of the second image from the first image. In the context of the ROSALIS SNSF project[35], we showed that gaze coordination priors involved in object manipulations and speaker interaction could be leveraged for obtaining weakly labeled gaze samples and robustly estimate gaze calibration models [3]. In addition, we also demonstrated that by building an explicit visual representation of the field of view of a person (people location, their speaking status, own speaking status), we could develop DNNs which defacto handle variable number of people in interactions and generalize much better across scenes and situations.

[34] Mise au point, TSR, 18/10/2020.
[35] https://www.idiap.ch/en/scientific-research/projects/ROSALIS

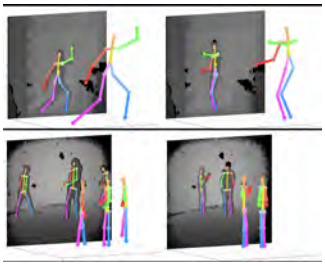
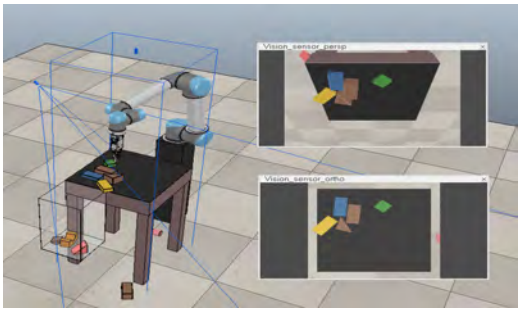
a)  b)  c) 

Figure 6: Human activity analysis. a) 3D pose estimation from depth images. b) accurate and robust 3D head and gaze tracking in adverse conditions. c) example of redirected gaze (left image: original; right one: redirected gaze towards the nose).

Multimedia and multimodal analysis, robotics

Audio analysis. Within the EU MuMMER project on social robotics, we investigated different DNN architectures and unsupervised methods for sound processing . This year, following-up on our efficient multi-task approach for the joint localization and categorization (speech vs non-speech) of multiple sources from a microphone array (see illustration in Fig. 7), we demonstrated that in addition we could jointly infer the relevant speaker embeddings, an information which would be much useful for the long term tracking of people [5].

Manipulation task policy learning. In the context of the EU HEAP project, we are investigating novel learning methods for Deep Q-learning. In particular, we studied the case of pushing tasks, which are often used in sinergy with other manipulation tasks, and demonstrated that the use of proper image-to-image translation architecture were improving current techniques. Current research involves the better exploitation of the reward structure during training, as well methods for encoding task parameters.



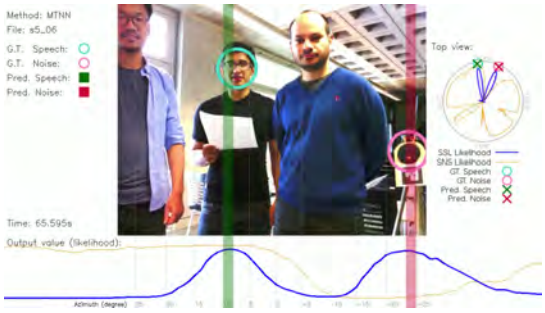


Figure 7: Left: investigating novel visual architectures to learn how to push objects (here, pushing into a box). Right: Multiple sound source localization and discrimination with a multi-task network learning to predict for each direction (circle on the right) the sound source likelihood. Sound sources are then mapped to the visual data.

Key publications

[1] A. Martínez, M. Villamizar, and J.-M. Odobez, "Pose Transformers (POTR): Human Motion Prediction With Non-Autoregressive Transformers," in *IEEE International Conference on Computer Vision (ICCV) Workshops*, 2021.

[2] G. Liu, Y. Yu, K. Funes and J.-M Odobez. "A Differential Approach for Gaze Estimation," in *IEEE Trans. Pattern Anal. Machine Intelligence (PAMI)* , Vol 43(3), 2021.

[3] R. Siegfried and J.-M. Odobez. "Robust Unsupervised Gaze Calibration using Conversation and Manipulation Attention Priors," *ACM Trans. on Multimedia Computing, Communications, and Applications*, 2021.

[4] M. Ewerton, A. Martínez and J.-M. Odobez. "An Efficient Image-to-Image Translation HourGlass-based Architecturefor Object Pushing Policy Learning," *International Conference on Intelligent Robots and Systems conference (IROS)*, 2021.

[5] W. He, P. Motlicek and J.-M. Odobez. "Multi-Task Neural Network for Robust Multiple Speaker Embedding Extraction," *Proc. of Interspeech*, 2021.

3.5 Genomics and Health Informatics

Head: Dr. Raphaëlle Luisier (Master of Science in Bioengineering and Biotechnology, EPFL, 2009 & PhD in Bioinformatics from the Basel University, 2013)

Group overview: The Genomics & Health Informatics Group was created in 2019 and is developing statistical and machine learning methods, including deep learning, to extract key information from multimodal and longitudinal biological data of various types such as genomic, clinical, and imaging data. Specifically, the group aims to address salient biological questions related to poorly understood human disorders by interpreting and integrating complex high-content data sets. The group works in close collaboration with biologists and clinicians (Patani laboratory, Francis Crick Institute, London; Serio lab, Kings College, London), thereby maintaining a fertile ground for innovation, learning and discoveries with real therapeutic prospects. The group has expertise in genomics, bioinformatics, RNA biology, neuroscience, data science, and data visualisation. Since December 2020, the group is an associate member of the *NCCR RNA & Disease*. The group is also taking part to frequent outreach activities in order to engage the public on the scientific activities led at IDIAP and also encourage young women to pursue a career in science. Such activities include, but are not limited to, interviews for radio channels (Radio Chablais, RTS CQDF), mentoring young ladies within the SwissTech ladies (SATW) framework, interview for TTC Le valais Innovant, and speaker at the Table Ronde *La Science du Bonheur* of the 2021 annual Science Valais network day.

Key scientific outputs: Since its establishment in December 2019, the Genomics & Health Informatics group has developed computational and machine learning methods to histopathological images and cellular imaging that enable the automated extraction of key information from such complex images leading to key discoveries related to neurodegenerative disorders. The methods and associated discoveries were published in leading peered-review journals in neuroscience and pathology, as briefly described below.

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

Deep learning method to robustly and efficiently test biological hypotheses

By deploying transfer learning with a pre-trained deep-learning image classifier to a rich multichannel fluorescence microscopy high content imaging data-set, the group revealed novel neurite-associated morphological perturbations in ALS disease. This method, published in *Neuropathology and Applied Neurobiology* (Verzat et al. 2021) allows to robustly and efficiently test biological hypotheses in an unbiased fashion and therefore is potentially transformational in terms of the novel phenotypic information that can be gleaned from image analysis of in vitro neurodegeneration models. In line with IDIAP incentive to promote reproducibility in research, the method has been publicly released on Gitlab (ALS classification) and the large-scale imaging dataset of 156,577 images has been publicly released in the Image Data Resource (IDR) under study idr0112.

Advance in histopathological tissue sections analysis

Histopathological analysis of tissue sections is an invaluable resource in neurodegeneration research. However, cell-to-cell variation in both the presence and severity of a given phenotype is a key limitation of this approach. The Genomics & Health Informatics group directly addressed these issues by combining automated image processing with machine learning methods to substantially improve the speed and reliability of identifying phenotypically diverse *Motor Neurons* (MNs) populations. The developed method enabled automatic identification of sick cells with unprecedented accuracy and at single cell resolution in histopathological tissue sections, and was published in *Brain Pathology* (Hagemann et al., 2021). This study showcase the potential of combining histopathology with automated image processing and machine learning and might prove transformational in our understanding of ALS and neurodegenerative diseases more broadly.

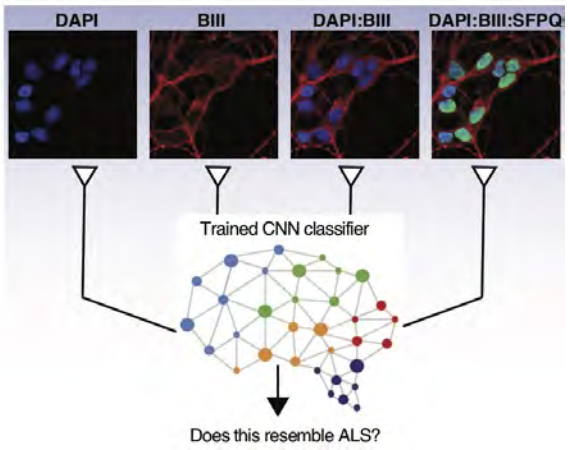


Figure 8: Consecutive training of deep learning image classifiers using a combination of several fluorescent biomarkers to discriminate control from ALS motor neuron cultures enables to identify disease cellular compartment. (fig. from C.V. et al. *Neuropathology and Applied Neurobiology, 2021)*

Discovery of a novel hallmark of Amyotrophic Lateral Sclerosis disease

Amyotrophic lateral sclerosis (ALS) is a rapidly progressive and incurable neurodegenerative disease, characterized by the progressive death of motor neurons (MNs) for which the early cellular and molecular events remain poorly understood. As a consequence, no effective treatments have been developed for this devastating disease where most patients die within 3-5 years (with 30% dying in just 12 months of diagnosis). The Genomics & Health Informatics group is working in close collaboration with the laboratory of Rickie Patani (The Francis Crick Institute, London) in order to identify the cellular and molecular events underlying the disease. In the last two years, the group has developed bioinformatic pipelines to RNA-sequencing data generated in Patani's laboratory. These methods enabled to demonstrate a progressive accumulation of intronic sequences in the cytoplasm of developing motor neurons *Brain* (Tyzack et al, 2021), together with a potential mechanism in the disease pathogenesis with real therapeutic potential *BioaRχiv* (Petric-Howe et al, 2021).

Key publications

[1] M Petric-Howe, H Crerar, J Neeves, GE Tyzack, R Patani* and R Luisier*. "Diminished miRNA activity is associated with aberrant cytoplasmic intron retention in ALS pathogenesis," in BioaRiv. *These authors contributed equally.

[2] C Verzat, J Harley, R Patani and R Luisier. "Image-based deep learning reveals the responses of human motor neurons to stress and VCP-related ALS," in *Neuropathology and Applied Neurobiology* (2021).

[3] C Andreassi*, R Luisier*, H Crerar, S Franke, NM Luscombe, G Cuda, M Gaspari and A Riccio. "3' UTR Remodelling of Localized Transcripts in Sympathetic Neurons Axons," in *Cell Reports* (2021). *These authors contributed equally.

[4] GE Tyzack, J Neeves, H Crerar, P Klein, O Ziff, DM Taha, R Luisier*, NM Luscombe* and R Patani*. "Aberrant cytoplasmic intron retention is a blueprint for RNA binding protein mislocalization in VCP-related amyotrophic lateral sclerosis," in *Brain* (2021). *These authors contributed equally.

[5] C Hagemann, GE Tyzack, DM Taha, H Devine, L Greensmith, J Newcombe, R Patani*, A Serio* and R Luisier*. "Automated and unbiased classification of motor neuron phenotypes with single cell resolution in ALS tissue," in *Brain Pathology* (2021). *These authors contributed equally.

3.6 Robot Learning and Interaction

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

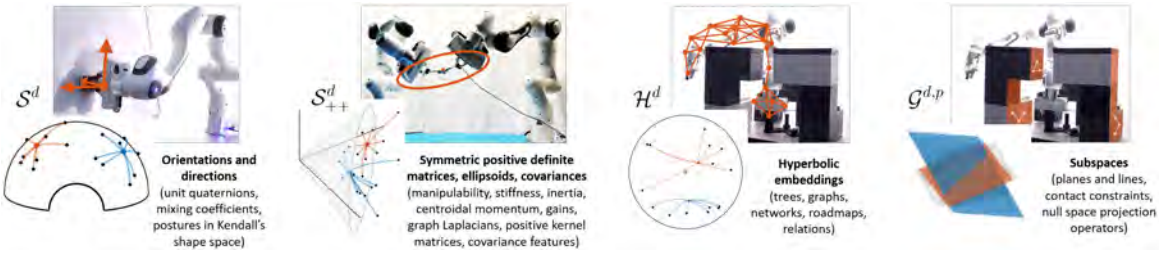
Group overview: The Robot Learning and Interaction group, created in 2014, focuses on human-centered robotics applications in which the robots can acquire new skills from only few demonstrations and interactions. It requires the development of models that can exploit data structures in an efficient way, the development of optimal control techniques that can exploit the learned task variations and coordination patterns, and the development of intuitive interfaces to acquire meaningful demonstrations. The developed approaches can be applied to a wide range of manipulation skills, with robots that are either close to us (assistive and industrial robots), parts of us (prosthetics and exoskeletons), or far away from us (teleoperation).

The Robot Learning & Interaction group in 2021 was composed of 3 postdoctoral fellows, 8 PhD students and 2 visiting PhD students.

Key scientific outputs: Development of learning techniques that only need a small number of demonstrations, by exploiting structures that can be found in a wide range of robotic tasks, and by exploiting bidirectional human-robot interaction as a way to collect better data.

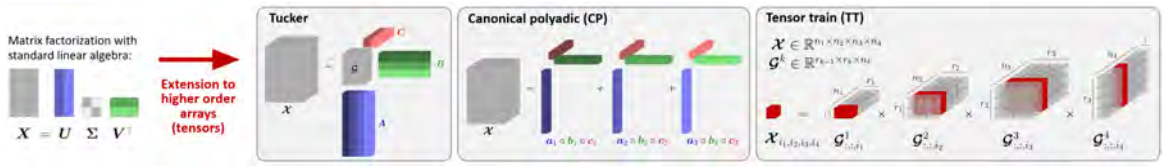
Additional information and a list of projects are available from www.idiap.ch/rli.

Geometry-aware learning and control



Data in robotics are characterized by simple but varied geometries, which are often under-exploited when developing learning and control algorithms. We exploit Riemannian geometry to extend algorithms initially developed for standard Euclidean data, by taking into account the structures of these manifolds in a unified manner.

Tensor factorization in robotics applications



Another structure that we exploit relates to the organization of data as multidimensional arrays (also called tensors). Such data appear in robotic tasks either as the natural organization of sensory/motor data (tactile arrays, images, kinematic chains), or as the result of standardized preprocessing steps (moving time windows, covariance descriptors). Tensor methods extend linear factorizations such as singular value decomposition to multilinear decompositions, without requiring the transformation of the tensors into matrices or vectors. We exploit these techniques to provide robots with the capability to learn tasks from only few tensor datapoints.

This factorization approach can be applied to learning and optimization problems characterized by two types of variables: 1) *task parameters* representing the situation that the robot encounters (locations of objects, users or obstacles); and 2) *decision variables* related to actions that the robot takes (controller acting within a given time window). For each change of task parameters, decision variables need to be recomputed as fast as possible, so that the robot can fluently collaborate with users and can swiftly react to changes in its environment. We investigate the roles of offline and online learning optimization to attain such objective, which is formalized as an *optimal control problem* with a cost function to minimize, parameterized by task parameters and decision variables. We investigate the use of *tensor train (TT) decomposition* as a model to learn the structure between the task parameters, the decision variables and the resulting cost expressed in the form of a probability distribution, which then allows solutions to be sampled from a conditional distribution. We exploit this structure to gather prior knowledge in an offline phase, which is further used for fast online decision making, with local Gauss-Newton optimization in the form of *iterative linear quadratic regulators (iLQR)*.

Combination of controllers as a product of experts

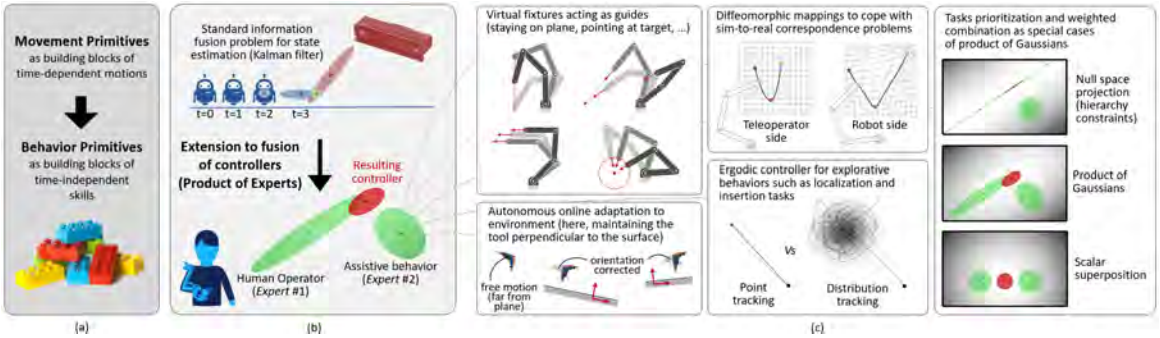


Figure 9: (a) *We extend the principle of movement primitives (superposition of basis functions as time-dependent trajectories) to behavior primitives (superposition of time-independent controllers).* (b) *State estimation in robotics is typically formulated as an information fusion problem. We propose to combine controllers in the same principled way by using a product of experts (PoE) formulation in which the user is part of the shared control problem, with the other controllers assisting the user to achieve the task.* (c) *We develop behavior primitives in both position and force domains, including the autonomous adaptation to object locations, the consideration of tasks variations and tasks prioritization, or the use of ergodic controllers for exploration behaviors.*

We formulate the problem of combining controllers as a product of experts (see see Fig. 9), where each expert takes care of a specific aspect of the task to achieve, which can be learned separately or altogether (by variational inference). With this probabilistic formulation, the robot counteracts perturbations that have an impact on the task, while ignoring other perturbations. This approach creates bridges with research in biomechanics and motor control, including minimal intervention principles, uncontrolled manifolds or optimal feedback control.

Key publications

[1] Jaquier, N., Rozo, L., Caldwell, D.G. and Calinon, S. (2021). Geometry-aware Manipulability Learning, Tracking and Transfer. International Journal of Robotics Research (IJRR), 40:2-3, 624-650.

[2] Shetty, S., Silvério, J. and Calinon, S. (2021). Ergodic Exploration using Tensor Train: Applications in Insertion Tasks. IEEE Trans. on Robotics (T-RO).

[3] Pignat, E., Silvério, J. and Calinon, S. (2021). Learning from Demonstration using Products of Experts: Applications to Manipulation and Task Prioritization. International Journal of Robotics Research (IJRR).

3.7 Natural Language Understanding

Head: Dr. James Henderson (BSc, Massachusetts Inst. Technology, USA, 1987; MSE & PhD, Univ. Pennsylvania, USA, 1991,1994; MER & Chargé de Cours, Univ. Geneva, 2008–2012,2012–2018)

Group overview: The Natural Language Understanding group (NLU) works at the intersection of machine learning and natural language processing, with an emphasis on representation learning for the meaning of language, attention-based deep learning models, and structured prediction. We model summarisation, abstraction (textual entailment), machine translation, knowledge extraction, syntactic and semantic structure, and lexical semantics, among other natural language processing problems (NLP). We develop deep learning models of the discovery and prediction of entities and their relations at multiple levels of representation for multiple tasks.

During 2021, the NLU group had the following members: the head of the group, 7 PhD students, and one postdoctoral researchers at the beginning of the year.

Key scientific outputs: During 2021, the work of the NLU group has produced several key publications. We published our work on a novel version of the Transformer deep learning architecture which maps graphs to graphs, and its application to the iterative refinement of syntactic structures. This is now the most accurate model of syntactic parsing on standard benchmarks. Three papers were published on extension of the Transformer architecture to improve its efficiency and improve multi-task learning and out-of-domain generalisation. Our novel model of summarisation which predicts embeddings for each sentence before generating the words for that sentence was also published.

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

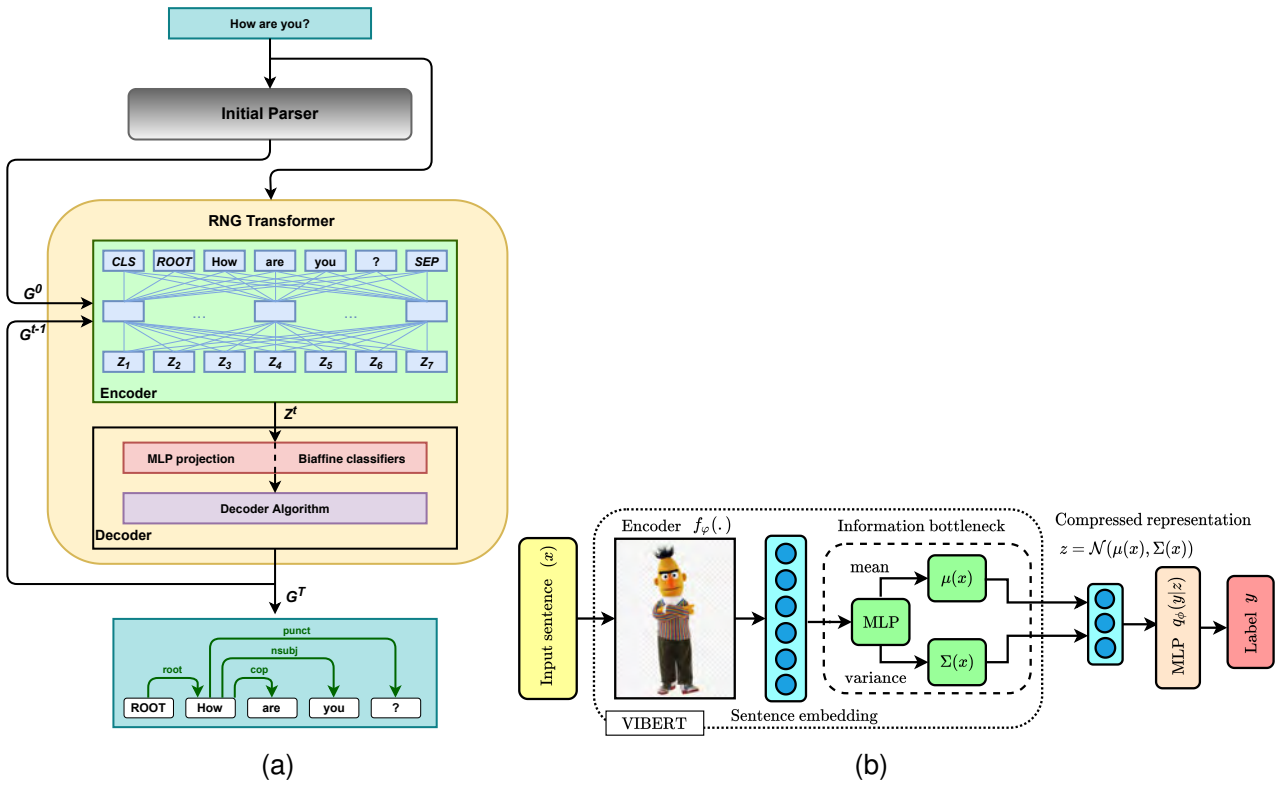


Figure 10: (a) Recursive Non-autoregressive Graph2Graph Transformer for syntactic parsing. (b) VIBERT architecture for low-resource fine-tuning of pretrained models.

Deep Learning Architectures for Graphs

Deep learning models based on self-attention, in particular Transformer, have revolutionised the state-of-the-art in many NLP tasks, particularly when combined with pretraining (e.g. BERT). We have developed a version of Transformer which maps graphs to graphs, instead of sequences to sequences. Given the nodes of a graph, such as the words of a sentence, Graph2Graph Transformer can input arbitrary graphs and output arbitrary graphs over these nodes, using the self-attention mechanism. We also developed Recursive Non-autoregressive Graph2Graph Transformer (Figure 10(a)) for the iterative refinement of predicted graphs. When combined with BERT pretraining, these models are now the most accurate models for syntactic parsing, a core benchmark for structured prediction and NLU. State-of-the-art results are also achieved in coreference resolution. We are currently investigating other structure prediction tasks and other NLP tasks which can benefit from structured inputs. An important current topic of research is developing effective architectures for inducing the nodes of the graph, instead of having them pre-specified.

Representation Learning for NLP Tasks

Deep learning models of natural language induce hidden representations which can be effectively transferred across tasks. One topic we have addressed is how to improve this transfer by reducing the models' reliance on idiosyncrasies of the training data which don't generalise to the real task, don't generalise to other domains, and are misleading given only a few training examples. We have proposed several methods to improve these kinds of generalisation, including a variational information bottleneck regulariser (Figure 10(b)), and methods using an efficient fine-tuning method called adapters.

One task which we have focused on in developing novel deep learning models is summarisation. Analogously to the hierarchical encoder models we have worked on previously, we developed an architecture for hierarchical decoding when generating summaries. It first predicts the sentence embedding of the next summary sentence, then predicts the words of that sentence, resulting in less exact copying from the source document.

Key publications

[1] Alireza Mohammadshahi and James Henderson. "Recursive Non-Autoregressive Graph-to-Graph Transformer for Dependency Parsing with Iterative Refinement". *Transactions of the Association for Computational Linguistics* (TACL), 9:120–138, 2021.

[2] Rabeeh Karimi Mahabadi, James Henderson, and Sebastian Ruder. "Compacter: Efficient Low-Rank HyperComplex Adapter Layers". In *Proc. NeurIPS 2021*, Online.

[3] Rabeeh Karimi Mahabadi, Sebastian Ruder, Mostafa Dehghani, and James Henderson. "Parameter-efficient Multi-task Fine-tuning for Transformers via Shared Hypernetworks". In *Proc. 59th Annual Meeting of the Association for Computational Linguistics* (ACL 2021), Online.

[4] Rabeeh Karimi Mahabadi, Yonatan Belinkov, and James Henderson. "Variational Information Bottleneck for Effective Low-Resource Fine-Tuning". In *Proc. Tenth International Conference on Learning Representations* (ICLR 2021) Online.

[5] Andreas Marfurt and James Henderson. "Sentence-level Planning for Especially Abstractive Summarization". In *Proc. NewSum workshop* at EMNLP 2021.

[6] Christos Theodoropoulos, James Henderson, Andrei C. Coman, Marie-Francine Moens. "Imposing Relation Structure in Language-Model Embeddings Using Contrastive Learning". In *Proc. CoNLL 2021*.

3.8 Computational Bioimaging

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

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

In 2021, the Computational Bioimaging Group was composed of the head of the group, one PhD students, and four interns.

Key scientific outputs: Recent milestones include temporal super-resolution methods for sensitive fluorescence and thermal cameras and approaches for quantitative heart development imaging in animal models.

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

Acquisition of microscopy image sequences and particle flow-estimation

As part of the SNSF project COMPBIO "Computational Biomicroscopy: Advanced Image Processing Methods to Quantify Live Biological Systems," we developed image capture techniques specifically adapted for optical microscopy of biological samples. Our research has focused on in vivo microscopy, an important tool to study developing organs such as the heart of the zebrafish embryo but which is often limited by slow image frame acquisition speed. For the application to cardiac imaging, while we previously showed that collections of still images of the beating heart at arbitrary phases can be sorted to obtain a virtual heartbeat, the presence of identical heart configurations at two or more heartbeat phases could derail this approach. We proposed a dual illumination method to encode movement in alternate frames to disambiguate heartbeat phases in the still frames. We proposed to alternately acquire images with a ramp and pulse illumination (Fig. 11) then sort all successive image pairs based on the ramp-illuminated data but use the pulse-illuminated images for display and analysis. We characterized our method on synthetic data, and showed its applicability on experimental data. We found that an exposure time of about 7% of the heartbeat or more is necessary to encode the movement reliably in a single heartbeat with a single redundant node. Our method opens the possibility to use sorting algorithms without prior information on the phase, even when the movement presents redundant frames.

In a complementary line of research, we focused on optical flow, a method aimed at predicting the movement velocity of any pixel in the image, which is used in medicine and biology to estimate flow of particles in organs or organelles. Since precise optical flow measurements requires images taken at high speed and low exposure time, collecting such images can induce phototoxicity due to the increase in illumination power. As an alternative approach, we aimed to estimate the three-dimensional movement vector field of moving out-of-plane particles using normal light conditions and a standard microscope camera. We presented a method to predict, from a single textured wide-field microscopy image, the movement of out-of-plane particles using the local characteristics of the motion blur. We estimated the velocity vector field from the local estimation of the blur model parameters using an deep neural network and achieved a prediction with a regression coefficient of 0.92 between the ground truth simulated vector field and the output of the network. This method could enable microscopists to gain insights about the dynamic properties of samples without the need for high-speed cameras or high-intensity light exposure.

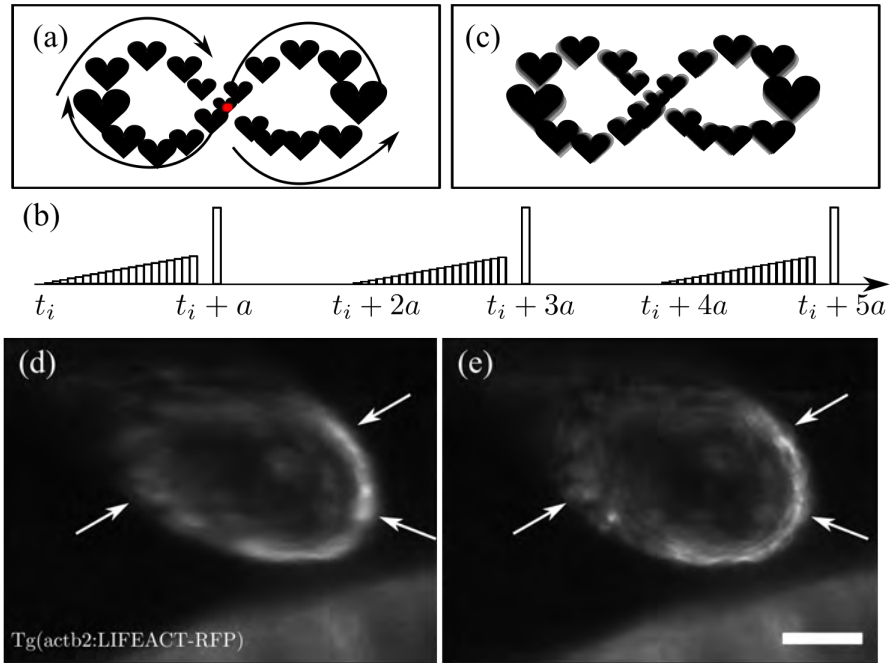


Figure 11: Structured illumination pattern removes sorting ambiguity in cardiac microscopy (a) Repeating poses within periodic trajectories (red dot) can derail image sorting methods. (b) We propose an alternating illumination pattern, where the ramp-illuminated images (c) are used to sort the pulse-illuminated images (a). (d) Image acquired with a 70ms ramp illumination. (e) Image acquired with a 4ms pulse. Arrows point to single cells that are visible in the pulse-illuminated data, but cannot be seen in the blurred ramp-illuminated data. Scalebar is 30μm.. (adapted from O.M. et al. *ISBI, 2021*

Studying the potential of visible-thermal image pair synthesis using generative adversarial networks

Many applications rely on thermal imaging to complement or replace visible light sensors in difficult imaging conditions. Recent advances in machine learning have opened the possibility of analyzing or enhancing images, yet these methods require large annotated databases. Training approaches that leverage data augmentation via simulated and synthetically-generated images could offer promising prospects. We explored the use of on a method that uses generative adversarial nets (GANs) to synthesize images of a complementary contrast. Starting from a dual-modality dataset of co-registered visible and thermal images, we trained a GAN to generate synthetic thermal images from visible images and vice versa. Our results show that the procedure yields sharp synthesized images that might be used to augment dual-modality datasets or assist in visual interpretation, yet are also subject to the limitations imposed by contrast independence between thermal and visible images.

Key publications

[1] A. Shajkofci and M. Liebling “Estimating Nonplanar Flow from 2D Motion-blurred Widefield Microscopy Images via Deep Learning,”, in: 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), Nice, France,

[2] F. Marelli and M. Liebling, “Optics Versus Computation: Influence of Illumination and Reconstruction Model Accuracy in Focal-Plane-Scanning Optical Projection Tomography,” in: 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), Nice, France, pages 567-570, IEEE, 2021

[3] D. Panchard, F. Marelli, E. De Moura Presa, P. Wellig and M. Liebling, “Perspectives and limitations of visible-thermal image pair synthesis via generative adversarial networks,” in: Security + Defence, Target and Background Signatures VII, Proc. of SPIE, online only, pages 1186509-1–1186509-8, SPIE, 2021

[4] O. Mariani, F. Marelli, C. Jaques, A. Ernst and M. Liebling “Unequivocal cardiac phase sorting from alternating ramp- and pulse-illuminated microscopy image sequences,” in: 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), pages 868-872, 2021

3.9 Biometrics Security and Privacy

Head: Prof. Sébastien Marcel (PhD, University of Rennes, France, 2000; Visiting Professor, University of Cagliari, 2010; Lecturer, EPFL since 2013; Lecturer, UNIL since 2018; Lecturer, Master AI since 2019, IEEE Senior Member)

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

- Biometric recognition: We investigate the development of accurate and fair (unbiased) recognition algorithms, notably for face, voice, and vein biometric modalities.
- Presentation attack detection (PAD): We look for new and better ways of detecting presentation attacks (direct attacks) on face, voice, and vein biometric recognition systems.
- Morphing attack detection and Deepfakes detection: We investigate the threat of morphing attacks (indirect attacks) and more generally *Deepfakes* to develop effective morphing and Deepfake detection techniques.
- Biometric template protection: We research effective methods of preserving both the security of biometric recognition systems and the privacy of their users by protecting the biometric models ("templates") that are employed by the system for recognition purposes.

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

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

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

The BSP group in 2021 was composed of 1 head of group, 6 research associates, 2 postdocs, 3 PhD students, 2 interns.

Key scientific outputs: The BSP group has been pioneering the work on mobile biometrics (face and speaker recognition) and on PAD in face and speaker recognition by sharing the first open databases, organising the first International competitions and producing the first reproducible research studies in the domain. Regarding face PAD, the group confirmed that the current trend using discriminant classifiers is prone to over-fitting hence resulting in a lack of generalisation on unseen presentation attacks. These results question the efficiency and practicality of the existing PAD systems, as well as, call for creation of databases with larger variety of realistic presentation attacks. The BSP group also investigated approaches for heterogeneous face recognition, vein recognition and more recently Deepfakes to assess its threat to face recognition. More recently, the BSP group proposed a new scheme to evaluate the fairness of AI-driven face recognition systems.

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

Deepfakes: a New Threat to Face Recognition ?

It is becoming increasingly easy to automatically replace a face of one person in a video with the face of another person by using a pre-trained generative adversarial network (GAN). Recent public scandals call for automated ways to detect these Deepfake videos. To help developing such methods, we produced and analysed Deepfake videos. We showed that the state of the art face recognition systems based on Neural Networks are vulnerable to Deepfake videos, with about 90% false acceptance rates , which means methods for detecting Deepfake videos are necessary. By considering several baseline approaches, we found that audio-visual approach based on lip-sync inconsistency detection was not able to distinguish Deepfake videos. The best performing method, which is based on visual quality metrics and is often used in presentation attack detection domain, resulted in 8.97% equal error rate on high quality Deepfakes. Our experiments demonstrate that GAN-generated Deepfake videos are challenging for both face recognition systems and existing detection methods, and the further development of face swapping technology will make it even more so.

Figure 12: Illustration of Deepfakes

Presentation attack detection

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

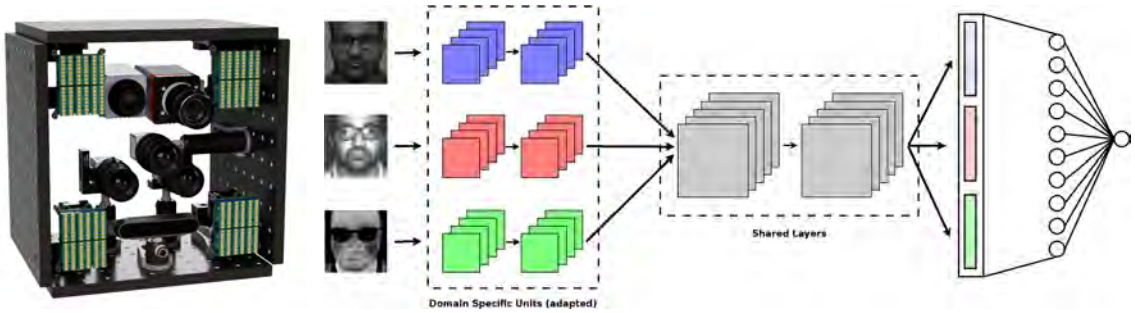


Figure 13: Illustration of multi-spectral face Presentation Attack Detection (PAD) device (left) and architecture (right).

We developed a multi-spectral face device (Figure 13) to capture synchronised, high speed and high resolution image sequences under different image domains (VIS, NIR, SWIR, Thermal and 3D). The main hypothesis is that bona fide samples are easier to discriminate from presentation attacks with the appropriate combination of image domains. We proposed a novel Deep Convolutional Neural Network architecture (Figure 13) to learn multi-spectral complementary information.

Heterogeneous face recognition

The task of Heterogeneous Face Recognition (Figure 14) consists in to match face images that were sensed in different modalities, such as sketches to photographs, thermal images to photographs or near infrared to photographs. We demonstrated that high level features of Deep Convolutional Neural Networks trained on visual spectra images are domain independent and can be used to encode faces sensed in different image domains.

Figure 14: Illustration of the heterogeneous face recognition problem: matching Visible spectra and Near-Infrared spectra (left), matching Visible spectra and sketch (middle), matching Visible spectra and Thermal spectra (right).

Remote photoplethysmography

Photoplethysmography (PPG) consists in measuring the variation in volume inside a tissue, using a light source. The aim of remote photoplethysmography (rPPG) is to measure the same variations, but using ambient light instead of structured light and widely available sensors such as a simple webcam (Figure 15).

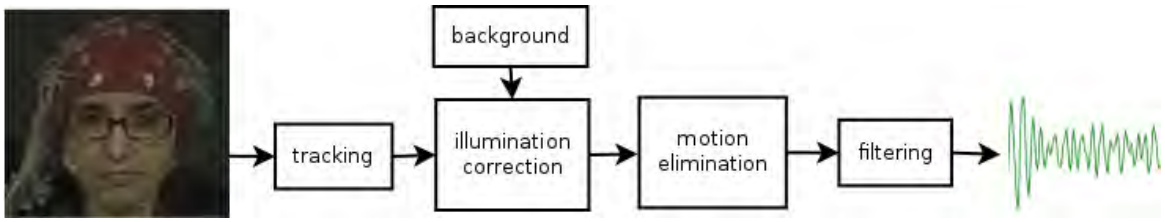


Figure 15: Illustration of remote photoplethysmography: colours from the video signal are filtered to produce an estimation of the heart beat signal.

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

Swiss Center for Biometrics Research and Testing

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

The centre attracted the attention of large companies (license, research and testing agreements) and led to many new projects (DARPA, IARPA, InnoSuisse). In 2020, the centre has developed over three directions:

- Maintaining and evolving the BEAT platform: The platform is now used in multiple research projects such as ALLIES, LEARN-REAL and the major H2020 AI4EU project. In parallel we are working towards improving the platform with the aim to create the Idiap AI platform.
- Evaluation and Testing: we engaged with the FIDO Alliance (https://fidoalliance.org), focused on providing open and free authentication standards to help reduce the world's reliance on passwords, and became an accredited FIDO lab able to perform certification of biometrics products. Since 2020 we are also an Android accredited lab and we conducted several evaluations of biometric products.
- Joining the CITeR cooperative research center: we joined the US CITeR (Center for Identification Technology Research) National Science Foundation (NSF) Industry/University Cooperative Research Center (I/UCRC) as the first non-US academic site with two affiliates (IDEMIA and SCIPA) to fund our research activities.

Key publications

[1] A. George and S. Marcel, "Cross Modal Focal Loss for RGBD Face Anti-Spoofing", *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.

[2] T.d.F. Pereira and S. Marcel, "Fairness in Biometrics: a figure of merit to assess biometric verification systems", *IEEE Transactions on Biometrics, Behavior, and Identity Science (TBIOM)*, 2021.

[3] V. Krivokuca and S.Marcel, "Biometric Template Protection for Neural-Network-based Face Recognition Systems: A Survey of Methods and Evaluation Techniques", *arXiv*, 2021.

[4] L. Colbois, T.d.F. Pereira and S. Marcel, "On the use of automatically generated synthetic image datasets for benchmarking face recognition", *International Joint Conference on Biometrics (IJCB)*, 2021.

[5] H. Otroshi and S. Marcel, "Towards Protecting and Enhancing Vascular Biometric Recognition methods via Biohashing and Deep Neural Networks". *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2021

[6] P. Korshunov and S. Marcel, "Subjective and objective evaluation of deepfake videos", *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2021.

[7] G. Heusch and S. Marcel, "Remote Blood Pulse Analysis for Face Presentation Attack Detection", *Handbook of Biometric Anti-Spoofing (2nd Edition)*, Springer, Chapter 10, pp 267–289, 2019.

[8] T.d.F. Pereira, A. Anjos and S. Marcel, "Heterogeneous Face Recognition Using Domain Specific Units", *IEEE Transactions on Information Forensics and Security*, Vol 14(7), pp 1803–1816, 2019.

3.10 Biosignal Processing

Head: Dr. André Anjos (BSc, MSc & PhD, Federal University of Rio de Janeiro, Brazil, 1999, 2001, 2006; Visiting Professor, State University of Bauru, 2015; Lecturer, EPFL, 2013–; Lecturer, Master of AI, 2019–)

Group overview: The Biosignal Processing group sits in the intersection between AI and Healthcare. It studies the use of Machine Learning for computer-aided diagnosis, prognosis, decision support, and risk analysis from a variety of bio-signals, including, but not limited to, imaging, clinical, demographical, and temporal data. The group has a special focus in reproducibility and the responsible use of AI, with an interest on fairness, explainability, and safety. It leverages on Idiap's expertise on human subject handling, data acquisition, open science and data processing.

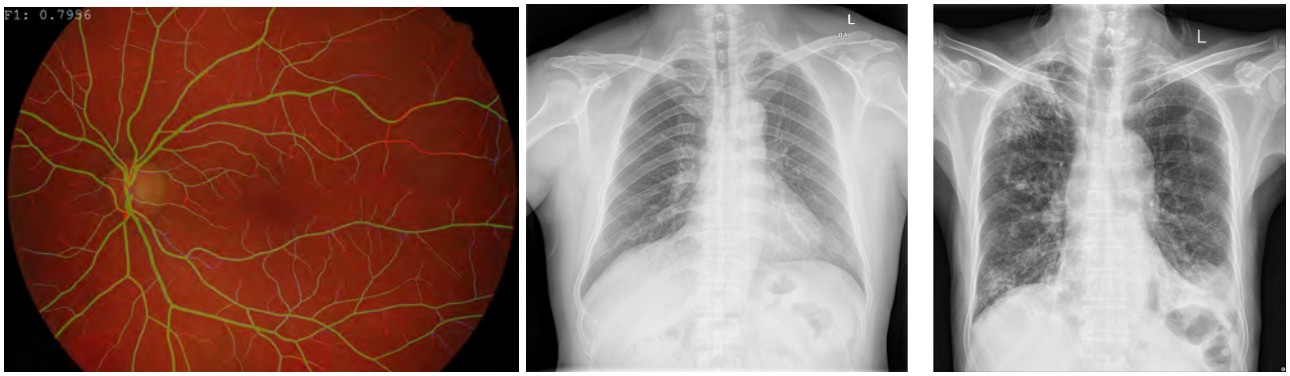
In 2021, the Biosignal Processing Group was composed of the head of the group, and 3 master students. A postdoctoral researcher will join the team in 2022.

Key scientific outputs: The group currently develops 4 thematic areas of research at the intersection between Artificial Intelligence and Medicine: computer-aided diagnosis (CAD) and decision support from imaging data (e.g. ophthalmic, radiography, histopathology), computer-aided risk prediction, the analysis of time sequences (e.g. vital signs or electroencephalo/cardiography), and, finally, reproducibility in AI research. We are particularly interested in the responsible use (explainability, fairness) of AI models we build, as well as challenges associated to data scarcity in the medical domain. The milestones for 2022 include continued work in above cited areas, and the development of new projects on computed tomography (CT) scans for the detection of rare diseases, reproducibility, and, finally, continued participation in EPFL courses and the Idiap's AI master program. Further funding was requested to the Swiss National Science Foundation to expand the team.

Additional information and a list of projects are available from www.idiap.ch/bio-sig.

Semantic Segmentation for Medical Imaging
Since the introduction of U-Nets in 2015, the field of medical image segmentation has seen renewed interest bringing in a variety of fully convolutional (deep) neural network (FCN) architectures for binary and multi-class segmentation problems promising very attractive results, with applications in computed tomography, retinography, and histopathology to cite a few. Despite the incredible progress, the lack of annotated images (due to cost), and rigor in the comparison of trained models has led the community to believe larger and more dense networks provide better results. We addressed these gaps in different ways. The first was to conduct and publish rigorous (open source, reproducible) benchmarks with popular retinography datasets and state-of-the-art FCN models in which we throw a new light over some published figures, opening space for new developments. In addition, we proposed a simple extension, of a compact U-Net architecture, dubbed W-Net, by concatenating two U-Nets together, which reaches outstanding performance on several popular datasets, still using orders of magnitude less learnable weights than any previously published approach.

Computer-aided Diagnosis for Tuberculosis
Tuberculosis (TB) is one of the leading causes of death from a single infectious agent. In many high-burden regions around the world, which often lack specialized healthcare professionals, Chest X-Ray (CXR) exams continue to be of vital importance in the diagnosis and follow-up of the various presentations of the disease. In this context, we investigate the benefits of automatic Pulmonary Tuberculosis (PTB) detection methods based on radiological signs found on CXR. Contrary to direct scoring from images, implemented in most related work, indirect detection offers natural interpretability of automated reasoning. We identify generalization difficulties for direct detection models trained exclusively on the modest amount of publicly available CXR images from PTB patients. We subsequently show that a model, pre-trained on tens of thousands of CXR images using automatically annotated radiological signs, offers a more adequate base for development. By relaying radiological signs through a simple linear classifier, one is able to obtain



(a) Vessel segmentation for Retinography – predicted vessel maps vs. ground truths for a DL model evaluated on the high-resolution HRF test-set. True positives, false positives and false negatives are displayed in green, blue and red respectively.

(b) Radiological signs on healthy (left) and PTB-affected lungs (right).

state-of-the-art results on all three publicly available datasets. We further discuss limitations imposed by the limited number of PTB-specific radiological signs available on public datasets, and evaluate possible performance gains that could be obtained if more were available. This work is fully reproducible.

Reproducible Research

Since the last decade, we have been actively looking at the reproducibility of published work and how to lower the entrance barrier of publication readers. We argue it is insufficient, in most cases, to only publish software leading to results if original data remains inaccessible. Reproducibility should imply in the following characteristics: repeatability, share-ability, extensibility and stability, which is not guaranteed by most published material to date. We propose a software suite called Bob[36] that possesses such characteristics, demonstrating its flexibility to various tasks. From another perspective, there are legitimate cases in which raw data leading to research conclusions cannot be published. Furthermore, in a growing number of use-cases, the availability of both software does not translate to an accessible reproducibility scenario. To bridge this gap, we built an open platform for research[37] in computational sciences related to pattern recognition and machine learning, to help on the development, reproducibility and certification of results obtained in the field.

Key publications

[1] A. Galdran, A. Anjos, J. Dolz, H. Chakor, H. Lombaert, and I. Ben Ayed, “State-of-the-Art Retinal Vessel Segmentation with Minimalistic Models”, Submitted to Nature Scientific Reports, 2022.

[2] M. A. Renzo, N. Fernandez, A. A. Baceti, N. N. Moura Junior, and A. Anjos. “Development of a lung segmentation algorithm for analog imaged chest x-ray: preliminary results”, XV Brazilian Congress on Computational Intelligence. 2021.

[3] A. C. B. H. Ferreira, D. D. Ferreira, H. C. Oliveira, I. C. de Resende, A. Anjos, and M. H. B. de M. Lopes, “Competitive neural layer-based method to identify people with high risk for diabetic foot”, Computers in Biology and Medicine, vol. 120, p. 103744, May 2020.

[4] A. Anjos, M. Günther, T. de Freitas Pereira, P. Korshunov, A. Mohammadi, and S. Marcel, “Continuously Reproducing Toolchains in Pattern Recognition and Machine Learning Experiments”, at Thirty-fourth International Conference on Machine Learning, 2017.

[36] https://www.idiap.ch/software/bob
[37] https://www.idiap.ch/software/beat

3.11 Energy Informatics

Head: Dr. Jérôme Kämpf (BSc, University of Kent at Canterbury; MSc, University of Lausanne, 2001 and 2003; PhD, Ecole Polytechnique Fédérale de Lausanne, 2009)

Group overview: The Energy Informatics concepts are to exploit state-of-the-art Information and Communication Technologies to tackle global warming and climate change challenges. Our aim is to increase the integration of renewable and distributed energy sources by making energy systems smarter, and to increase energy efficiency beyond what improvements at component level can achieve. In that vein, the Energy Informatics Group at Idiap researches into ways of simulating energy transition pathways with intelligent control and adjustment mechanisms of evolving shelters with retrofitting and use, renewable energy production and energy storage in a changing climate. Two key application areas (see §2.2) are more deeply studied: the building automation with its adaptation to human behavior and the energy management with its indirect modeling of human activity to anticipate energy needs.

In 2021, the Energy Informatics Group was composed of 1 head of group, 3 postdocs, 2 exchange PhD students, 1 exchange MSc student and 1 intern.

Key scientific outputs: In building automation, surrogate models were developed to evaluate glare in the eye of the occupants and work plane illuminance. The proposed machine learning models being light and fast, allowed for their implementation in embarked computers to control in quasi real-time blinds and electric lighting in two offices of our Idiap building.
In energy management, the open-source Urban Energy Simulation tool CitySim was improved to simulate the hydraulics and thermal behaviour in the pipes of District Heating Networks. Comparisons with monitoring data taken on partner networks showed a good agreement. Physics Informed Neural Networks were explored as a track to build AI surrogate models.

Additional information and a list of projects are available from http://www.idiap.ch/energy.

Building automation
Building occupants tend to negatively perceive building automation as it may alter their comfort for the sake of energy savings. The main challenge identified is therefore to control (or suggest actions on) the building infrastructure in order to minimise the energy intake while maintaining a sufficient comfort for the occupants. Noteworthy, human comfort has multiple facets and depends on each individual's history and preferences. The use of smart sensors is a unique opportunity to learn and adapt the automation to the users.

Through the project LUCIDELES, we study the trade-off between energy and visual comfort in buildings. Machine learning based surrogate models were developed to simulate the daylight penetration through Venetian blinds and its effect on work-plane illuminance and glare perceived by the occupants. These predictive models were then used to control optimally the blinds' state to avoid glare while maximising daylight in buildings, thus minimising the need for electric lighting. Over the summer 2021, with the concourse of Idiap's infrastructure team, we have upgraded two offices at Idiap with dimmable LED luminaires and automated (KNX) blinds control systems. Both offices labelled 103 and 107 were further equipped with Task Lights, controlled for their intensity and colour temperature. The predictive models and control architecture were adapted and improved with solar protection and thermal saving measures for the two rooms at Idiap. Questionnaires to evaluate the satisfaction of the users were prepared and a longitudinal experiment with office workers started in October 2021. The experimental set-up was demonstrated to the public during Idiap's open doors in September 2021.

The project LUCIDELES is giving interesting insights within the Internation Energy Agency (IEA) Solar Heating and Cooling (SHC) Task 61 (task61.iea-shc.org) on Integrated Solutions for Daylighting and Electric Lighting: From component to user centered system efficiency.

Energy management

While urban areas cover about 2% of the planet surface, they consume 3/4 of the total planetary resources. According to the Pareto rule, major energy-related efforts should target cities and their inhabitants. The efforts comprises the integration of renewable and distributed energy sources, which in dense urban centers remains a challenge. An efficient operation of existing infrastructures and planning of new ones is needed to mitigate and adapt to climate change.

The management of the energy transition depends on the availability and the quality of a large range of data. However, data is often difficult to find, mixed in different repositories, as well as fragmented. Through the project EnerMaps (enermaps.eu), we aim to support scientific data management in the field of energy by the development of 1) a user-friendly Gateway centralising all research articles and publications on the same platform (beta.enermaps.openaire.eu) and 2) a data visualisation platform with calculation modules (lab.idiap.ch/enermaps). We developed the calculation module named HeatLearn based on a Convolutional Neural Network which determines the density of heating needs as a function of the European Settlement Map. HeatLearn, based on open-data, goes in the direction of an efficient planning of district heating network extensions.

Meanwhile in the project Eguzki, we aim to develop a simulation tool for District Heating Networks (DHN) based on Artificial Intelligence (AI) to quickly assess, predict and optimise the performance of looped networks. So far, we have improved the tool CitySim (github.com/idiap/CitySim-Solver) to simulate in a physically-based way the hydraulics and thermal behaviour of fluids in the pipes and serve as a baseline for our AI surrogate models. Regarding the latter, we have focused our developments on Physics Informed Neural Networks (PINNs) giving promising preliminary results on a first case-study.

CREM - Centre de Recherches Energétiques et Municipales

We maintain a close scientific collaboration with CREM (www.crem.ch), an actor specialized in the field of energy sustainability in urban areas. CREM has a large network of communes, public utilities and companies making it an essential partner for case-studies and dissemination. We share the project EnerMaps,

Key publications

[1] C. Basurto, R. Boghetti, M. Colombo, M. Pappinutto, J. Nembrini and J. Kämpf, Implementation of machine learning techniques for the quasi real-time blind and electric lighting optimization in a controlled experimental facility, in: Journal of Physics: Conference Series, IOP Publishing, 2021

[2] G. Peronato, R. Boghetti and J. Kämpf, A machine-learning model for the prediction of aggregated building heating demand from pan-European land-use maps, in: Journal of Physics: Conference Series, 2021

[3] P. Florio, G. Peronato, A. T. D. Perera, A. Di Blasi, Kin Ho Poon and J. Kämpf, Designing and assessing solar energy neighborhoods from visual impact, in: Sustainable Cities and Society, 2021

[4] V. Todeschi, R. Boghetti, J. Kämpf and G. Mutani, Evaluation of Urban Scale Building Energy-Use Models and Tools – Application for the City of Fribourg, Switzerland, in: Sustainability, 13(7), 2021

[5] F. Haneef, G. Pernigotto, A. Gasparella and J. Kämpf, Application of Urban Scale Energy Modelling and Multi-Objective Optimization Techniques for Building Energy Renovation at District Scale, in: Sustainability, 13(20), 2021

3.12 Computation, Cognition, and Language

Head: Prof. Lonneke van der Plas (MPhil Cambidge, UK, 2001; PhD, University of Groningen, The Netherlands, 2008; Junior professor, University of Stuttgart, Germany, since 2012, Associate professor, University of Malta, since 2014)

Group overview: The Computation, Cognition and Language (CCL) group was newly created in 2021. It is composed of the head, Lonneke van der Plas, and one intern since October 2021. The group aims to build models of language that are inspired by human cognition and linguistic theory thereby working on the frontiers of what current models are capable of.

The group currently focuses on two areas. Firstly, **cross-lingual transfer**: We investigate to what extend language technology tools for one language can be transferred to another language in low-resource scenarios, for example, when transferring a model that automatically determines the sentiment (positive, negative, or neutral) of a sentence in English to a language such as Maltese, for which limited amounts of data are available. Secondly, **modelling creative processes**: We research models that are able to reproduce aspects of human creative processes with the aim of creating technology that supports human creativity.

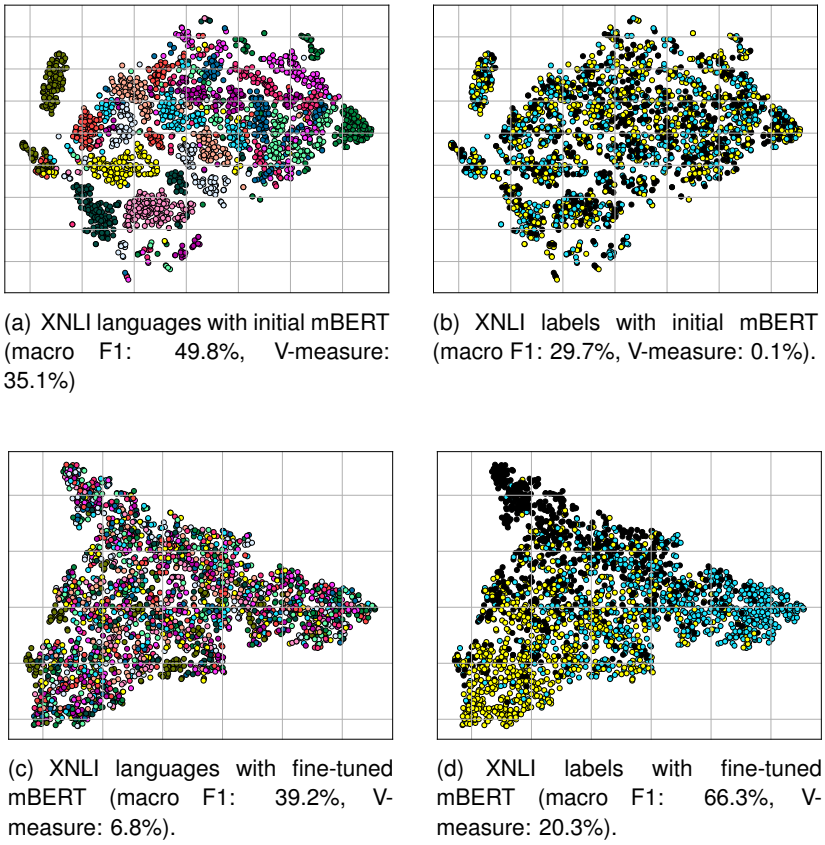
Key scientific outputs: Shortly after the fresh start of the group, we reached several scientific achievements. We published a paper that contributes to the recent endeavours to try and to better understand large-scale language models such as BERT. The head-of-group was anonymously accepted as associate investigator of the NCCR Evolving Language with a project on modelling lexical change. We have also invested ourselves to position the group both locally and internationally, and took part in several events such as the International Create Challenge, which we won , and resulted in two collaborations with major industry partners. The head-of-group was invited to give two keynote talks at the AICon conference in Schwyz, and she gave an invited talk at Cardiff University.

Researching the language-specificity of multilingual BERT and the impact of fine-tuning

Recent work has shown evidence that the knowledge acquired by multilingual BERT (mBERT) has two components: a language-specific and a language-neutral one. We analysed the relationship between them, in the context of fine-tuning on the task of natural language inference, for example. Visualisations reveal that mBERT loses the ability to cluster representations by language after fine-tuning, a result that is supported by evidence from language identification experiments. However, further experiments on 'unlearning' language-specific representations using gradient reversal and iterative adversarial learning are shown not to add further improvement to the language-independent component over and above the effect of fine-tuning. The results presented suggest that the process of fine-tuning causes a reorganisation of the model's limited representational capacity, enhancing language-independent representations at the expense of language-specific ones. Figure 16 shows the t-SNE projections of the text-based XNLI representations, together with the corresponding V-measure. For labels in both target tasks, mBERT starts off with no discernible structure, whereas fine-tuning results in clear clusters by label (compare figures (b) and (d)). At the same time, fine-tuning leads to languages being less clearly identifiable (compare figures (a) and (c)).

Modelling creative processes

In earlier work, we introduced temporally and contextually-aware models for the novel task of predicting unseen but plausible concepts, as conveyed by noun-noun compounds in a time-stamped corpus. We trained compositional models on observed compounds, more specifically the composed distributed representations of their constituents across a time-stamped corpus. The model captures generalisations over this data and learns what combinations give rise to plausible compounds and which ones do not.



(a) XNLI languages with initial mBERT (macro F1: 49.8%, V-measure: 35.1%)

(b) XNLI labels with initial mBERT (macro F1: 29.7%, V-measure: 0.1%).

(c) XNLI languages with fine-tuned mBERT (macro F1: 39.2%, V-measure: 6.8%).

(d) XNLI labels with fine-tuned mBERT (macro F1: 66.3%, V-measure: 20.3%).

Figure 16: 2D t-SNE projections before and after fine-tuning, with XNLI as target task. Macro F1 scores for label/language classification of mBERT embeddings are also given.

Since then, we have also considered the ethical implications of systems that lack creativity. In a paper written in collaboration with a philosopher from the University of Zurich, Michele Loi, which received the ELSI best paper award, we aimed to put an issue on the agenda of AI ethics that in our view is overlooked in the current discourse that is dominated by topics such as trustworthiness and bias. In a second paper (Vigano et al , 2020), we provide a philosophical account of the value of creative systems for individuals and society. After, these philosophical considerations, we have looked at the topic of computational creativity from a more practical angle, during a recently held International Create Challenge, where we won the competition.

Key publications

[1] Marc Tanti, Lonneke van der Plas, Claudia Borg and Albert Gatt, On the Language-specificity of Multilingual BERT and the Impact of Fine-tuning, in: Proceedings of the Workshop on Analyzing and Interpreting Neural Networks for NLP, 2021

[2] Carlos Mena, Andrea DeMarco, Claudia Borg, Lonneke van der Plas, Albert Gatt, Data Augmentation for Speech Recognition in Maltese: A Low-Resource Perspective, in: ArXiv preprint, 2021

[3] Patrick Ziering and Lonneke van der Plas, Compound or phrase or in between? Testing Linguistic Criteria for Compoundhood, in: Compounds Between Phrases and Words, Special issue Word Structure, 2020

[4] Gianina Iordăchioaia, Lonneke van der Plas, Glorianna Jagfeld Compositionality in English deverbal compounds: The role of the head, in: Sabine Schulte im Walde and Eva Smolka (eds.) The role of constituents in multiword expressions: An interdisciplinary, cross-lingual perspective. Phraseology and Multiword Expressions. Language Science Press, 2020

[5] Michele Loi, Eleonora Viganò, and Lonneke van der Plas The societal and ethical relevance of computational Creativity, in: Proceedings of the International Conference on Computational Creativity, 2020

3.13 Reasoning and Explainable AI

Head: Dr. André Freitas (PhD, NUI Galway, Ireland, 2016; Senior Lecturer (Associate Professor), University of Manchester, UK; AI Group Leader, CRUK Manchester Institute, UK)

Group overview:

The Reasoning & Explainable AI (ExplAIn) group aims at developing AI models which are capable of complex, abstract and flexible inference. The group operates at the interface between neural and symbolic AI methods aiming to enable the next generation of explainable, data-efficient and safe AI models. These themes closely dialogue with the need for mitigating some of the risks for the adoption of AI models in real world scenarios.

Key scientific outputs: The group is pioneering the development of natural language inference (NLI) models targeting abstract and scientific domains, including areas such as Cancer Research, Mathematics and Physics. Recent contributions include: improving the formal understanding of inference and explanatory patterns in scientific discourse, the development of novel neuro-symbolic architectures to support better inference control, the development of probing and metamorphic testing frameworks for natural language inference; assessment of the interaction between end-users and AI models, aiming for critical transparency and clinical embedding; development of safe and explainable AI models to support cancer research and industrial applications. Currently, the group focuses on 6 core research streams, which are outlined below.

Building AI models capable of complex abstract inference. *How to develop AI models which can deliver complex, expert-level (e.g. scientific) reasoning and explanations?* In this area, the ExplAIn group focuses on building models capable of encoding complex and abstract inference, with a particular emphasis in the scientific domain. Figure 17 outlines an example of a complex expert-level explanation/argumentation in the context of oncology. Our contributions to this area affirm a balance between the flexibility provided by contemporary deep learning based models and the explicit inference controls delivered by symbolic methods. In linguistic terms, scientific explanations are still poorly defined objects. Our research this year contributed to define some of the critical components behind scientific inference and explanations, and the recurring linguistic and inference patterns expressed on them. Eliciting these components can inform the construction of models which are capable of abstract inference, better generalisation and which are natively explainable.

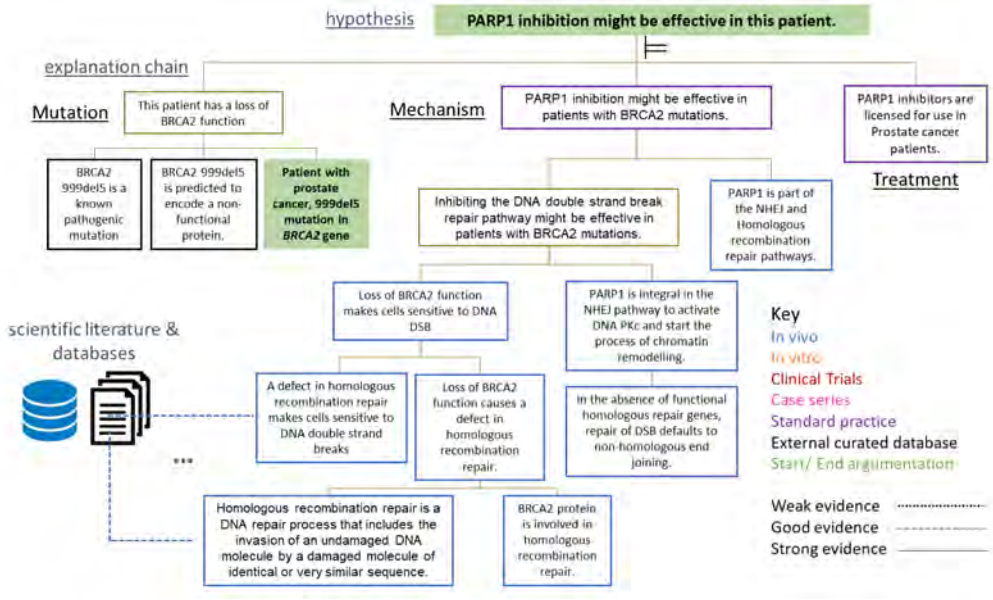


Figure 17: Example of expert-level explanation in the context of oncology.

Reasoning over natural language & equations. *How to develop AI models which can reason over mathematical text?* A significant part of scientific discourse is expressed as mathematics. The ExplAIn group is currently pioneering new AI methodologies which allows automated inference over mathematical text, where models need to encode both natural language and mathematical expressions. This year, the group established new modeling mechanisms to jointly encode these two symbolic modalities, including modelling the interpretation of variables within mathematical discourse. Our investigation extended to the domain of Physics, where we developed the first corpus for Natural Language Inference over Physics texts.

Safe & controlled inference. *How to develop inference models which balance flexibility, rigour and safety?* We investigated new methods to integrate additional logical constraints into neural-based natural language inference (NLI) models. Our contributions to this area include the analysis and improvement of abstract logical feature modelling within neural NLI models. In order to establish the internal linguistic and inference consistency properties of neural NLI models, we introduced the notion of metamorphic testing in Natural Language Processing (NLP), and systematically applied semantic and inference probing within different domains. An open source framework (Probe-Ably) was developed by the group to support this process.

Transparency & explainability. *How to develop efficient communication mechanisms between end-users and AI models?* The application of machine learning (ML) models in domains such as healthcare requires the introduction of mechanisms which allow end-users to have a critical understanding of the models' underlying inference process. The design of explainabilty mechanisms allows for the communication of these model properties to non-AI experts. During 2021, we pragmatically embedded different standard models of explainability in a clinical decision setting, and found that, despite delivering transparency, these models still lack the usability required by healthcare professionals. The underlying complexity of AI models brings transparency problems even for AI experts. One of the research areas investigated by the group is on how to improve the efficiency of the communication between experts using optimised diagrammatic representations of neural network based models.

Clinical applications & Industrial Collaborations. The group collaborates closely with a major cancer research centre in the UK (CRUK Manchester Institute), on the development of AI models in the oncology domain. Areas of collaboration involve the development of AI models to support treatment recommendation, toxicity prediction and the development of novel biomarkers, with an emphasis on safety, explainability and ethics (SEE). During the COVID outbreak, the group supported the development of CORONET, an explainable ML prototype to support the decision on the clinical admission of cancer patients with COVID-19 (in partnership with collaborators at the University of Manchester). Current industrial collaborations involve the use of natural language inference to support the discovery of new antibiotic substances (together with Inflamalps) and a cross research group collaboration on Safety and Explainable AI for decision support in orthopaedics (in collaboration with the Biosignal Processing Group and Med4CAST).

Key publications

[1] Marco Valentino, Mokanarangan Thayaparan, Deborah Ferreira, André Freitas. "Hybrid Autoregressive Inference for Scalable Multi-hop Explanation Regeneration". In *Proc. of AAAI*, 2022.

[2] Marco Valentino, Mokanarangan Thayaparan, André Freitas. "Unification-based Reconstruction of Multi-hop Explanations for Science Questions". In *Proc. of EACL*, 2021.

[3] Deborah Ferreira, Mokanarangan Thayaparan, Marco Valentino, Julia Rozanova, André Freitas "STAR: Cross-modal Statement Representation for Selecting Relevant Mathematical Premises". In *Proc. of EACL*, 2021.

[4] Mokanarangan Thayaparan, Marco Valentino, André Freitas. "Explainable Inference Over Grounding-Abstract Chains for Science Questions". In *Findings of the ACL*, 2021.

[5] Giangiacomo Mercatali, André Freitas. "Disentangling Generative Factors in Natural Language with Discrete Variational Autoencoders". In *Findings of EMNLP*, 2021.

3.14 Signal Processing for Communication

Head: Dr. Ina Kodrasi (MSc, Jacobs University Bremen, Germany, 2010; PhD, Carl von Ossietzky Universität Oldenburg, Germany, 2015)

Group overview:

The Signal Processing for Communication group focuses on tackling challenges arising in acoustic communication such as speaking impairments, hearing impairments, and undesired interferences (e.g., additive noise and reverberation). The high-level objectives of the group are to establish novel digital signal processing and pattern recognition approaches for speech, audio, and multi-modal signals to automatically detect speaking and hearing impairments, provide speaking and hearing assistance, and improve the communication experience in the presence of undesired interferences.

The group was established in 2021 and Dr. Kodrasi is currently building the team to address the topics outlined above. In 2021, the Signal Processing for Communication group consisted of one PhD student jointly affiliated with the Speech and Audio Processing Group.

Key scientific outputs: Key scientific outputs of the group in 2021 include i) novel automatic pathological speech detection approaches and ii) multi-task single-channel speech enhancement.

Additional information and a list of projects are available at www.idiap.ch/spc.

Pathological speech detection

As part of the SNSF Sinergia projects MoSpeeDi and ChaSpeePro on motor speech disorders, we have developed novel deep learning automatic pathological speech detection approaches.

Recently proposed deep learning approaches for pathological speech detection use unsupervised auto-encoders to obtain a high-level representation of speech. Since there is no guarantee that these representations are robust to pathology-unrelated cues (e.g., speaker identity information) or that they are discriminative for pathology detection, we have proposed to use supervised auto-encoders instead (cf. Figure 18) [1]. To reduce the influence of speaker variability unrelated to pathology, we have constructed speaker identity-invariant representations by adversarial training of an auto-encoder and a speaker identification task. To obtain a discriminative representation, we have jointly trained an auto-encoder and a pathological speech classifier.

Furthermore, state-of-the-art approaches typically learn discriminative representations by processing time-frequency input representations such as the magnitude spectrum of the short-time Fourier transform. Although these approaches are expected to leverage perceptual pathological cues, such representations do not necessarily convey perceptual aspects of complex sounds. Inspired by the temporal processing mechanisms of the human auditory system, we have proposed to factor speech signals into the product of a slowly varying envelope and a rapidly varying fine structure similarly to the decomposition occurring in the cochlea [2]. Two discriminative representations separately learned from the temporal envelope and fine structure signals using convolutional neural networks (CNNs) are then exploited for automatic pathological speech detection (cf. Figure 19).

A remaining challenge in successfully exploiting deep learning approaches in pathological speech assessment is alleviating over-fitting issues associated with the typically limited training data that is available. To tackle this challenge, we have proposed to use a CNN-based pathological speech detection framework exploiting pairwise distance matrices [3]. The distance matrix between a reference (i.e., healthy) representation and a phonetically-balanced test representation is computed and used as input to the CNN-based binary classifier as schematically illustrated in Figure 19. Such a framework results in a five-fold increase of available training samples in typically used databases.

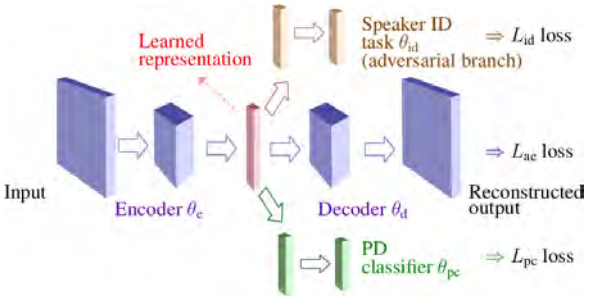


Figure 18: Supervised autoencoder for automatic pathological speech detection.

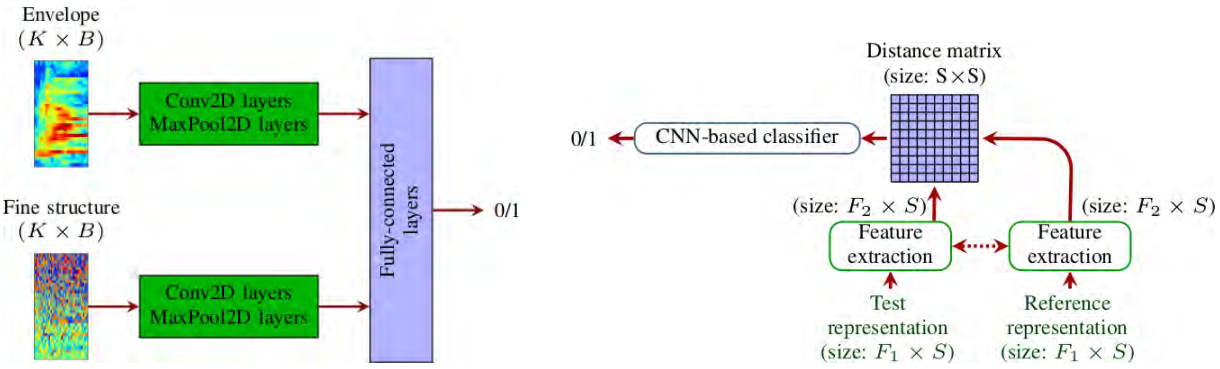


Figure 19: Auditory-inspired CNN (left) and pairwise distance-based CNN (right) for automatic pathological speech detection.

Single-channel speech enhancement.

To cope with reverberation and noise in single-channel acoustic scenarios, typical supervised deep learning approaches learn a mapping from reverberant and noisy input features to a user-defined target such as the clean signal magnitude or a time-frequency mask. We have proposed to incorporate multi-task learning in such approaches by using speech presence probability estimation as a secondary task assisting the target estimation in the main task [5]. The advantage of multi-task learning lies in sharing domain-specific information between the two tasks (i.e., target and speech presence probability estimation), yielding more generalizable and robust representations for single-channel speech enhancement.

Key publications

[1] P. Janbakhshi and I. Kodrasi, "Supervised speech representation learning for Parkinson's Disease classification," in *Proc. ITG Conference on Speech Communication*, Kiel, Germany, Sep. 2021, pp. 1–5.

[2] I. Kodrasi, "Temporal envelope and fine structure cues for dysarthric speech detection using convolutional neural networks," in *IEEE Signal Processing Letters*, vol. 28, pp. 1853–1857, 2021.

[3] P. Janbakhshi, I. Kodrasi, and H. Bourlard, "Automatic dysarthric speech detection exploiting pairwise distance-based convolutional neural networks," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, Toronto, Canada, June 2021, pp. 7328–7332.

[4] I. Kodrasi, M. Pernon, M. Laganaro, and H. Bourlard, "Automatic and perceptual discrimination between dysarthria, apraxia of speech, and neurotypical speech," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, Toronto, Canada, June 2021, pp. 7308–7312.

[5] L. Wang, J. Zhu, and I. Kodrasi, "Multi-task single channel speech enhancement using speech presence probability as a secondary task training target," in *Proc. European Signal Processing Conference*, Dublin, Ireland, Aug. 2021, pp. 296–300.

4 Researchers (in addition to permanent staff)

4.1 Research Associates

Last name	First name	Nationality	Supervisor		Start	Estimated End
BHATTACHARJEE	Sushil	CH	Sébastien	MARCEL	2016-01-01	
BRAUN	Rudolf	DE	Hervé	BOURLARD	2021-03-01	2026-02-28
GEISSBUHLER	David	CH	Sébastien	MARCEL	2018-06-01	
GEORGE	Anjith	IN	Sébastien	MARCEL	2021-01-01	
IMSENG	David	CH	Hervé	BOURLARD	2016-04-01	2021-02-28
KORSHUNOV	Pavel	EE	Sébastien	MARCEL	2017-01-01	
KOTWAL	Ketan	IN	Sébastien	MARCEL	2021-08-01	2022-07-31
KRIVOKUCA	Vedrana	HR	Sébastien	MARCEL	2021-01-16	
LABHART	Florian	CH	Daniel	GATICA-PEREZ	2017-11-01	2021-06-30
MADIKERI	Srikanth	IN	Petr	MOTLICEK	2018-03-01	
VILLAMIZAR	Michael	ES	Jean-Marc	ODOBEZ	2020-11-01	2022-10-31
VILLATORO TELLO	Esaú	MX	Hervé	BOURLARD	2021-11-01	

4.2 Post-doctoral Scholars

Last name	First name	Nationality	Supervisor		Start	Estimated End
ANTONELLO	Niccolò	IT	Hervé	BOURLARD	2020-04-16	2021-04-15
BASURTO DAVILA	Chantal	MX	Jérôme	KÄMPF	2020-04-01	2022-01-31
DE FREITAS PEREIRA	Tiago	BR	Sébastien	MARCEL	2019-03-01	2022-02-28
HE	Weipeng	CN	Petr	MOTLICEK	2020-11-01	2021-08-31
KHOSRAVANI	Abbas	IR	Philip	GARNER	2019-08-01	2021-10-31
KOTWAL	Ketan	IN	Sébastien	MARCEL	2018-01-01	2021-07-31
KRIVOKUCA	Vedrana	HR	Sébastien	MARCEL	2017-01-15	2021-01-15
LAZZARI	Gianrocco	IT	Jérôme	KÄMPF	2021-05-01	2021-08-03
MASSÉ	Benoit	FR	Daniel	GATICA-PEREZ	2019-03-01	2021-10-31
MICULICICH	Lesly	PE	James	HENDERSON	2020-11-01	2021-01-31
MOHAMMADI	Amir	IR	Sébastien	MARCEL	2020-02-01	2023-01-31
PARIDA	Shantipriya	IN	Petr	MOTLICEK	2019-02-01	2022-01-31
PERONATO	Giuseppe	IT	Jérôme	KÄMPF	2020-09-04	2023-11-30
PHAN	Trung	VN	Daniel	GATICA-PEREZ	2020-10-01	2021-06-30
PRASAD	Ravi	IN	Mathew	MAGIMAI DOSS	2018-10-01	2021-12-31
RACCA	Mattia	IT	Jean-Marc	ODOBEZ	2020-10-01	2022-03-31
SARFJOO	Saeed	IR	Petr	MOTLICEK	2018-01-01	2022-06-30
SIEGFRIED	Rémy	CH	Jean-Marc	ODOBEZ	2021-10-01	2022-03-31
SILVERIO	João	PT	Sylvain	CALINON	2019-07-01	2022-02-28
SOUSA EWERTON	Marco	DE	Jean-Marc	ODOBEZ	2019-08-15	2022-04-30
TORNAY	Sandrine	CH	Mathew	MAGIMAI DOSS	2021-03-01	2022-08-31
VLASENKO	Bogdan	DE	Mathew	MAGIMAI DOSS	2017-07-01	2022-02-15
ZUFFEREY	Marie	CH	André	FREITAS	2021-09-15	2022-09-15

4.3 PhD Students

Last name	First name	Nationality	Supervisor		Start	Estimated End
BEHJATI	Melika	IR	James	HENDERSON	2020-09-01	2023-08-31
BITTAR	Alexandre	CH	Philip	GARNER	2020-03-02	2024-03-02
BOGHETTI	Roberto	IT	Jérôme	KÄMPF	2021-06-01	2025-05-31
CHEN	Haolin	CN	Philip	GARNER	2021-08-30	2025-08-01
COLBOIS	Laurent	CH	Sébastien	MARCEL	2021-01-01	2024-12-31
COMAN	Andrei	IT	James	HENDERSON	2020-10-01	2024-09-30
COPPIETERS DE GIBSON	Louise	BE	Philip	GARNER	2020-10-01	2024-09-30
COURDIER	Evann	FR	François	FLEURET	2019-03-01	2023-02-28
DUBAGUNTA	Pavankumar	IN	Mathew	MAGIMAI DOSS	2017-04-01	2021-08-31
FABIEN	Maël	FR	Petr	MOTLICEK	2020-03-01	2024-02-29
FAJCÍK	Martin	SK	Petr	MOTLICEK	2021-11-01	2024-10-31
FEHR	Fabio	CH	James	HENDERSON	2021-02-01	2025-01-31
FRITSCH	Julian	DE	Mathew	MAGIMAI DOSS	2018-07-01	2022-06-30
GIRGIN	Hakan	TR	Sylvain	CALINON	2018-09-01	2022-08-31
GUPTA	Anshul	IN	Jean-Marc	ODOBEZ	2021-01-01	2024-12-31
HERMANN	Enno	DE	Mathew	MAGIMAI DOSS	2018-07-01	2022-06-30
JANBAKHSHI	Parvaneh	IR	Hervé	BOURLARD	2018-02-01	2022-04-30
JANKOWSKI	Julius	DE	Sylvain	CALINON	2020-04-01	2024-03-31
JOHARI	Seyed Mohammad Mahdi	IR	François	FLEURET	2020-01-01	2023-12-31
KABIL	Selen	TR	Hervé	BOURLARD	2018-05-01	2022-04-30
KARIMI MAHABADI	Rabeeh	IR	James	HENDERSON	2018-10-01	2022-09-30
KATHAROPOULOS	Angelos	GR	François	FLEURET	2017-03-01	2021-12-31
KHODABAKHSHANDEH	Hamid	DE	Philip	GARNER	2021-09-01	2021-10-31
KULAK	Thibaut	FR	Sylvain	CALINON	2017-12-01	2021-11-30
LEMBONO	Teguh	ID	Sylvain	CALINON	2018-07-23	2022-07-23
LÖW	Tobias	DE	Sylvain	CALINON	2021-01-04	2024-12-31
MAI	Florian	DE	James	HENDERSON	2018-10-01	2022-09-30
MARELLI	François	BE	Michael	LIEBLING	2018-10-01	2022-09-30
MARFURT	Andreas	CH	James	HENDERSON	2018-11-01	2022-10-31
MARTINEZ GONZALEZ	Angel	MX	Jean-Marc	ODOBEZ	2016-11-15	2021-08-15
MATOBA	Kyle	US	François	FLEURET	2020-08-01	2023-07-31
MEEGAHAPOLA	lakmal	LK	Daniel	GATICA-PEREZ	2019-06-24	2023-06-23
MOHAMMADSHAHI	Alireza	IR	James	HENDERSON	2019-09-01	2022-08-31
MOSTAANI	Zohreh	IR	Mathew	MAGIMAI DOSS	2020-02-01	2024-01-31
NIGMATULINA	Iuliia	RU	Petr	MOTLICEK	2021-01-18	2025-01-18
OTROSHI SHAHREZA	Hatef	IR	Sébastien	MARCEL	2020-03-01	2024-02-29
PANNATIER	Arnaud	CH	François	FLEURET	2020-03-01	2024-02-29
PRASAD	Amrutha	IN	Petr	MOTLICEK	2020-08-01	2024-07-31
PUROHIT	Tilak	IN	Mathew	MAGIMAI DOSS	2021-06-01	2025-05-31
RAZMJOO FARD	Amirreza	IR	Sylvain	CALINON	2020-12-01	2024-11-30
SAJADMANESH	Sina	IR	Daniel	GATICA-PEREZ	2019-05-15	2023-05-15
SARKAR	Eklavya	CH	Mathew	MAGIMAI DOSS	2021-03-01	2025-02-28
SCHNELL	Bastian	DE	Philip	GARNER	2017-05-01	2021-08-31
SHETTY	Suhan	IN	Sylvain	CALINON	2019-06-03	2023-06-03

SIEGFRIED	Rémy	CH	Jean-Marc	ODOBEZ	2017-01-01	2021-09-30
SIVAPRASAD	Prabhu	IN	François	FLEURET	2018-11-01	2022-10-31
SRINIVAS	Suraj	IN	François	FLEURET	2017-03-01	2021-09-30
TAFASCA	Samy	MA	Jean-Marc	ODOBEZ	2021-03-01	2025-02-28
TARIGOPULA	Neha	IN	Mathew	MAGIMAI DOSS	2021-03-08	2025-02-28
TORNAY	Sandrine	CH	Mathew	MAGIMAI DOSS	2016-08-01	2021-01-31
UNNERVIK	Alex	CH	Sébastien	MARCEL	2020-03-01	2024-02-29
VYAS	Apoorv	IN	Hervé	BOURLARD	2018-07-15	2022-07-15
XUE	Teng	CN	Sylvain	CALINON	2021-11-01	2025-10-31
ZULUAGA GOMEZ	Juan Pablo	CO	Petr	MOTLICEK	2020-01-01	2023-12-31

4.4 Interns

Last name	First name	Nationality	Supervisor		Start	Estimated End
ASSI	Karim	LB	Daniel	GATICA-PEREZ	2021-09-21	2022-01-31
BOGHETTI	Roberto	IT	Jérôme	KÄMPF	2020-12-01	2021-05-31
BOUTON-BESSAC	Emma	FR	Daniel	GATICA-PEREZ	2021-09-20	2022-01-31
BROS	Victor	FR	Sébastien	MARCEL	2021-02-01	2022-01-31
BRUDERMÜLLER	Lara	DE	Sylvain	CALINON	2021-03-01	2021-07-31
CARMINATI	Davide	IT	Sylvain	CALINON	2021-07-01	2021-12-31
CLERC	Rémi	CH	Michael	LIEBLING	2021-03-01	2021-05-31
FORNAROLI	Alessandro	IT	Daniel	GATICA-PEREZ	2021-09-21	2022-01-31
LANG	Inga	NO	Lonneke	VAN DER PLAS	2021-10-25	2022-04-25
NIGMATULINA	Iuliia	RU	Petr	MOTLICEK	2020-09-15	2021-01-17
NOBAR	Mahdi	IR	Sylvain	CALINON	2021-03-01	2021-08-27
PAUL	Oliver Joseph	GB	Jérôme	KÄMPF	2020-12-01	2021-02-28
PIRAS	Florian	BE	Michael	LIEBLING	2021-06-01	2021-11-30
POLVANI	Niccolò	IT	Michael	LIEBLING	2021-09-21	2021-12-31
SARKAR	Eklavya	CH	Mathew	MAGIMAI DOSS	2020-11-01	2021-02-28
THAYAPARAN	Mokanarangan	LK	André	FREITAS	2021-10-01	2022-09-30
THOO	Yong Joon	GB	Sylvain	CALINON	2021-02-01	2021-07-31
TI	Boyang	CN	Sylvain	CALINON	2021-02-24	2022-02-21
TOGNOLI	Marco	IT	Jérôme	KÄMPF	2021-07-01	2021-11-03
VALENTINO	Marco	IT	André	FREITAS	2021-09-01	2022-08-31
VERZAT	Colombine	FR	Olivier	BORNET	2020-08-01	2021-02-28
ELBANNA	Gasser	EG	Mathew	MAGIMAI DOSS	2021-07-12	2021-08-31
GIROUD	Mathieu	CH	Jérôme	KÄMPF	2021-06-28	2021-08-20
HADJMBAREK	Nadia	TN	Sylvain	CALINON	2021-07-12	2021-09-24
GEVERS	Louis	BE	Jean-Marc	ODOBEZ	2021-11-03	2022-03-31
HAEFLIGER	Garance	CH	Raphaëlle	LUISIER	2021-02-01	2021-02-19
JAIN	Anubhav	IN	Sébastien	MARCEL	2020-10-01	2021-07-31
LINKE	Julian	DE	Philip	GARNER	2021-09-01	2021-11-30
SALAMIN	Chloé	CH	Petr	MOTLICEK	2021-03-15	2021-08-27
VASQUEZ-CORREA	Juan Camilo	CO	Mathew	MAGIMAI DOSS	2021-03-01	2021-04-30

4.5 Visitors

Last name	First name	Nationality	Supervisor		Start	Estimated End
VILLATORO TELLO	Esaú	MX	Hervé	BOURLARD	2019-09-01	2021-10-31

5 Active and Granted Projects in 2021

An overview of the projects that have been active during the year 2021 is presented in Section 5.1. The projects are grouped in three categories, namely National Research Projects, European and International Research Projects and Industry-oriented Projects.

Section 5.2 presents the list of projects accepted during 2021 but starting in the following year.

5.1 Projects in Progress during 2021

5.1.1 National Research Projects

[1] Name **2000LAKES** (Alpine research and citizen science toward the microbial conservation of high-mountain lakes in Switzerland)
Funding EPFL-UNIL CLIMACT initiative
Coordinator Ecole Polytechnique Federale de Lausanne
Duration 2021.11.01 - 2022.10.31
Partner(s) Idiap Research Institute; University of Lausanne

[2] Name **ADEL** (Automatic Detection of Leadership from Voice and Body)
Funding UNIL
Coordinator University of Lausanne
Duration 2020.06.01 - 2021.12.31
Partner(s) Idiap Research Institute, IMD Switzerland, Ecole Polytechnique Fédérale de Lausanne

[3] Name **AI4AUTISM2** (Digital Phenotyping of Autism Spectrum Disorders in Children)
Funding SNSF Sinergia
Coordinator University of Geneva
Duration 2021.11.01 - 2025.10.31
Partner(s) Idiap Research Institute; University of Applied Sciences and Arts of Southern Switzerland (SUPSI)

[4] Name **ALLIES** (Autonomous Lifelong learnIng intelLigent Systems)
Funding SNF - ERA-NET
Coordinator Idiap Research Institute
Duration 2018.01.01 - 2021.09.30
Partner(s) Laboratoire national de métrologie et d'essais, Université du Maine, Universitat Politecnica de Catalunya

[5] Name **CHASPEEPRO** (Characterisation of motor speech disorders and processes)
Funding SNSF Sinergia
Coordinator University of Geneva
Duration 2021.12.01 - 2025.11.30
Partner(s) Idiap Research Institute; University Hospitals of Geneva; University Sorbonne Nouvelle

[6] Name **CODIMAN** (A future that works: Cobotics, digital skills and the re-humanization of the workplace)
Funding SNF - NRP77
Coordinator Berner Fachhochschule
Duration 2020.05.01 - 2024.04.30
Partner(s) Idiap Research Institute

[7] Name **COMPBIO** (Computational biomicroscopy: advanced image processing methods to quantify live biological systems)
Funding SNF - Division II
Coordinator Idiap Research Institute
Duration 2018.04.01 - 2022.03.31
Partner(s) -

[8] Name **CORTI** (Computational Reduction for Training and Inference)
Funding SNF - Division II
Coordinator Idiap Research Institute
Duration 2020.03.01 - 2022.06.30
Partner(s) -

[9] Name **DOMAT** (On-demand Knowledge for Document-level Machine Translation)
Funding SNF - Division II
Coordinator Idiap Research Institute
Duration 2018.01.01 - 2022.09.30
Partner(s) HES-SO Vaud

[10] Name **DUSK2DAWN** (Characterizing Youth Nightlife Spaces, Activities, and Drinks)
Funding SNF - Sinergia
Coordinator Idiap Research Institute
Duration 2017.07.01 - 2021.06.30
Partner(s) University of Zurich, La Trobe University

[11] Name **EMIL** (Emotion in the loop – a step towards a comprehensive closed-loop deep brain stimulation in Parkinson's disease)
Funding SNF - Bridge Discovery
Coordinator University of Bern
Duration 2021.05.01 - 2025.04.30
Partner(s) Idiap Research Institute, Centre Suisse d'Electronique et de Microtechnique

[12] Name **EVOLANG** (The Origins and Future of Language)
Funding SNF - NCCR
Coordinator University of Zurich
Duration 2020.06.01 - 2024.05.31
Partner(s) Idiap Research Institute, Ecole Polytechnique Fédérale de Lausanne, Eidgenoessische Technische Hochschule Zuerich, University of Basel, University of Fribourg, University of Geneva, University of Lausanne, University of Neuchâtel, Zurich University of the Arts

[13] Name **HEALTHVLOGGING** (Social media culture and the (re)shaping of health-related practices by Youtubers)
Funding SNF Spark
Coordinator University of Lausanne
Duration 2019.12.01 - 2021.03.31
Partner(s) Idiap Research Institute

[14] Name **HEAP** (Human-Guided Learning and Benchmarking of Robotic Heap Sorting)
Funding SNF - ERA NET
Coordinator University of Lincoln
Duration 2019.04.01 - 2022.03.31
Partner(s) Idiap Research Institute, Istituto Italiano di Tecnologia, Institut de Recherche en Informatique et en Automatique, Technische Universitaet Wien

[15] Name **INTREPID** (Automated interpretation of political and economic policy documents: Machine learning using semantic and syntactic information)
Funding SNF - Sinergia
Coordinator Graduate Institute of International and Development Studies
Duration 2019.01.01 - 2022.12.31
Partner(s) Idiap Research Institute

[16] Name **IPEQ** (Uncertainty quantification and efficient design of experiments for data- and simulation-driven inverse problem solving)
Funding SNF - Division II
Coordinator Idiap Research Institute
Duration 2018.11.01 - 2022.10.31
Partner(s) University of California at Davis

[17] Name **LAOS** (Learning Representations of Abstraction in Text)
Funding SNF - Division II
Coordinator Idiap Research Institute
Duration 2018.11.01 - 2022.10.31
Partner(s) -

[18] Name **LEARN-REAL** (LEARNing physical manipulation skills with simulators using REAListic variations)
Funding SNF - ERA NET
Coordinator Idiap Research Institute
Duration 2019.04.01 - 2022.03.31
Partner(s) Ecole Centrale de Lyon, Istituto Italiano di Tecnologia

[19] Name **MOSPEEDI** (Motor Speech Disorders: characterizing phonetic speech planning and motor speech programming/execution and their impairments)
Funding SNF - Sinergia
Coordinator University of Geneva
Duration 2017.10.01 - 2021.09.30
Partner(s) Idiap Research Institute, University Hospitals of Geneva, Université Paris 3

[20] Name **NAST** (Neural Architectures for Speech Technology)
Funding SNF - Division II
Coordinator Idiap Research Institute
Duration 2020.02.01 - 2024.01.31
Partner(s) -

[21] Name **NATAI** (The Nature of Artificial Intelligence)
Funding SNSF - Agora
Coordinator Idiap Research Institute
Duration 2020.10.01 - 2023.09.30
Partner(s) Musée de la Main UNIL/CHUVV

[22] Name **NKBP** (Deep Learning Models for Continual Extraction of Knowledge from Text)
Funding SNF - Division II Lead Agency
Coordinator Idiap Research Institute
Duration 2020.10.01 - 2024.09.30
Partner(s) Katholieke Universiteit Leuven

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

[24] Name **SHISSM** (Sparse and hierarchical Structures for Speech Modeling)
Funding SNF - Division II
Coordinator Idiap Research Institute
Duration 2018.03.01 - 2022.02.28
Partner(s) -

[25] Name **SMILE-II** (Scalable Multimodal sign language technology for sIgn language Learning and assessmEnt Phase-II)
Funding SNF - Sinergia
Coordinator Idiap Research Institute
Duration 2021.01.01 - 2024.12.31
Partner(s) University of Applied Sciences of Special Needs Education, University of Surrey, University of Zurich

[26] Name **STEADI** (Storytelling Algorithm for Digital Interviews)
Funding SNF - Division I
Coordinator University of Neuchâtel
Duration 2021.02.01 - 2025.01.31
Partner(s) Idiap Research Institute, University of Lausanne

[27] Name **SWITCH** (Learning by Switching Roles in Physical Human-Robot Collaboration)
Funding SNF - Division II
Coordinator Idiap Research Institute
Duration 2021.03.01 - 2024.02.29
Partner(s) Jozef Stefan Institute

[28] Name **TIPS** (Towards Integrated processing of Physiological and Speech signals)
Funding SNF Division II
Coordinator Idiap Research Institute
Duration 2019.12.01 - 2023.11.30
Partner(s) Centre Suisse d'Electronique et de Microtechnique, Coaching & Moderation

5.1.2 European and International Research Projects

[1] Name **AI4MEDIA** (A European Excellence Centre for Media, Society and Democracy)
Funding H2020-ICT
Coordinator Centre for Research and Technology Hellas
Duration 2020.09.01 - 2024.08.31
Partner(s) Idiap Research Institute, Athens Technology Center, Aristotelio Panepistimio Thessalonikis, Commissariat a l'Energie Atomique et aux Energies Alternatives, Communaute d'Universites et Etablissements Universite Cote D'Azur, Deutsche Welle, Institute of Information Science and Technologies, F6S Network Limited, Flemish Radio and Television Broadcasting Organization, Fraunhofer-Gesellschaft, Globalz SA, Grassroots Arts And Research UG, HES-SO Valais, IBM Ireland Limited, Imagga Technologies Ltd, Interdigital RD France, Institut de Recherche et de Coordination Acoustique Musique, Joanneum Research Forschungsgesellschaft, Katholieke Universiteit Leuven, Modl.ai ApS, Polytechnic University of Bucharest, Queen Mary University of London, RAI-Radiotelevisione Italiana SPA, The Netherlands Institute for Sound and Vision, University of Amsterdam, University of Florence, University of Malta, University of Trento

[2] Name **ATCO2** (Automatic collection and processing of voice data from air-traffic communications)
Funding H2020-CSJU
Coordinator Idiap Research Institute
Duration 2019.11.01 - 2022.02.28
Partner(s) Brno University of Technology, OpenSky Network, ReplayWell, Romagna Tech, Evaluations and Language Resources Distribution Agency, University of Saarland

[3] Name **BATL** (Biometric Authentification with Timeless Learner)
Funding USA IARPA
Coordinator University of Southern California
Duration 2017.03.01 - 2021.09.30
Partner(s) Idiap Research Institute

[4] Name **AI4EU** (A European AI On Demand Platform and Ecosystem)
Funding H2020-RIA-ICT
Coordinator Thales Services SA
Duration 2019.01.01 - 2021.12.31
Partner(s) Idiap Research Institute, ABB AS, AGI Research SRO, Allianz SE, Atos Spain SA, Aristotle University of Thessaloniki, Blumorpho SAS, Budapest University of Technology and Economics, Bureau de Recherches Géologiques et Miniéres, Barcelona Supercomputing Center, CARTIF Foundation, Commissariat á l'énergie atomique et aux énergies alternatives, CINECA - Consorzio Interuniversitario, Consiglio Nationale delle Ricerche, Centre National de la Recherche Scientifique, Agencia Estatal Consejo Superior De Investigaciones Cientifícas, National Center for Scientific Research "Demokritos", German Research Center for Artificial Intelligence, German Aerospace Center, EIT Digital, Eötvös Loránd University, European Organisation for Security, FundingBox Research, Fondazione Bruno Kessler, Fraunhofer Gesellschaft, France Digitale, Foundation for Research and Technology – Hellas, Forum Virium Helsinki, Grassroots Arts and Research, France Hub IA, Industrial Data Space e. V., Institut Mines-Télécom-IMT, Institut National de Recherche en Informatique et Automatique, Associacao Do Instituto Superior Tecnico Para A Investigacao E Desenvolvimento, Centre for Research and Technology Hellas, Jožef Stefan Institute, Karlsruhe Institute of Technology, Know-Center Gmbh Research Center for Datadriven Business & Big data Analytics, University of Leeds, Loupe 16, Università degli studi di Roma "La Sapienza", Norwegian University of Science and Technology, National University of Ireland Galway, Office National d'Etudes et Recherches Aérospatiales, Orange SA, Örebro University, PG WConsulting, Université Paris I Panthéon-Sorbonne, QWANT, Siemens AG, SAP SE, Smile, Smart Rural, Simula Research Laboratory, Thales Alenia Space, Thomson Licensing, Telenor ASA, Tilde SIA, Thales Research & Technology France, Technische Universität Berlin, Delft University of Technology, Centre for Intelligent Technologies, Technical University of Munich, Technical University of Vienna, Alma Mater Studiorum – University of Bologna, University College Cork, University of Coimbra, Université Grenoble Alpes, Unilever UK, National and Kapodistrian University of Athens, Universitat Politecnica de Catalunya, Technical University of Madrid, Universita Degli Studi Di Siena, Sorbonne University, Ca' Foscari University of Venice, Vrije Universiteit Brussel, Women in AI, Wavestone

[5] Name **COLLABORATE** (Co-production CeLL performing Human-Robot Collaborative AssEmbly)
Funding H2020-RIA-DT
Coordinator Aristotle University Of Thessaloniki
Duration 2018.10.01 - 2022.03.31
Partner(s) Idiap Research Institute, Arcelik A.S., Association pour la Recherche et le Développement des méthodes et processes, ASTI Mobile Robotics, Blue Ocean Robotics APS, Centre for Research and Technology Hellas, Centro Ricerche Fiat SCPA, Jozef Stefan Institute, Katholieke Universiteit Leuven, Kolektor, University of Patras, Pratt & Whitney Rzeszów, Universita Degli Studi di Genova

[6] Name **CRITERIA** (Comprehensive data-driven Risk and Threat Assessment Methods for the Early and Reliable Identification, Validation and Analysis of migration-related risks)
Funding H2020-SU-SEC
Coordinator Gottfried Wilhelm Leibniz University Hannover
Duration 2021.09.01 - 2024.08.31
Partner(s) Idiap Research Institute; Centre for Research and Technology Hellas; Hensoldt Analytics GMBH; University of Groningen; Swedish National Police Authority; General Inspectorate of Romanian Border Police; Ministry of Interior Croatia; University of Malta; Estonian Police and Border Guard Board; Malta Police Force; Weblyzard Technology GmBH; European University Cyprus; ARSIS Association for the Social Support of Youth; Knowledge and Innovation Srls

[7] Name **ENERMAPS** (Open Source Tools to Share, Compare, and Reuse Low-Carbon Energy Data)
Funding H2020-LC
Coordinator Centre de recherches energetiques et municipales
Duration 2020.04.01 - 2022.06.30
Partner(s) Idiap Research Institute, Zentrum Für Energiewirtschaft Und Umwelt (E-THINK), Accademia Europea di Bolzano, OpenAire Make, Revolve Media, Technische Universitaet Wien

[8] Name **ePartners4All** (A (personalized and) blended care solution with virtual buddy for child health)
Funding Eureka (Innosuisse)
Coordinator TNO (Netherlands)
Duration 2021.11.15 - 2024.05.15
Partner(s) Idiap Research Institute, Bern University of Applied Sciences (CH), Eyeware Tech SA (CH), Leids Universitair Medisch Centrum, NL, Topicus SA (NL), MedVision SA (NL), Interactive Robotics B.V. (NL), XpertHealth (NL), Delft University of Technology (NL), Thearpieland SA (NL).

[9] Name **GRAIL** (Generative Range and Altitude Identity Learning)
Funding USA IARPA
Coordinator University of Southern California
Duration 2021.11.12 - 2025.11.11
Partner(s) Idiap Research Institute

[10] Name **HAAWAII** (Highly Automated Air Traffic Controller Workstations with Artificial Intelligence Integration)
Funding H2020-SESAR
Coordinator Deutsches Zentrum Fuer Luft und Raumfahrt Ev
Duration 2020.06.01 - 2022.11.30
Partner(s) Idiap Research Institute, Austro Control, Croatia Control, Isavia OHF, NATS, Brno University of Technology

[11] Name **ICARUS** (Innovative AppRoach to Urban Security)
Funding H2020-SU-SEC
Coordinator Forum Européen Pour La Sécurité Urbaine
Duration 2020.06.01 - 2024.08.31
Partner(s) Idiap Research Institute, Ethical and Legal Plus S.L., Erasmus Universiteit Rotterdam, Eurocircle Association, Fachhochschule Salzburg GmbH, Globalz SA, Kentro Meleton Asfaleias, University of Leeds, Lisbon Municipal Police, Makesense, Commune de Nice, Panteion University Of Social And Political Sciences, Riga Municipal Police, City of Rotterdam, Landeshauptstadt Stuttgart, City of Torino, University of Salford

[12] Name **MEMMO** (Memory of Motion)
Funding H2020-RIA-ICT
Coordinator Centre national de la recherche scientifique
Duration 2018.01.01 - 2022.06.30
Partner(s) Idiap Research Institute, University of Edinburgh, Max Planck Society for the Advancement of Sciences, University of Oxford, PAL ROBOTICS SL, AIRBUS SAS, Wandercraft, Centre de médecine physique et de réadaptation, Costain Group PLC

[13] Name **ROXANNE** (Real time network, text, and speaker analytics for combating organized crime)
Funding H2020-SU-SEC
Coordinator Idiap Research Institute
Duration 2019.09.01 - 2022.12.31
Partner(s) Trilateral Research LTD, Brno University of Technology, Phonexia s.r.o., SAIL LABS Technology GmbH, Capgemini Technology Services, The International Criminal Police Organization, Saarland University, KENTRO MELETON ASFALEIAS, Gottfried Wilhelm Leibniz Universitaet Hannover, Universitá Cattolica del Sacro Cuore – Transcrime, AEGIS IT RESEARCH UG, AIRBUS Defence and Space SAS (Innovation Coordinator), Police of Czech Republic, Romanian Minister of Interior, Lithuanian Forensic Science Centre, Police Service of Northern Ireland, ADITESS Advanced Integrated Technology Solutions & Services LTD, Ministry of Interior Croatia, Netherlands Forensic Institute, Internet of Things applications and Multi-Layer development, Ministry Of Public Security - Israel National Police, Hellenic Police, An Garda Síochána

[14] Name **SARAL** (Summarization and domain-Adaptive Retrieval of Information Across Languages)
Funding USA IARPA
Coordinator University of Southern California
Duration 2017.10.01 - 2021.10.22
Partner(s) Idiap Research Institute, Massachusetts Institute of Technology, Raytheon Company, Reenselaer Polytechnic Institute, University of Massachusetts Amherst, Northeastern University

[15] Name **SOTERIA** (uSer-friendly digiTal sEcured peRsonal data and prIvacy plAtform)
Funding H2020-SU-DS
Coordinator Ariadnext
Duration 2021.10.01 - 2024.09.30
Partner(s) Idiap Research Institute; Institut de Recherche en Informatique et en Automatique; Katholieke Universiteit Leuven; Centro de Vision por Computador; Audencia Business School; Erdyn Atlantique; Asociatia Infocons; Secure Electronic Voting SA; Stelar Security Technology Law Research UG; Noria Onlus; ipcenter.at GmbH; Servicio Vasco de Salud Osakidetza

[16] Name **TAPAS** (Training Network on Automatic Processing of PAthological Speech)
Funding H2020-MSCA
Coordinator Idiap Research Institute
Duration 2017.11.01 - 2022.09.30
Partner(s) University of Sheffield, Philips, Radboud University Nijmegen - Stichting Katholieke Universiteit, Ludwig-Maximilians-Universität München, Institut de Recherche En Informatique de Toulouse, Antwerpen University Hospital, Friedrich-Alexander-Universität Erlangen Nuernberg, Instituto de Engenharia de Sistemas E Computadores, Investigacao E desenvolvimento Em Lisboa, Interuniversitair Micro-Electronica Centrum Imec Vzw, Stichting Het Nederlands Kanker Instituut - Antoni Van Leeuwenhoek Ziekenhuis, Universitaet Augsburg

[17] Name **TRESPASS-ETN** (TRaining in Secure and PrivAcy-preserving biometricS)
Funding H2020-MSCA
Coordinator Eurecom
Duration 2020.01.01 - 2023.12.31
Partner(s) Idiap Research Institute, Hochschule Darmstadt, Chalmers Tekniska Hoegskola AB, Katholieke Universiteit Leuven, Rijksuniversiteit Groningen, Universidad Autónoma de Madrid

[18] Name **WENET** (The Internet of US)
Funding H2020-RIA-FETPROACT
Coordinator University of Trento
Duration 2019.01.01 - 2022.12.31
Partner(s) Idiap Research Institute, Aalborg University, Amrita Vishwa Vidyapeetham, Ben-Gurion University of the Negev, University of Tübingen, Instituto Potosino de Investigacion Cientifica y Tecnologica, Jilin University, London School of Economics and Political Science, Martel GmbH, National University of Mongolia, Open University of Cyprus, U-Hopper SRL, Universidad Catolica Nuestra Senora de La Ascuncion

5.1.3 Industry-oriented Projects

[1] Name **AADES** (Adaptive and Asynchronous Detection and Segmentation)
Funding Armasuisse
Coordinator Idiap Research Institute
Duration 2018.10.01 - 2022.09.30
Partner(s) -

[2] Name **AMS_SPONSORSHIP** (Sensor Fusion and Active Sensing for World-View Understanding)
Funding Industrial
Coordinator Idiap Research Institute
Duration 2019.09.01 - 2024.01.31
Partner(s) ams

[3] Name **BOAT** (Automated Braces generation for Orthopaedic Anatomical Treatment of fractures)
Funding Innosuisse
Coordinator Idiap Research Institute
Duration 2019.04.05 - 2021.03.31
Partner(s) HES-SO Fribourg, Swibrace SA

[4] Name **CANDY** (ContActless finger veiN recognition and presentation attack Detection on-the-flY)
Funding Innosuisse
Coordinator Idiap Research Institute
Duration 2020.03.01 - 2022.08.31
Partner(s) Global ID SA

[5] Name **CMM** (Conversation Member Match)
Funding Innosuisse
Coordinator Idiap Research Institute
Duration 2020.07.01 - 2022.06.30
Partner(s) Speak & Lunch SA

[6] Name **COBHOOK** (COB'HOOK)
Funding Innosuisse
Coordinator Idiap Research Institute
Duration 2019.05.01 - 2021.01.31
Partner(s) Richemont

[7] Name **COMINT** (Speech Technology for COMINT)
Funding Armasuisse
Coordinator Idiap Research Institute
Duration 2020.09.01 - 2021.01.31
Partner(s) -

[8] Name **DAHL** (Domain Adaptation via Hierarchical Lexicons)
Funding Industrial
Coordinator Idiap Research Institute
Duration 2019.11.01 - 2021.10.31
Partner(s) Swisscom

[9] Name **EGUZKI** (Programme de simulation de réseaux de chauffage à distance basé sur l'intelligence artificielle pour la résolution rapide et prédictive de réseaux complexes bouclés.)
Funding OFEN
Coordinator RWB Valais SA
Duration 2020.09.01 - 2023.08.30
Partner(s) Idiap Research Institute, Altis Groupe SA, Oiken SA, SATOM SA

[10] Name **GAZESENSESCREEN** (GazeSense Screen)
Funding Fondation The Ark
Coordinator Eyeware
Duration 2020.01.01 - 2022.01.31
Partner(s) Idiap Research Institute

[11] Name **HARDENING** (Heterogeneous face recognition for unified identity management)
Funding Innosuisse
Coordinator Idiap Research Institute
Duration 2021.02.01 - 2022.07.31
Partner(s) Facedapter Sarl

[12] Name **ICU** (Ilots de chaleur en ville de Fribourg : identification, anticipation et stratégie d'adaptation et de valorisation)
Funding OFEN
Coordinator HES-So Fribourg
Duration 2019.01.01 - 2021.03.31
Partner(s) Idiap Research Institute

[13] Name **INNO-MOBILET** (Smart Battery Network)
Funding Innosuisse
Coordinator Idiap Research Institute
Duration 2020.09.01 - 2021.05.31
Partner(s) Mobi-Let Sarl

[14] Name **IVECT** (Impact of greening on the energy balance and thermal comfort of buildings and districts)
Funding OFEN
Coordinator HES-So Valais
Duration 2020.12.01 - 2023.11.30
Partner(s) Idiap Research Institute, Centre de recherches energetiques et municipales, Etat du Valais, City of Zurich, Zurich University of Applied Sciences

[15] Name **LUCIDELES** (Leveraging User-Centric Intelligent Daylight and Electric Lighting for Energy Saving)
Funding OFEN
Coordinator University of Fribourg
Duration 2020.02.01 - 2022.07.31
Partner(s) Idiap Research Institute, Regent lighting AG

[16] Name **MALAT** (Machine Learning for Air Traffic)
Funding Innosuisse
Coordinator Idiap Research Institute
Duration 2020.03.01 - 2023.02.28
Partner(s) SkySoft ATM

[17] Name **MARGIN** (Multi-modal federated age verification)
Funding Innosuisse
Coordinator Idiap Research Institute
Duration 2020.07.01 - 2022.04.30
Partner(s) Privately SA, EPFL

[18] Name **P3** (P3: Press Pressure Prediction)
Funding Innosuisse
Coordinator HES-SO Valais
Duration 2021.06.01 - 2022.11.30
Partner(s) Idiap Research Institute; Constellium Valais SA

[19] Name **RISE** (Rich Interpersonal Skill analytics for rEcruitment)
Funding Innosuisse
Coordinator Idiap Research Institute
Duration 2018.05.01 - 2021.07.31
Partner(s) University of Lausanne

[20] Name **SPORTPROFILING** (Sport Profiling)
Funding Fondation The Ark
Coordinator Idiap Research Institute
Duration 2019.08.01 - 2021.11.15
Partner(s) Action Types Swiss Sàrl, ProKey Coach

[21] Name **3D2Cut** (Machine Learning for Tailor Made Vine Pruning)
Funding Fondation The Ark
Coordinator Idiap Research Institute
Duration 2020.08.01 - 2022.08.31
Partner(s) 3D2Cut SA

[22] Name **SRML** (Super-resolution through Machine Learning)
Funding Armasuisse
Coordinator Idiap Research Institute
Duration 2020.07.01 - 2021.01.31
Partner(s) -

[23]	Name	**STARFISH** (Safety and Speech Recognition with Artificial Intelligence in the Use of Air Traffic Control)
	Funding	Industrial
	Coordinator	Idiap Research Institute
	Duration	2020.10.31 - 2022.09.30
	Partner(s)	

[24]	Name	**USP** (Unique Stability Plates: Advanced Aluminium Solution for High Precision Milling)
	Funding	Innosuisse
	Coordinator	HES-SO Valais/Wallis
	Duration	2019.05.01 - 2021.02.28
	Partner(s)	Idiap Research Institute, Allega GmbH, Constellium Valais SA

[25]	Name	**WAVE2-96** (H2020-SESAR-PJ.10-W2-Solution 96)
	Funding	Industrial
	Coordinator	Idiap Research Institute
	Duration	2020.10.01 - 2022.07.31
	Partner(s)	-

5.2 Projects Awarded in 2021 and Starting in 2022

[1] Name **NEUMATH** (NeuMath: Neural Discourse Inference over Mathematical Texts)
Funding SNF - Division II
Coordinator Idiap Research Institute
Duration 2022.03.01 - 2025.02.28
Partner(s) -

[2] Name **SAFER** (reSponsible fAir FacE Recognition)
Funding Hasler Foundation
Coordinator Idiap Research Institute
Duration 2022.03.01 - 2026.02.28
Partner(s) University of Zurich; SICPA SA

[3] Name **COBHOOK-DEV** (CobHook Development Project)
Funding Industrial
Coordinator Idiap Research Institute
Duration 2022.01.01 - 2022.04.30
Partner(s) -

6 List of Publications in 2021

6.1 Book Chapters

[1] A. George and S. Marcel, "Multi-channel face presentation attack detection using deep learning," in *Deep Learning-Based Face Analytics*, Springer International Publishing, 2021.

[2] A. Mohanty, D. K. Mallick, S. Parida, and S. R. Dash, "Semantic behavior analysis of covid-19 patients: A collaborative framework," in *Machine Learning for Healthcare Applications*, John Wiley Sons, Inc. USA and Scrivener Publishing LLC, USA, 2021.

6.2 Articles in Scientific Journals

[1] C. Andreassi, R. Luisier, H. Crerar, M. Darsinou, S. Blokzijl-Franke, L. Tchern, N. M. Luscombe, G. Cuda, M. Gaspari, A. Saiardi, and A. Riccio, "Cytoplasmic cleavage of impa1 3' utr is necessary for maintaining axon integrity," *Cell Reports*, 2021.

[2] J. R. E. L. Carmona, S. González-López, E. VILLATORO-TELLO, and J. M. García-Gorrostieta, "Analysis of vector representations in maintenance logs in the industry: Towards an information retrieval system," *Journal of Research in Computing Science*, May 2021.

[3] K. Dimitropoulos, P. Daras, S. Manitsaris, F. F. Leymarie, and S. Calinon, "Editorial: Artificial intelligence and human movement in industries and creation," *Frontiers in Robotics and AI*, vol. 8, p. 712 521, 2021.

[4] P. Florio, G. Peronato, A. T. D. Perera, A. Di Blasi, K. H. Poon, and J. Kämpf, "Designing and assessing solar energy neighborhoods from visual impact," *Sustainable Cities and Society*, 2021.

[5] T. de Freitas Pereira and S. Marcel, "Fairness in biometrics: A figure of merit to assess biometric verification systems," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2021.

[6] J. Fritsch and M. Magimai.-Doss, "Utterance verification-based dysarthric speech intelligibility assessment using phonetic posterior features," *IEEE Signal Processing Letters*, vol. 28, pp. 224 –228, 2021.

[7] X. Gao, J. Silverio, S. Calinon, M. Li, and X. Xiao, "Bilateral teleoperation with object-adaptive mapping," *Complex Intelligent Systems*, 2021.

[8] X. Gao, J. Silverio, E. Pignat, S. Calinon, M. Li, and X. Xiao, "Motion mappings for continuous bilateral teleoperation," *IEEE Robotics and Automation Letters*, vol. 6, no. 3, pp. 5048–5055, 2021.

[9] P. N. Garner and S. Tong, "A bayesian approach to recurrence in neural networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 8, pp. 2527–2537, 2021.

[10] A. Gomez-Alanis, J. A. Gonzalez-Lopez, S. P. Dubagunta, A. M. Peinado, and M. Magimai.-Doss, "On joint optimization of automatic speaker verification and anti-spoofing in the embedding space," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 1579–1593, 2021.

[11] C. Hagemann, G. E. Tyzack, D. M. Taha, H. Devine, L. Greensmith, J. Newcombe, R. Patani, A. Serio, and R. Luisier, "Automated and unbiased discrimination of als from control tissue at single cell resolution," *Brain Pathology*, 2021.

[12] F. Haneef, G. Pernigotto, A. Gasparella, and J. Kämpf, "Application of urban scale energy modelling and multi-objective optimization techniques for building energy renovation at district scale," *Sustainability*, vol. 13, no. 20, 2021.

[13] W. He, P. Motlicek, and J.-M. Odobez, "Neural network adaptation and data augmentation for multi-speaker direction-of-arrival estimation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 1303–1317, 2021.

[14] E. Hermann, H. Kamper, and S. Goldwater, "Multilingual and unsupervised subword modeling for zero-resource languages," *Computer Speech and Language*, vol. 65, Jan. 2021.

[15] J. Jankowski, H. Girgin, and S. Calinon, "Probabilistic adaptive control for robust behavior imitation," *IEEE Robotics and Automation Letters*, Jan. 2021.

[16] N. Jaquier, R. Haschke, and S. Calinon, "Tensor-variate mixture of experts for proportional myographic control of a robotic hand," *Robotics and Autonomous Systems*, vol. 142, p. 103 812, 2021.

[17] N. Jaquier, L. Rozo, D. G. Caldwell, and S. Calinon, "Geometry-aware manipulability learning, tracking and transfer," *International Journal of Robotic Research*, vol. 40, no. 2-3, pp. 624–650, 2021.

[18] D. Khashabi, A. Cohan, S. Shakeri, P. Hosseini, P. Pezeshkpour, M. Bitaab, F. Brahman, S. Ghazarian, A. Kabiri, rabeeh karimi mahabadi, O. Memarrast, A. Mosallanezhad, E. Noury, S. Raji, M. S. Rasooli, S. Sadeghi, E. S. Azer, N. S. Samghabadi, M. Shafaei, S. Sheybani, A. Tazarv, and Y. Yaghoobzadeh, "Parsinlu: A suite of language understanding challenges for persian," *TACL*, 2021.

[19] I. Kodrasi, "Temporal envelope and fine structure cues for dysarthric speech detection using convolutional neural networks," *IEEE Signal Processing Letters*, Sep. 2021.

[20] P. Korshunov and S. Marcel, "Improving generalization of deepfake detection with data farming and few-shot learning," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, Dec. 2021.

[21] T. Kulak and S. Calinon, "Combining social and intrinsically-motivated learning for multi-task robot skill acquisition," *IEEE Transactions on Cognitive and Developmental Systems*, 2021.

[22] T. Kulak, H. Girgin, J.-M. Odobez, and S. Calinon, "Active learning of bayesian probabilistic movement primitives," *IEEE Robotic and Automation Letters*, 2021.

[23] F. Labhart, S. Muralidhar, B. Massé, L. B. Meegahapola, E. Kuntsche, and D. Gatica-Perez, "Ten seconds of my nights: Exploring methods to measure brightness, loudness and attendance and their associations with alcohol use from video clips," *PLOS ONE*, 2021.

[24] R. Lee, O. Wysocki, A. Freitas, and e. al., "Longitudinal characterisation of haematological and biochemical parameters in cancer patients prior to and during covid-19 reveals features associated with outcome," *ESMO Open*, Feb. 2021.

[25] T. S. Lembono, E. Pignat, J. Jankowski, and S. Calinon, "Learning constrained distributions of robot configurations with generative adversarial network," *IEEE Robotics and Automation Letters*, 2021.

[26] G. Liu, Y. Yu, and J.-M. Odobez, "A differential approach for gaze estimation," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 43, no. 3, pp. 1092–1098, 2021.

[27] G. Marshall, C. Jay, and A. Freitas, "Number and quality of diagrams in scholarly publications is associated with number of citations," *Diagrams*, 2021.

[28] J. Meadows and A. Freitas, "Similarity-based equational inference in physics," *Physics Review Research*, 2021.

[29] L. B. Meegahapola and D. Gatica-Perez, "Smartphone sensing for the well-being of young adults: A review," *IEEE Access*, Dec. 2021.

[30] L. B. Meegahapola, F. Labhart, T.-T. Phan, and D. Gatica-Perez, "Examining the social context of alcohol drinking in young adults with smartphone sensing," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)*, vol. 5, no. 3, p. 26, Sep. 2021.

[31] L. B. Meegahapola, S. Ruiz-Correa, V. del Carmen Robledo-Valero, E. E. Hernandez-Huerfano, L. Alvarez-Rivera, R. Chenu-Abente, and D. Gatica-Perez, "One more bite? inferring food consumption level of college students using smartphone sensing and self-reports," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)*, vol. 5, no. 1, Mar. 2021.

[32] A. Mohammadshahi and J. Henderson, "Recursive non-autoregressive graph-to-graph transformer for dependency parsing with iterative refinement," *Transactions of the Association for Computational Linguistics (2021)*, vol. 9, p. 18, Mar. 2021.

[33] A. Naceri, T. Schumacher, Q. Li, S. Calinon, and H. Ritter, "Learning optimal impedance control during complex 3d arm movements," *IEEE Robotics and Automation Letters (RA-L)*, vol. 6, no. 2, pp. 1248–1255, 2021.

[34] V. S. Nallanthighal, Z. Mostaani, A. Härmä, H. Strik, and M. Magimai.-Doss, "Deep learning architectures for estimating breathing signal and respiratory parameters from speech recordings," *Neural Networks*, vol. 141, pp. 211–224, 2021.

[35] C. Oertel, P. Jonell, D. Kontogiorgos, K. A. Funes Mora, J.-M. Odobez, and J. Gustafson, "Towards an engagement-aware attentive artificial listener for multi-party interactions," *Frontiers in Robotics and AI*, vol. 8, p. 189, 2021.

[36] H. Otroshi Shahreza and S. Marcel, "Towards protecting and enhancing vascular biometric recognition methods via biohashing and deep neural networks," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2021.

[37] D. Panda, D. Panda, S. R. Dash, and S. Parida, "Extreme learning machines with feature selection using ga for effective prediction of fetal heart disease: A novel approach," *Informatica*, vol. 45, no. 3, Oct. 2021.

[38] J. Pidoux, P. Kuntz, and D. Gatica-Perez, "Declarative variables in online dating: A mixed-method analysis of a mimetic-distinctive mechanism," *Proceedings of the ACM on Human-Computer Interaction*, vol. 5, no. CSCW1, Apr. 2021.

[39] E. Pignat, J. Silverio, and S. Calinon, "Learning from demonstration using products of experts: Applications to manipulation and task prioritization," *International Journal of Robotics Research*, 2021.

[40] M. del Río Carral, L. Volpato, C. Michoud, T.-T. Phan, and D. Gatica-Perez, "Professional youtubers' health videos as research material: Formulating a multi-method design in health psychology," *Methods in Psychology, Special Issue on Innovations in Qualitative Research*, vol. 5, Dec. 2021.

[41] L. A. A. Romero, G. Ramírez-de-la Rosa, and E. VILLATORO-TELLO, "Topic analysis and tracking from mexico's president daily press briefing," *Journal of Research in Computing Science*, May 2021.

[42] R. Sahu, S. R. Dash, L. A. Cacha, R. R. Poznanski, and S. Parida, "Classifier implementation for spontaneous eeg activity during schizophrenic psychosis," *Computacion y Sistemas (CyS)*, vol. 25, no. 3, Sep. 2021.

[43] N. Schütz, A. Botros, S. B. Hassen, H. Saner, P. Buluschek, P. Urwyler, B. Pais, V. Santschi, D. Gatica-Perez, R. M. Müri, and T. Nef, "A sensor-driven visit detection system in older adults' homes: Towards digital late-life depression marker extraction," *IEEE Journal of Biomedical And Health Informatics*, Sep. 2021.

[44] N. Schütz, H. Saner, A. Botros, B. Pais, V. Santschi, P. Buluschek, D. Gatica-Perez, P. Urwyler, R. M. Müri, and T. Nef, "Contactless sleep monitoring for early detection of health deteriorations in community-dwelling older adults: Exploratory study," *JMIR Mhealth Uhealth*, vol. 9, no. 6, Jun. 2021.

[45] J. Sebastian, M. Sur, H. A. Murthy, and M. Magimai.-Doss, "Signal-to-signal neural networks for improved spike estimation from calcium imaging data," *PLoS Computational Biology*, vol. 17, no. 3, pp. 1–19, Mar. 2021.

[46] S. Shetty, J. Silverio, and S. Calinon, "Ergodic exploration using tensor train: Applications in insertion tasks," *IEEE Transactions on Robotics*, Jun. 2021.

[47] A. K. Tanwani, A. Yan, J. Lee, S. Calinon, and K. Goldberg, "Sequential robot imitation learning from observations," *International Journal of Robotics Research (IJRR)*, 2021.

[48] V. Todeschi, R. Boghetti, J. Kämpf, and G. Mutani, "Evaluation of urban scale building energy-use models and tools – application for the city of fribourg, switzerland," *Sustainability*, vol. 13, no. 7, 2021.

[49] G. E. Tyzack, J. Neeves, H. Crerar, P. Klein, O. Ziff, D. M. Taha, R. Luisier, N. M. Luscombe, and R. Patani, "Aberrant cytoplasmic intron retention is a blueprint for rna binding protein mislocalization in vcp-related amyotrophic lateral sclerosis," *Brain*, 2021.

[50] C. Verzat, J. Harley, R. Patani, and R. Luisier, "Image-based deep learning reveals the responses of human motor neurons to stress and vcp-related als," *Neuropathology and Applied Neurobiology*, 2021.

[51] E. VILLATORO-TELLO, S. Parida, S. Kumar, and P. Motlicek, "Applying attention-based models for detecting cognitive processes and mental health conditions," *Cognitive Computation*, p. 18, Jul. 2021.

[52] A. Wierling, V. J. Schwanitz, S. Altinci, M. Bałazińska, M. J. Barber, M. E. Biresselioglu, C. Burger-Scheidlin, M. Celino, M. H. Demir, R. Dennis, N. Dintzner, A. el Gammal, C. M. Fernández-Peruchena, W. Gilcrease, P. Gładysz, C. Hoyer-Klick, K. Joshi, M. Kruczek, D. Lacroix, M. Markowska, R. Mayo-García, R. Morrison, M. Paier, G. Peronato, M. Ramakrishnan, J. Reid, A. Sciullo, B. Solak, D. Suna, W. Süß, A. Unger, M. L. Fernandez Vanoni, and N. Vasiljevic, "Fair metadata standards for low carbon energy research—a review of practices and how to advance," *Energies*, vol. 14, no. 20, Oct. 2021.

[53] Q. Zhan, X. Xie, H. Chenguang, Z.-G. Juan, J. Wang, and H. Cheng, "Domain-adversarial based model with phonological knowledge for cross-lingual speech recognition," *Electronics*, vol. 10, no. 24, pp. 1–15, Dec. 2021.

6.3 PhD Theses

[1] S. P. Dubagunta, “Novel methods for incorporating prior knowledge for automatic speech assessment,” PhD thesis, École polytechnique fédérale de Lausanne (EPFL), Sep. 2021.

[2] W. He, “Deep learning approaches for auditory perception in robotics,” PhD thesis, École polytechnique fédérale de Lausanne, Mar. 2021.

[3] T. Kulak, “Learning strategies and representations for intuitive robot learning from demonstration,” PhD thesis, EPFL, Dec. 2021.

[4] O. Mariani, “Computational methods for live heart imaging with speed-constrained microscopes,” PhD thesis, EPFL, 2021.

[5] A. Martínez-González, “Efficient depth-based deep learning methods for multi-party pose estimation,” PhD thesis, École polytechnique fédérale de Lausanne, 2021.

[6] R. Siegfried, “Modeling and inferring attention between humans or for human-robot interactions,” PhD thesis, Ecole Polytechnique Federale de Lausanne, Dec. 2021.

[7] S. Srinivas, “Gradient-based methods for deep model interpretability,” PhD thesis, École polytechnique fédérale de Lausanne, 2021.

[8] S. Tornay, “Explainable phonology-based approach for sign language recognition and assessment,” PhD thesis, Ecole Polytechnique Fédérale de Lausanne, 2021.

6.4 Articles in Conference Proceedings

[1] C. Basurto, R. Boghetti, M. Colombo, M. Pappinutto, J. Nembrini, and J. Kämpf, "Implementation of machine learning techniques for the quasi real-time blind and electric lighting optimization in a controlled experimental facility," in *Journal of Physics: Conference Series*, ser. 2042, vol. 2021, IOP Publishing, Sep. 2021.

[2] C. Basurto, O. Paul, and J. Kämpf, "Machine learning techniques for the daylight and electric lighting performance predictions," in *Proceedings of Building Simulation 2021*, Sep. 2021.

[3] A. Bittar and P. N. Garner, "A bayesian interpretation of the light gated recurrent unit," in *Proceedings IEEE International Conference on Acoustics, Speech and Signal Processing*, Jun. 2021.

[4] A. Bogatu, N. Paton, M. Douthwaite, S. Davie, and A. Freitas, "Cost–effective variational active entity resolution," in *37th IEEE International Conference on Data Engineering (ICDE)*, 2021.

[5] R. Braun, S. Madikeri, and P. Motlicek, "A comparison of methods for oov-word recognition on a new public dataset," in *2021 IEEE International Conference on Acoustics, Speech and Signal Processing*, IEEE Signal Processing Society, Toronto, Ontario, Canada, Jun. 2021.

[6] V. Bros, K. Kotwal, and S. Marcel, "Vein enhancement with deep auto-encoders to improve finger vein recognition," in *Biometrics Special Interest Group (BIOSIG 2021)*, 2021.

[7] L. Brudermuller, T. S. Lembono, S. Shetty, and S. Calinon, "Trajectory prediction with compressed 3d environment representation using tensor train decomposition," in *International Conference on Advanced Robotics*, Dec. 2021.

[8] L. Colbois, T. de Freitas Pereira, and S. Marcel, "On the use of automatically generated synthetic image datasets for benchmarking face recognition," in *International Joint Conference on Biometrics (IJCB 2021)*, Accepted for Publication in IJCB2021, 2021.

[9] C. Dancette, R. Cadene, D. Teney, and M. Cord, "Beyond question-based biases: Assessing multimodal shortcut learning in visual question answering," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021.

[10] E. Dantec, R. Budhiraja, A. Roig, T. S. Lembono, G. Saurel, O. Stasse, P. Fernbach, S. Tonneau, S. Vijayakumar, S. Calinon, M. Taix, and N. Mansard, "Whole body model predictive control with a memory of motion:experiments on a torque-controlled talos," in *IEEE International Conference on Robotics and Automation*, 2021.

[11] C. Dromart, L. Puthod, J. Kämpf, and D. von Gunten, "District heating network modelling for future integration of solar thermal energy," in *Journal of Physics: Conference Series*, vol. 2042, IOP Publishing, Nov. 2021, p. 012 089.

[12] M. Ewerton, S. Calinon, and J.-M. Odobez, "An attention mechanism for deep q-networks with applications in robotic pushing," in *Proc. of Workshop on Emerging paradigms for robotic manipulation: From the lab to the productive world, ICRA*, 2021.

[13] M. Ewerton, A. Martínez-González, and J.-M. Odobez, "An efficient image-to-image translation hourglass-based architecture for object pushing policy learning," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2021.

[14] M. Fabien and P. Motlicek, "Open-set speaker identification pipeline in live criminal investigations," in *1st ISCA Symposium on Security and Privacy in Speech Communication*, 2021.

[15] M. Fabien, S. Parida, D. Zhu, P. Motlicek, A. Krishnan, and H. H. Nguyen, "Roxanne research platform: Automate criminal investigations," in *Interspeech Show and Tell 2021*, Jun. 2021.

[16] M. Fabien, S. S. Sarfjoo, S. Madikeri, and P. Motlicek, "Graph2speak: Improving speaker identification using network knowledge in criminal conversational data," in *1st ISCA Symposium on Security and Privacy in Speech Communication*, 2021.

[17] A. George and S. Marcel, “Cross modal focal loss for rgbd face anti-spoofing,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.

[18] A. George and S. Marcel, “On the effectiveness of vision transformers for zero-shot face anti-spoofing,” in *International Joint Conference on Biometrics (IJCB 2021)*, 2021.

[19] H. Girgin, T. S. Lembono, R. Cirligeanu, and S. Calinon, “Optimization of robot configurations for motion planning in industrial riveting,” in *Proc. IEEE Intl Conf. on Advanced Robotics (ICAR)*, 2021.

[20] M. Granero-Moya, T.-T. Phan, and D. Gatica-Perez, “Zurich like new: Analyzing open urban multimodal data,” in *Proceedings of the 1st International Workshop on Multimedia Computing for Urban Data*, Oct. 2021.

[21] B. Halpern, J. Fritsch, E. Hermann, R. Van Son, O. Scharenborg, and M. Magimai.-Doss, “An objective evaluation framework for pathological speech synthesis,” in *Proceedings of ITG Conference on Speech Communication*, 2021.

[22] W. He, P. Motlicek, and J.-M. Odobez, “Multi-task neural network for robust multiple speaker embedding extraction,” in *Proceedings of Interspeech 2021*, 2021.

[23] H. Helmke, M. Kleinert, S. Shetty, O. Ohneiser, h. Ehr, H. Arilíusson, T. S. Simiganoschi, A. Prasad, P. Motlicek, K. Vesely, K. Ondřej, P. Smrz, J. Harfmann, and C. Windisch, “Readback error detection by automatic speech recognition to increase atm safety,” in *Fourteenth USA/Europe Air Traffic Management Research and Development Seminar (ATM2021)*, The United States Federal Aviation Administration (FAA), EUROCONTROL, Sep. 2021, p. 10.

[24] H. Helmke, S. Shetty, M. Kleinert, O. Ohneiser, h. Ehr, A. Prasad, P. Motlicek, C. Aneta, and C. Windisch, “Measuring speech recognition and understanding performance in air traffic control domain beyond word error rates,” in *11th SESAR Innovation Days*, 2021.

[25] E. Hermann and M. Magimai.-Doss, “Handling acoustic variation in dysarthric speech recognition systems through model combination,” in *Proceedings of Interspeech*, 2021.

[26] A. Jain, P. Korshunov, and S. Marcel, “Improving generalization of deepfake detection by training for attribution,” in *International Workshop on Multimedia Signal Processing*, Oct. 2021.

[27] P. Janbakhshi and I. Kodrasi, “Supervised speech representation learning for parkinson’s disease classification,” in *ITG Conference on Speech Communication*, Sep. 2021.

[28] P. Janbakhshi, I. Kodrasi, and H. Bourlard, “Automatic dysarthric speech detection exploiting pairwise distance-based convolutional neural networks,” in *45th International Conference on Acoustics, Speech, and Signal Processing*, Toronto, Canada, May 2021, 7328–7332.

[29] M. M. Johari, C. Carta, and F. Fleuret, “Depthinspace: Exploitation and fusion of multiple video frames for structured-light depth estimation,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, Oct. 2021, pp. 6039–6048.

[30] Z.-G. Juan, N. Iuliia, A. Prasad, P. Motlicek, K. Vesely, M. Kocour, and I. Szoke, “Contextual semi-supervised learning: An approach to leverage air-surveillance and untranscribed atc data in asr systems,” in *Interspeech 2021*, Aug. 2021.

[31] A. Khosravani, P. N. Garner, and A. Lazaridis, “An evaluation benchmark for automatic speech recognition of german-english code-switching,” in *IEEE Automatic Speech Recognition and Understanding Workshop*, Dec. 2021.

[32] A. Khosravani, P. N. Garner, and A. Lazaridis, “Learning to translate low-resourced swiss german dialectal speech into standard german text,” in *IEEE Automatic Speech Recognition and Understanding Workshop*, Colombia, Cartagena: IEEE, Dec. 2021.

[33] A. Khosravani, P. N. Garner, and A. Lazaridis, “Modeling dialectal variation for swiss german automatic speech recognition,” in *Proceedings of Interspeech*, Aug. 2021.

[34] M. Kleinert, H. Helmke, S. Shetty, O. Ohneiser, h. Ehr, A. Prasad, P. Motlicek, and J. Harfmann, "Automated interpretation of air traffic control communication: The journey from spoken words to a deeper understanding of the meaning," in *2021 IEEE/AIAA 40th Digital Avionics Systems Conference (DASC)*, San Antonio, TX, USA: IEEE, Oct. 2021, pp. 1–9.

[35] M. Kocour, K. Vesely, B. Alexander, Z.-G. Juan, I. Szoke, J. Cernocky, D. Klakow, and P. Motlicek, "Boosting of contextual information in asr for air-traffic call-sign recognition," in *Interspeech 2021*, Aug. 2021.

[36] M. Kocour, K. Vesely, I. Szoke, S. Kesiraju, Z.-G. Juan, B. Alexander, A. Prasad, N. Iuliia, P. Motlicek, and et al., "Automatic processing pipeline for collecting and annotating air-traffic voice communication data," in *Proceedings of 9th OpenSky Symposium 2020*, OpenSky Network, Brussels, Belgium: MDPI, Nov. 2021, pp. 1–9.

[37] I. Kodrasi, M. Pernon, M. Laganaro, and H. Bourlard, "Automatic and perceptual discrimination between dysarthria, apraxia of speech, and neurotypical speech," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, Jun. 2021.

[38] P. Korshunov and S. Marcel, "Subjective and objective evaluation of deepfake videos," in *The international Conference on Acoustics, Speech, and Signal Processing*, Jun. 2021.

[39] T. Kulak and S. Calinon, "Intrinsically-motivated robot learning of bayesian probabilistic movement primitives," in *ICRA workshop: "Towards Curious Robots: Modern Approaches for Intrinsically-Motivated Intelligent Behavior"*, 2021.

[40] T. S. Lembono and S. Calinon, "Probabilistic iterative lqr for short time horizon mpc," in *International Conference on Intelligent Robots and Systems*, 2021.

[41] D. Lindner, K. Matoba, and A. Meulemans, "Challenges for using impact regularizers to avoid negative side effects," in *SafeAI 2021 - AAAI's Workshop on Artificial Intelligence Safety*, 2021.

[42] Z. Liu, C. Rodriguez-Opazo, D. Teney, and S. Gould, "Image retrieval on real-life images with pre-trained vision-and-language models," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021.

[43] T. Löw, T. Bandyopadhyay, J. Williams, and P. Borges, "Prompt: Probabilistic motion primitives based trajectory planning," in *Proceedings of Robotics: Science and Systems*, Virtual, Jul. 2021.

[44] S. Madikeri, P. Motlicek, and H. Bourlard, "Multitask adaptation with lattice-free mmi for multi-genre speech recognition of low resource languages," in *Proceedings of Interspeech 2021*, 2021.

[45] rabeeh karimi mahabadi, y. belinkov, and J. Henderson, "Variational information bottleneck for effective low-resource fine-tuning," in *ICLR*, 2021.

[46] rabeeh karimi mahabadi, J. Henderson, and sebastian ruder, "Compacter: Efficient low-rank hypercomplex adapter layers," in *NeurIPS*, 2021.

[47] rabeeh karimi mahabadi, sebastian ruder, d. mostafa, and J. Henderson, "Parameter-efficient multi-task fine-tuning for transformers via shared hypernetworks," in *ACL*, 2021.

[48] F. Marelli and M. Liebling, "Optics versus computation: Influence of illumination and reconstruction model accuracy in focal-plane-scanning optical projection tomography," in *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, Nice, France: IEEE, Apr. 2021, pp. 567–570.

[49] A. Marfurt and J. Henderson, "Sentence-level planning for especially abstractive summarization," in *Proceedings of the Third Workshop on New Frontiers in Summarization*, Online and in Dominican Republic: Association for Computational Linguistics, Nov. 2021, pp. 1–14.

[50] O. Mariani, F. Marelli, C. Jaques, A. Ernst, and M. Liebling, "Unequivocal cardiac phase sorting from alternating ramp- and pulse-illuminated microscopy image sequences," in *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, Apr. 2021, pp. 868–872.

[51] G. Marshall, C. Jay, and A. Freitas, “Scholarly ai system diagrams as an access point to mental models,” in *Diagrams*, 2021.

[52] G. Marshall, C. Jay, and A. Freitas, “Structuralist analysis for neural network system diagrams,” in *Diagrams*, 2021.

[53] G. Marshall, M. Thayaparan, P. Osborne, and A. Freitas, “Switching contexts: Transportability measures for nlp,” in *14th International Conference on Computational Semantics*, 2021.

[54] A. Martínez-González, M. Villamizar, and J.-M. Odobez, “Pose transformers (potr): Human motion prediction with non-autoregressive transformers,” in *International Conference in Computer Vision - Workshops*, 2021.

[55] D. Mendes and A. Freitas, “Star: Cross-modal statement representation for selecting relevant mathematical premises,” in *16th conference of the European Chapter of the Association for Computational Linguistics (EACL)*, 2021.

[56] D. Mendes, J. Rozanova, M. Thayaparan, M. Valentino, and A. Freitas, “Does my representation capture x? probe-ably,” in *59th Annual Meeting of the Association for Computational Linguistics (Demonstration track)*, ser. Demonstration paper, 2021.

[57] G. Mercatali and A. Freitas, “Disentangling generative factors in natural language with discrete variational autoencoders,” in *The 2021 Conference on Empirical Methods in Natural Language Processing*, 2021.

[58] A. Mohammadshahi and J. Henderson, “Syntax-aware graph-to-graph transformer for semantic role labelling,” in *Arxiv*, Apr. 2021.

[59] Z. Mostaani, V. S. Nallanthighal, A. Härmä, H. Strik, and M. Magimai.-Doss, “On the relationship between speech-based breathing signal prediction evaluation measures and breathing parameters estimation,” in *Proc. of ICASSP*, 2021.

[60] T. Nakazawa, H. Nakayama, C. Ding, R. Dabre, S. Higashiyama, H. Mino, I. Goto, W. Pa Pa, A. Kunchukuttan, S. Parida, O. Bojar, C. Chu, A. Eriguchi, K. Abe, Y. Oda, and S. Kurohashi, “Overview of the 8th workshop on asian translation,” in *Proceedings of the 8th Workshop on Asian Translation (WAT2021)*, Association for Computational Linguistics, Aug. 2021, pp. 1–45.

[61] V. S. Nallanthighal, A. Härmä, H. Strik, and M. Magimai.-Doss, “Phoneme based respiratory analysis of read speech,” in *Proceedings of European Signal Processing Conference (EUSIPCO)*, 2021.

[62] H. H. Nguyen, M. Fabien, P. Motlicek, S. Parida, and K. Maly, “Roxsd: A simulated dataset of communication in organized crime,” in *1st ISCA Symposium on Security and Privacy in Speech Communication*, 2021.

[63] O. Ohneiser, S. S. Sarfjoo, H. Helmke, S. Shetty, P. Motlicek, M. Kleinert, h. Ehr, and Š. Murauskas, “Robust command recognition for lithuanian air traffic control tower utterances,” in *Interspeech*, 2021.

[64] H. Otroshi Shahreza, V. Krivokuca, and S. Marcel, “On the recognition performance of biohashing on state-of-the-art face recognition models,” in *Proceedings of the 13th IEEE International Workshop on Information Forensics and Security (WIFS)*, Montpellier, France: IEEE, Dec. 2021.

[65] H. Otroshi Shahreza and S. Marcel, “Deep auto-encoding and biohashing for secure finger vein recognition,” in *Proceedings of the 2021 IEEE International Conference on Acoustics, Speech, and Signal Processing*, IEEE, Toronto, Canada, 2021.

[66] D. Panchard, F. Marelli, E. De Moura Presa, P. Wellig, and M. Liebling, “Perspectives and limitations of visible-thermal image pair synthesis via generative adversarial networks,” in *Security + Defence, Target and Background Signatures VII, Proc. of SPIE*, vol. 11865, online only: SPIE, Sep. 2021, pp. 1 186 509–1–1186509–8.

[67] S. Parida, S. Panda, S. P. Biswal, K. Kotwal, A. Sen, S. R. Dash, and P. Motlicek, "Multimodal neural machine translation system for english to bengali," in *Proceedings of the First Workshop on Multimodal Machine Translation for Low Resource Languages (MMTLRL 2021)*, Online (Virtual Mode): INCOMA Ltd., Sep. 2021, pp. 31–39.

[68] S. Parida, S. Panda, A. R. Dash, E. VILLATORO-TELLO, A. S. Dogruöz, R. M. Ortega-Mendoza, A. Hernández, Y. Sharma, and P. Motlicek, "Open machine translation for low resource south american languages (americasnlp 2021 shared task contribution)," in *Proceedings of the First Workshop on Natural Language Processing for Indigenous Languages of the Americas*, Association for Computational Linguistics, Jun. 2021, 218–223.

[69] S. Parida, S. Panda, K. Kotwal, A. R. Dash, S. R. Dash, Y. Sharma, P. Motlicek, and O. Bojar, "Nlphut's participation at wat2021," in *Proceedings of the 8th Workshop on Asian Translation (WAT2021)*, Association for Computational Linguistics, Aug. 2021, pp. 146–154.

[70] G. Peronato, R. Boghetti, and J. Kämpf, "A machine-learning model for the prediction of aggregated building heating demand from pan-european land-use maps," in *Journal of Physics: Conference Series*, vol. 2042, 2021.

[71] R. Prasad and M. Magimai.-Doss, "Identification of f1 and f2 in speech using modified zero frequency filtering," in *Proceedings of Interspeech*, 2021.

[72] S. Purnapatra, N. Smalt, K. Bahmani, P. Das, D. Yambay, A. Mohammadi, A. George, T. Bourlai, S. Marcel, and S. Schuckers, "Face liveness detection competition (livdet-face) - 2021," in *International Joint Conference on Biometrics*, 2021.

[73] M. Ramirez, A. Bogatu, N. Paton, and A. Freitas, "Natural language inference over tables: Enabling explainable data exploration on data lakes," in *18th Extended Semantic Web Conference (ESWC)*, 2021.

[74] A. Razmjoo, T. S. Lembono, and S. Calinon, "Optimal control combining emulation and imitation to acquire physical assistance skills," in *Proc. IEEE Intl Conf. on Advanced Robotics (ICAR)*, 2021.

[75] M. A. Renzo, N. Fernandez, A. A. Baceti, N. N. de Moura Junior, and A. Anjos, "Development of a lung segmentation algorithm for analog imaged chest x-ray: Preliminary results," in *XV Brazilian Congress on Computational Intelligence*, Joinville, Brazil, Oct. 2021.

[76] D. Ribes Lemay, N. Henchoz, H. Portier, L. Defayes, T.-T. Phan, D. Gatica-Perez, and A. Sondereger, "Trust indicators and explainable ai: A study on user perceptions," in *Proc. Int. Conf. on Human-Computer Interaction*, Bari, Italy, Aug. 2021.

[77] J. Rozanova, D. Mendes, M. Thayaparan, M. Valentino, and A. Freitas, "Supporting context monotonicity abstractions in neural nli models," in *Natural Logic Meets Machine Learning Workshop*, 2021.

[78] S. K. Sahoo, B. K. Mishra, S. Parida, S. R. Dash, J. N. Besra, and E. VILLATORO-TELLO, "Automatic dialect detection for low resource santali language," in *Proceeding of International Conference on Information Technology (OCIT)*, 2021.

[79] S. Sajadmanesh and D. Gatica-Perez, "Locally private graph neural networks," in *ACM Conference on Computer and Communications Security (CCS)*, 2021.

[80] S. S. Sarfjoo, S. Madikeri, and P. Motlicek, "Speech activity detection based on multilingual speech recognition system," in *Interspeech*, 2021.

[81] L. Schelenz, I. Bison, M. Busso, A. de Götzen, D. Gatica-Perez, F. Giunchiglia, L. B. Meegahapola, and S. Ruiz-Correa, "The theory, practice, and ethical challenges of designing a diversity-aware platform for social relations," in *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, ACM, May 2021, p. 11.

[82] B. Schnell and P. N. Garner, "Improving emotional tts with an emotion intensity input from unsupervised extraction," in *11th ISCA Speech Synthesis Workshop*, Aug. 2021.

[83] A. Shajkofci and M. Liebling, “Estimating nonplanar flow from 2d motion-blurred widefield microscopy images via deep learning,” in *International Symposium on Biomedical Imaging, 2021*, 2021.

[84] R. Siegfried and J.-M. Odobez, “Visual focus of attention estimation in 3d scene with an arbitrary number of targets,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, IEEE, Jun. 2021, p. 9.

[85] J. Silverio, G. Clivaz, and S. Calinon, “A laser-based dual-arm system for precise control of collaborative robots,” in *IEEE International Conference on Robotics and Automation*, 2021.

[86] P. T. Sivaprasad and F. Fleuret, “Test time adaptation through perturbation robustness,” in *Workshop on Distribution Shifts, 35th Conference on Neural Information Processing Systems (NeurIPS 2021)*, 2021.

[87] P. T. Sivaprasad and F. Fleuret, “Uncertainty reduction for model adaptation in semantic segmentation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.

[88] S. Srinivas and F. Fleuret, “Rethinking the role of gradient-based attribution methods for model interpretability,” in *International Conference on Learning Representations*, 2021.

[89] M. Tanti, L. van der Plas, C. Borg, and A. Gatt, “On the language-specificity of multilingual bert and the impact of fine-tuning,” in *Proceedings of the Fourth BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks for NLP*, 2021.

[90] D. Teney, E. Abbasnejad, and A. van den Hengel, “Unshuffling data for improved generalization in visual question answering,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021.

[91] D. Teney, E. Abbasnejad, S. Lucey, and A. van den Hengel, “Evading the simplicity bias: Training a diverse set of models discovers solutions with superior ood generalization,” in *ArXiv*, May 2021.

[92] M. Thayaparan, M. Valentino, and A. Freitas, “Explainable inference over grounding-abstract chains for science questions,” in *59th Annual Meeting of the Association for Computational Linguistics (ACL Findings)*, 2021.

[93] C. Theodoropoulos, J. Henderson, A. C. Coman, and M.-F. Moens, “Imposing relation structure in language-model embeddings using contrastive learning,” in *Proceedings of the 25th Conference on Computational Natural Language Learning*, Online: Association for Computational Linguistics, Nov. 2021, pp. 337–348.

[94] M. Valentino, I. Pratt-Hartmann, and A. Freitas, “Do natural language explanations represent valid logical arguments? verifying entailment in explainable nli gold standards,” in *14th International Conference on Computational Semantics*, 2021.

[95] M. Valentino, M. Thayaparan, and A. Freitas, “Explainable natural language reasoning via conceptual unification,” in *16th conference of the European Chapter of the Association for Computational Linguistics (EACL)*, 2021.

[96] M. Valentino, M. Thayaparan, and A. Freitas, “Unification-based reconstruction of multi-hop explanations for science questions,” in *16th conference of the European Chapter of the Association for Computational Linguistics*, 2021.

[97] J. C. Vasquez-Correa, J. Fritsch, J. R. Orozco-Arroyave, E. Nöth, and M. Magimai.-Doss, “On modeling glottal source information for phonation assessment in parkinson’s disease,” in *Proceedings of Interspeech*, 2021.

[98] E. VILLATORO-TELLO, S. P. Dubagunta, J. Fritsch, G. Ramírez-de-la Rosa, P. Motlicek, and M. Magimai.-Doss, “Late fusion of the available lexicon and raw waveform-based acoustic modeling for depression and dementia recognition,” in *Proceedings of Interspeech 2021*, ISCA-International Speech Communication Association 2021, Aug. 2021.

[99] E. VILLATORO-TELLO, G. Ramírez-de-la Rosa, D. Gatica-Perez, M. Magimai.-Doss, and H. Jiménez-Salazar, "Approximating the mental lexicon from clinical interviews as a support tool for depression detection," in *Proceedings of the 2021 International Conference on Multimodal Interaction*, ACM, Oct. 2021.

[100] B. Vlasenko, R. Prasad, and M. Magimai.-Doss, "Fusion of acoustic and linguistic information using supervised autoencoder for improved emotion recognition," in *2nd Multimodal Sentiment Analysis Challenge (MuSe '21), October 24, 2021, Virtual Event, China*, 2021.

[101] A. Vyas, S. Madikeri, and H. Bourlard, "Comparing ctc and lfmmi for out-of-domain adaptation of wav2vec 2.0 acoustic model," in *Proceedings of Interspeech*, Sep. 2021.

[102] A. Vyas, S. Madikeri, and H. Bourlard, "Lattice-free mmi adaptation of self-supervised pretrained acoustic models," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Jun. 2021.

[103] H. Wang, J. Henderson, and P. Merlo, "Multi-adversarial learning for cross-lingual word embeddings," in *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Online, Jun. 2021, pp. 463–472.

[104] L. Wang, J. Zhu, and I. Kodrasi, "Multi-task single channel speech enhancement using speech presence probability as a secondary task training target," in *European Signal Processing Conference, EUSIPCO 2021*, Aug. 2021.

[105] O. Wysocki, M. Florea, D. Landers, and A. Freitas, "What is semeval evaluating? a systematic analysis of evaluation campaigns in nlp," in *EMNLP, 2nd Workshop on Evaluation and Comparison of NLP Systems*, 2021.

[106] Z. Zhou, M. Valentino, D. Landers, and A. Freitas, "Encoding explanatory knowledge for zero-shot science question answering," in *14th International Conference on Computational Semantics*, 2021.

7 List of Patents

While evaluating the success of technology transfer is a difficult task, invention disclosures and patents are usually key metrics in the field. As a necessary first step to any patent filing, an invention disclosure aims to identify a technology with a certain level of maturity and some promising economic potential.

In 2021, 16 invention and software disclosures have been filled. A patent committee examines each invention disclosure and decides to move forward with a patent filing or not. The committee is composed of the head of technology transfer, two representatives of the direction and one representative of the researchers. Since 2012, 11 patents have been filed: 6 patents have been granted, 3 applications are published and pending (1 of them has been sold) and 2 application were abandoned. In 2021, 3 patents have been granted: IDIAP-5, IDIAP-10 and IDIAP-11.

- Granted patents

  IDIAP-1 [US 9,689,959 B2] A. Asaei, H. Bourlard, V. Cevher, “Method, apparatus and computer program product for determining the location of a plurality of speech sources”

  IDIAP-2 [US 9,058,541 B2] C. Dubout, F. Fleuret, “Object detection method, object detector and object detection computer program”

  IDIAP-5 [EP3154407] K. A. Funes Mora, J-M. Odobez, “A gaze estimation method and apparatus”

  IDIAP-8 [US 9,973,503 B2] S. Marcel, A. Anjos, P. Abbet, “Method and internet-connected server for reviewing a computer-executable experiment”

  IDIAP-10 [EP 3 691 258 B1] M. Liebling, C. Jacques, “System and method for acquiring images”

  IDIAP-11 [EP 3 719 679 B1] S. Marcel, V. Krivokuca, “A method for protecting biometric templates, and a system and method for verifying a speaker’s identity”

- Pending applications

  IDIAP-9 [WO 2017/221049 A1] A. Anjos, S. Marcel, “A data-network connected server, a device, a platform and a method for conducting computer-executable experiments”

  globID-1 [WO 2019/150254 A1] L. Sonna Momo, L. Cerqueira Torres, S. Marcel, A. Anjos, M. Liebling, A. Shajkofci, S. Amoos, A. Woeffray, A. Sierro, “Method and device for biometric vascular recognition and/or identification”

- Sold applications

  IDIAP-4 [WO 2016/023582 A1] S. Marcel, “A method of detecting a falsified presentation to a vascular recognition system”. This patent has been sold.

- Abandonned applications

  IDIAP-6 [US 2017/0069306 A1] A. Asaei, M. Cernak, H. Bourlard, “Signal processing method and apparatus based on structured sparsity of phonological features”

  SAMSUNG-1 [US 2014/0149104 A1, EP 2736042 A1] N-H. Kim, P. Motlicek, P. N. Garner, D. Imseng, J-W. Lee, J-M. Cho, “Apparatus and method for constructing multilingual acoustic model and computer readable recording medium for storing program for performing the method”