IDIAP COMMUNICATION

Activity Report 2001

IDIAP–Com 2002-01

February 2002

Dalle Molle Institute
for Perceptual Artificial
Intelligence P.O.Box 592
Martigny Valais Switzerland

phone 41 27 721 77 11
fax 41 27 721 77 12
e-mail secretariat@idiap.ch
internet http://www.idiap.ch

Est. 1991

INTERACTIVE
MULTIMODAL
INFORMATION
MANAGEMENT

Est. 2001

Institut Dalle Molle d'Intelligence Artificielle Perceptive

“Villa Tissières”, one of the IDIAP facilities.

Contents

1 Introduction (in English) 1

2 Introduction (en français) 3

3 Major events 5
3.1 IDIAP Leading House of the IM2 National Centre of Competence in Research 5
3.2 First Tuesday, special on Artificial Intelligence 5
3.3 OSCAR PME Award 6
3.4 Tata Infotech 6
3.5 Intl. Computer Science Institute, Berkeley (CA, USA) 6
3.6 IEEE Neural Networks for Signal Processing, NNSP'02 7
3.7 Eurospeech'03 7
3.8 Tenth Anniversary 7

4 Staff 8
4.1 Scientific Staff 8
4.2 Students 10
4.3 System and Development Staff 10
4.4 Administrative Staff 10

5 Research Activities 11
5.1 Speech Processing Group 12
5.1.1 Research Themes 12
5.1.2 Application Examples 13
5.2 Computer Vision Group 13
5.2.1 Research Themes 14
5.2.2 Application Examples 15
5.3 Machine Learning Group 15
5.3.1 Research Themes 16
5.3.2 Application Examples 16

6 Current Projects 17

7 Educational Activities 33
7.1 Current PhD Theses 33
7.2 PhD Defenses 33
7.3 Participation in PhD Thesis Committees 33
7.4 Courses 35
7.5 Other student projects 35

8 Scientific Activities 36
8.1 Editorship 36
8.2 Scientific and Technical Committees 36
8.3 Short Term Visits 37
8.4 Scientific Presentations (other than conferences) 37

9 Publications (2000 and 2001) 39
9.1 Books and Book Chapters 39
9.2 Articles in International Journals 39
9.3 Articles in Conference Proceedings (refereed) 40
9.4 IDIAP Research Reports 42
9.5 IDIAP Communications 45
9.6 Other Documents 45

1 Introduction (in English)

Created in 1991 by the Dalle Molle Foundation for the Quality of Life, the Dalle Molle Institute for Perceptual Artificial Intelligence (IDIAP, http://www.idiap.ch), located in Martigny (Valais, Switzerland), is a not-for-profit research institute affiliated with the Swiss Federal Institute of Technology in Lausanne (EPFL) and the University of Geneva.

IDIAP is primarily funded by long-term support from the Swiss Confereration (Federal Office for Education and Science), the State of Valais, and the City of Martigny. The "Loterie Romande" also provides additional financial support to our research activities. In 2001, this long term funding amounted to approximately **30%** of the total IDIAP budget.

In addition to its long-term funding, IDIAP receives substantial research grants from the Swiss National Science Foundation (SNSF) for national (basic research and PhD) projects (representing about **25%** of the annual budget) and the Federal Office for Education and Science (OFES) for European projects (representing about **30%** of the budget). The rest of the funding (about **15%**) comes from collaboration with industry, and one CTI (Commission for Technology and Innovation) project.

In 2001, IDIAP numbered an average of 40-45 collaborators, including permanent scientific staff, postdoctoral fellows, PhD students (around 18), system and development engineers, and short-term to medium-term visitors.

The activities carried out at IDIAP can be described as follows: research and development activities, participation in European and national research projects, collaborations with organizations and companies, and teaching and training activities. IDIAP's mission therefore consists in:

> Carrying out fundamental and applied research activities aiming at long and medium term industrial transfer.
>
> Teaching and training activities.

In 2001, IDIAP's activities have continued to flourish, with a reasonable growth of the number of collaborative projects and publications, together with a constant increase of the quality of the research, now recognized at the international level. For example, the number of **national and international projects** has significantly increased, and many new projects were granted or started in 2001. As of this writing, there are about 14 SNSF and 13 European (EC/OFES) projects active at IDIAP. InfoVOX, a national project from CTI (Commission for Technology and Innovation), done in collaboration with EPFL, and involving companies like Swisscom, VOX-Access S.A. (the IDIAP spin-off company) and Omedia S.A., is also exploiting some of the IDIAP research results.

The value of a research institution is also assessed on the basis of its publications (number, but mainly quality). Here also, the average number of **international publications** is also consistently growing, resulting for the last two years in the following: 5 books or book chapters, 14 journal papers (compared to 11 in the previous Activity Report), 51 international conference papers (compared to 67 in previous Activity Report), and 46 unpublished (or not yet published) internal research reports (compared to 42 in previous Research Report).

Partnerships with academic institutions have also significantly been strengthened. In this framework, a major success for IDIAP in 2001 was its selection as "Leading House" of a **National Centre of Competence in Research** (NCCR). Resulting of a very strict and particularly competitive selection process, the NCCR on "Interactive Multimodal Information Management (IM)2", proposed by IDIAP and centered on many of its research activities, officially started on January 1, 2002. This long-term NCCR project (with funding planned for 10 to 12 years) was set up in collaboration with, and will involve, many national (EPFL, ETHZ, Univ. of Geneva, Univ. of Fribourg, Univ. of Bern) and international (ICSI in Berkeley, and Eurecom in Sophia-Antipolis) organizations. While strengthening further the links between IDIAP and many of its university partners, this project also confirms the leading role of IDIAP at the national level in the targeted research areas.

With an SNSF funding of CHF 15'400'000 and Self/Third party funding of CHF 16'220'000 for the next 4 years, this NCCR will certainly boost the activities at IDIAP, while also reinforcing its academic and industrial links.

Thanks to the continued support of our authorities, and to our most competent personnel, motivated to the highest level, IDIAP is thus recognized as a highly sought partner in the areas they decided to focus on (i.e., speech processing, computer vision and machine learning). It is now our job to continue to concentrate our research and development activities on those areas, while fostering technology transfer through industrial partnerships.

2 Introduction (en français)

Créé en 1991 par la Fondation Dalle Molle pour la Qualité de la Vie, l'Institut Dalle Molle d'Intelligence Artificielle Perceptive (IDIAP, http://www.idiap.ch), situé à Martigny (Valais, Switzerland), est un institut de recherche à but non lucratif affilié à l'Ecole Polytechnique Fédérale de Lausanne (EPFL) et à l'Université de Genève.

L'IDIAP est principalement financé à long terme par la confédération suisse (Office Fédéral de l'Education et de la Science – OFES), l'Etat du Valais, et la Ville de Martigny. La Loterie Romande supporte également nos activités de recherche au travers de soutiens financiers réguliers. En 2001, ce financement représentait environ **30%** du budget total de l'IDIAP.

En plus de son financement de base, l'IDIAP bénéficie de nombreux subsides de recherche au travers du Fonds National Suisse de la Recherche Scientifique (représentant environ **25%** du budget annuel) pour des projets de recherche fondamentale (étudiants doctorants) ainsi que de l'OFES pour les projets européens (représentant environ **30%** du budget). Le reste du financement de l'IDIAP (environ **15%**) provient de collaborations avec l'industrie et d'un projet CTI (Commission pour la Technologie et l'Innovation).

En 2001, l'IDIAP employait environ 40-45 collaborateurs, composés essentiellement de chercheurs permanents, de chercheurs post-doctoraux, d'ingénieurs doctorants (environ 18), d'ingénieurs systèmes et de développement, et de visiteurs à court ou moyen terme.

Les activités de l'IDIAP peuvent se répartir selon différentes catégories: les activités de recherche et développement, la participation à de nombreux projets de recherche européens et nationaux, les collaborations avec diverses organisations et sociétés, et les activités d'enseignement et de formation. La mission de l'IDIAP consiste donc en:

- La poursuite d'activités de recherche fondamentale et appliquée, dans le but de transfert technologique à moyen et long terme.
- L'enseignement et la formation.

En 2001, les activités de l'IDIAP ont été des plus florissantes, avec une bonne croissance du nombre de projets et de collaborations, ainsi que du nombre de publications, associé à une progression croissante de la qualité de sa recherche, maintenant reconnue au niveau international. Par exemple, le nombre de **projets nationaux et internationaux** a significativement augmenté, et plusieurs nouveaux projets ont démarré en 2001. A ce jour, environ 14 projets du Fonds National Suisse de la Recherche Scientifique et 13 projets européens (EC/OFES) sont actifs à l'IDIAP. InfoVOX, un projet national de la CTI (Commission pour la Technologie et l'Innovation), en partenariat avec l'EPFL et les sociétés Swisscom, VOXCom S.A. (société "spin-off" de l'IDIAP) et Omedia S.A., exploite aussi certains des résultats de recherche de l'IDIAP.

La valeur d'une institution de recherche scientifique est également jaugée à ses publications (nombre, mais surtout qualité). Ici aussi, le nombre moyen de **publications internationales** a continué à augmenter régulièrement, générant sur les deux dernières années les publications suivantes: 5 livres ou chapitres de livre, 14 articles dans des revues internationales (comparé à 11 pour le rapport d'activité précédent), 51 articles dans des conférences internationales (comparé à 67 pour le rapport d'activité précédent), et 46 rapports scientifiques internes non publiés ou en cours de publication (comparé à 42 pour le rapport d'activité précédent),.

Finalement, les **collaborations avec les institutions académiques** se sont également fortement renforcées. Dans ce contexte, un succès important pour l'IDIAP en 2001 a été sa sélection comme "Leading House" d'un **Pôle de Recherche National** (PRN). Résultant d'un long processus de sélection, strict et particulièrement sélectif, le PRN "Interactive Multimodal Information Management (IM)2" proposé par l'IDIAP, et centré sur de nombreuses activités de notre institut, a officiellement démarré au 1er janvier 2002. Ce projet PRN à long terme (avec un financement prévu pour 10 à 12 ans) a été élaboré en collaboration avec de nombreuses institutions nationales et internationales qui y participeront, dont notamment: EPFL, ETHZ, Université de Genève, Université de Fribourg, Université de Bern, ICSI (Berkeley) et Eurecom (Sophia-Antipolis). Tout en renforcant d'avantage les liens entre l'IDIAP et ces partenaires universitaires, ce projet devrait aussi confirmer la reconnaissance de l'IDIAP au niveau national dans ses domaines d'activité. Avec un financement du Fonds National de CHF 15'400'000 et un financement institutionnel ou de tiers de CHF 16'220'000 sur les 4 prochaines années, ce PRN va encore renforcer les activités de l'IDIAP, tout en consolidant ses liens académiques et industriels.

Grâce au support continu de nos autorités, ainsi qu'aux efforts de notre personnel des plus compétents et des plus motivés, l'IDIAP est maintenant reconnu comme un partenaire essentiel pour tous les développements touchant à ses domaines d'activité (à savoir le traitement de la parole, la vision par ordinateur, et l'apprentissage automatique). Notre mission est maintenant de continuer à concentrer nos activités de recherche et développement dans ces domaines de compétence, tout en favorisant le transfert technologique et les partenariats industriels.

3 Major events

3.1 IDIAP Leading House of the IM2 National Centre of Competence in Research

The major event for the year 2001 is of course the selection of the NCCR on Interactive Multimodal Information Management, (IM)2 in short. IDIAP will act as the Leading House of the NCCR, responsible for its scientific and administrative management. Research in this domain will be conducted within the network composed of teams from IDIAP, EPFL, the Universities of Geneva, Fribourg and Bern, as well as ETHZ and HEVs in Sion. Several industrial partners are also involved, and international collaboration will take place among other with a young researcher exchange program agreement signed with ICSI in Berkeley.

A presentation of the IM2 NCCR, including the history of the project, can be found in Appendix 1. Figures 1 and 2 show the various academic, industrial and institutional partners involved in the IM2 NCCR.

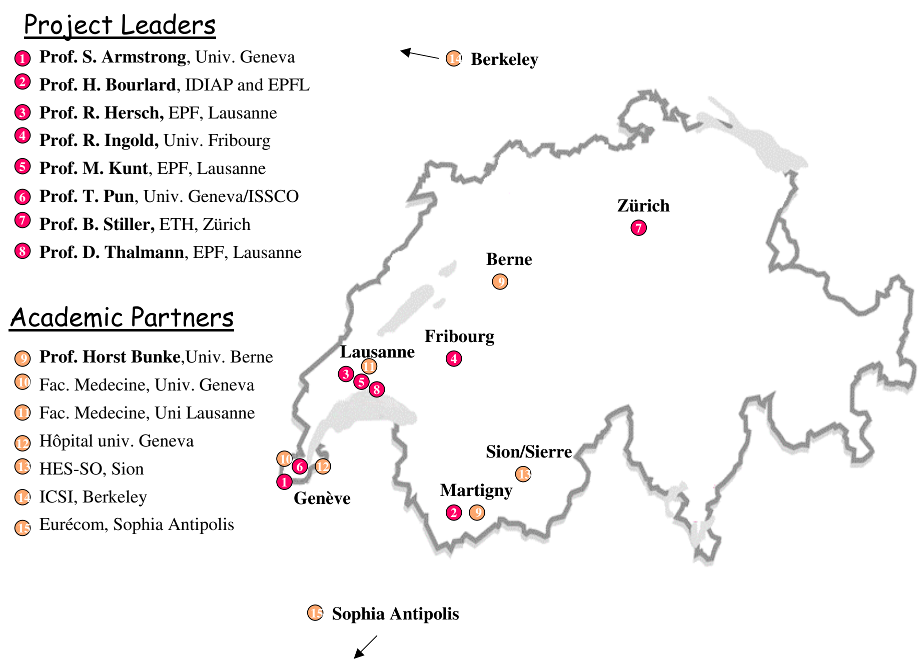


Figure 1: The academic partners of IM2.

3.2 First Tuesday, special on Artificial Intelligence

Based on a tradition of bringing together start-ups, entrepreneurs, investors, scientists, and professionals, First Tuesday with IDIAP, Icare, Sodeval, Cimtec-Valais and Centre du Parc gathered some of the top players in Artificial Intelligence and Multimodal Human-Computer Interaction to organize a must-attend event in Martigny, Switzerland on November, 16.

More than 20 venture capital companies, startups selected for an elevator pitch and several hundred professionals attended the day as well as companies from Canada, Europe and the US. An online video conference took place with the Swiss House in Boston and the Director of the MIT Artificial Intelligence Laboratory. Full details can be found at http://www.rezonance.ch/

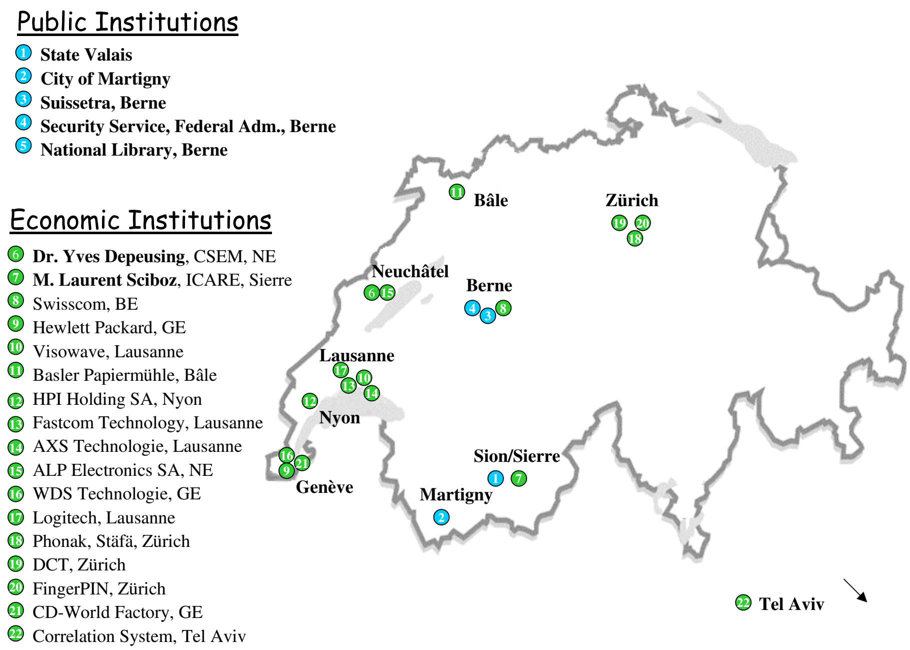


Figure 2: The public and private institutions involved in IM2.

3.3 OSCAR PME Award

On June 23, IDIAP received the OSCAR PME Award (of the Swiss PRD) as one of the most active enterprise in its field.

3.4 Tata Infotech

On November 1, a Memorandum of Understanding was signed between IDIAP and Tata Infotech Ltd., an international leading Company in the field of Information Technology (see `http://www.tatainfotech.com`). Similarly to IDIAP, the Tata Infotech Cognitive Systems Research Lab of has been engaged in research and development in multimodal human-computer interaction, including speech, script, natural language processing and artificial intelligence.

In this agreement, the two institutions have expressed their desire to facilitate mutual cooperation in research and technology development activities in areas of common interest, also including a visitor exchange program and the possibility of joint activities in projects of common interest.

3.5 Intl. Computer Science Institute, Berkeley (CA, USA)

The International Computer Science Institute, a non-profit Institute closely affiliated with the University of California at Berkeley, is entering into a long-term partnership with the IM2 National Centre of Competence in Research (NCCR), represented by IDIAP as the “Leading House”. ICSI has been a key player in the international speech research arena for over a decade, and in particular has been working with speech data from informal meetings for the last 2 years. Similar problems will be a key focus for the new research network, and there is much potential for collaboration. A visitor program for Swiss researchers to go to Berkeley is a key part of the program, which will allow a limited number of PhD students and postdocs scientists involved in

IM2 to go to ICSI for collaborative research.

3.6 IEEE Neural Networks for Signal Processing, NNSP'02

In September 2002, IDIAP will organize in Martigny the 2002 IEEE International Workshop on Neural Networks for Signal Processing (NNSP02). Details at http://eivind.imm.dtu.dk/nnsp2002/.

3.7 Eurospeech'03

IDIAP is organizing the next Eurospeech'2003 international conference, which will be held at the Intl. Congress Centre of Geneva in September 2003. Eurospeech is the premiere conference on speech and language technology, attracting more than 1000 scientists every two years. Details at http://www.eurospeech2003.org.

3.8 Tenth Anniversary

Last but not least, IDIAP celebrated in 2001 its 10^{th} anniversary. The official celebration was synchronised with the First Tuesday Special on Artificial Intelligence on November, 16 (see above).

To mark this major milestone, IDIAP has a new logo. Its web site http://www.idiap.ch/, which is one of the most important vector of communication with more than 500 visits every day, will be refreshed in early 2002, adopting a more fashionable graphic line.

4 Staff

Mail: IDIAP — Institut Dalle Molle d'Intelligence Artificielle Perceptive
Simplon 4, CP 592
CH–1920 Martigny (VS)
Switzerland

Phone: +41 - 27 - 721 77 11

Fax: +41 - 27 - 721 77 12

Internet: http://www.idiap.ch/

4.1 Scientific Staff

Persons at IDIAP in 2001 or as of this writing

Mr.	Jitendra AJMERA Jitendra.Ajmera@idiap.ch	Research assistant +41 27 721 77 48	1.1.2001
Mr.	Mark BARNARD Mark.Barnard@idiap.ch	Research assistant +41 27 721 77 29	15.3.2001
Dr.	Samy BENGIO Samy.Bengio@idiap.ch	Machine Learning group leader +41 27 721 77 39	
Mr.	Mohamed Faouzi BENZEGHIBA Mohamed.Benzeghiba@idiap.ch	Research assistant +41 27 721 77 41	
Prof.	Hervé BOURLARD Herve.Bourlard@idiap.ch	Director, Professor EPFL +41 27 721 77 20	
Mr.	Fabien CARDINAUX Fabien.Cardinaux@idiap.ch	Research assistant +41 27 721 77 55	1.10.2001
Mr.	Datong CHEN Datong.Chen@idiap.ch	Research assistant +41 27 721 77 56	
Ms.	Silvia CHIAPPA Silvia.Chiappa@idiap.ch	Research assistant +41 27 721 77 30	1.11.2001
Mr.	Ronan COLLOBERT Ronan.Collobert@idiap.ch	Research assistant	30.9.2001
Mr.	Christos DIMITRAKAKIS Christos.Dimitrakakis@idiap.ch	Research assistant +41 27 721 77 40	1.10.2001
Mr.	Beat FASEL Beat.Fasel@idiap.ch	Research assistant +41 27 721 77 23	
Mr.	Nicolas GILARDI Nicolas.Gilardi@idiap.ch	Research assistant +41 27 721 77 47	
Dr.	Astrid HAGEN astrid.hagen@idiap.ch	Research assistant	31.12.2001
Mr.	Shajith IKBAL Shajith.Ikbal@idiap.ch	Research assistant +41 27 721 77 46	

Prof.	Mikhael KANEVSKI Mikhael.Kanevski@idiap.ch	Research scientist +41 27 721 77 49	
Dr.	Sacha KRSTULOVIC sacha.krstulovic@idiap.ch	Research assistant	30.09.2001
Mr.	Quan LE Quan.Le@idiap.ch	Research assistant +41 27 721 77 36	1.2.2001
Mr.	Mathew MAGIMAI DOSS Mathew@idiap.ch	Research assistant +41 27 721 77 51	
Ms.	Christine MARCEL Christine.Marcel@idiap.ch	Development engineer +41 27 721 77 50	1.6.2001
Dr.	Sebastien MARCEL Sebastien.Marcel@idiap.ch	Research scientist +41 27 721 77 27	
Mr.	Johnny MARIÉTHOZ Johnny.Mariethoz@idiap.ch	Development engineer +41 27 721 77 44	
Dr.	Iain MCCOWAN Iain.Mccowan@idiap.ch	Research scientist +41 27 721 77 32	15.4.2001
Mr.	Hemant MISRA Hemant.Misra@idiap.ch	Research assistant +41 27 721 77 57	24.7.2001
Dr.	Andrew MORRIS Andrew.Morris@idiap.ch	Research scientist +41 27 721 77 35	
Dr.	Jean-Marc ODOBEZ Jean-Marc.Odobez@idiap.ch	Research scientist +41 27 721 77 26	1.9.2001
Ms.	Maja POPOVIC maja.popovic@sezampro.yu	Research assistant	31.8.2001
Dr.	Kim SHEARER tangowolf@bigfoot.com	Research scientist	31.3.2001
Mr.	Todd STEPHENSON Todd.Stephenson@idiap.ch	Research assistant +41 27 721 77 52	
Mr.	Alex TRUTNEV Alex.Trutnev@idiap.ch	Research assistant +41 27 721 77 38	
Mr.	Vivek TYAGI Vivek.Tyagi@idiap.ch	Research assistant +41 27 721 77 58	1.6.2001
Mr.	Alessandro VINCIARELLI Alessandro.Vinciarelli@idiap.ch	Research assistant +41 27 721 77 24	
Ms.	Katrin WEBER Katrin.Weber@idiap.ch	Research assistant +41 27 721 77 37	

4.2 Students

Mr.	Francois BELISLE belislfr@iro.umontreal.ca	1.5.2001	31.8.2001
Mr.	David BONNEVILLE david_bonneville@yahoo.fr	15.3.2001	15.9.2001
Mr.	Eric MCSWEEN Eric.Mcsween@idiap.ch	1.5.2001	31.8.2001
Ms.	Susagna POL FONT susagnapol@hotmail.com	18.09.2000	31.8.2001
Mr.	Pere PUJOL pere.pujol@upcnet.es	18.09.2000	31.3.2001

4.3 System and Development Staff

Mr.	Thierry COLLADO Thierry.Collado@idiap.ch	Development engineer +41 27 721 77 42	
Mr.	Laurent DEFAGO Laurent.Defago@idiap.ch	System engineer +41 27 721 77 25	1.3.2001
Mr.	Frank FORMAZ Frank.Formaz@idiap.ch	System Management group leader +41 27 721 77 28	
Ms.	Haiyan WANG Haiyan.Wang@idiap.ch	Development engineer +41 27 721 77 54	

4.4 Administrative Staff

Mr.	Pierre DAL PONT Pierre.DalPont@idiap.ch	Financial Manager +41 27 721 77 45	1.11.2001
Dr.	Jean-Albert FERREZ Jean-Albert.Ferrez@idiap.ch	Program Manager +41 27 721 77 19	1.11.2001
Ms.	Sylvie MILLIUS Sylvie.Millius@idiap.ch	Secretary +41 27 721 77 21	
Ms.	Nadine ROUSSEAU Nadine.Rousseau@idiap.ch	Secretary +41 27 721 77 22	

5 Research Activities

The focus of our activities is on the development of advanced (multimodal) natural input and output interfaces to a computer through speech and vision, as well on new ways to access multimedia documents.

The field of multimodal interaction covers a wide range of critical activities and applications, including recognition and interpretation of spoken, written and gestural language, particularly when used to interface with multimedia information systems. Other key subthemes include the biometric protection of information access (through speaker and/or face recognition and verification), and the structuring, retrieval and presentation of multimedia information.

The resulting multimodal interfaces are expected to represent a new direction for computing, providing people (including non-specialists) with access to complex information systems (e.g., incorporating multimedia content). Ultimately, these multimodal interfaces should flexibly accommodate a wide range of users, tasks, and environments for which any single mode may not suffice. The ideal interface should primarily be able to deal with more comprehensive and realistic forms of data, including mixed data types (i.e., data from different input modalities such as image and audio).

Although all the IDIAP research and development activities are structured in three groups (speech processing, computer vision, and machine learning) briefly described later, these activities can also be summarized as follow:

> **Spoken language input**: Covering speech signal processing and multilingual robust speech recognition. **Research issues** include: improved robustness, portability across new applications, language modeling, automatic adaptation (of acoustic and language models), confidence measures, out-of-vocabulary words, spontaneous speech, prosody, modeling dynamics.
>
> **Written language input**: Including document image analysis; OCR (printed and handwritten, off-line recognition); handwriting as computer interface (on-line recognition). **Research issues** include: analysis of documents with complex layout, recognition of degraded printed text, recognition of running handwriting.
>
> **Visual input**: Shape tracking (including lips tracking, face tracking); gesture recognition; facial expressions; images (e.g., sketches, signatures, photos) used as input. **Research issues** include: robustness of the algorithms; combination of colour, motion, texture, and shape in the analysis; more accurate model-based analysis; computational complexity.
>
> **Input (spoken, written, visual) analysis and understanding**, involving parsing and syntactic and semantic analysis and modeling. **Research issues** include: specification and formalism of unimodal and multimodal syntactical and semantic constraints, using these constraints into unimodal and multimodal input signal processing, merging modalities through multimodal “grammars”.
>
> **Protecting information access**, involving: speaker verification, signature recognition, face recognition; bio-metric (multimodal) user authentication. **Research issues** include: increasing robustness of user authentication techniques, multimodal user authentication (mixture of experts, confidence-based weighting of the different media, etc).
>
> **Modality integration**, involving, e.g.: Speech and gestures, facial movement and speech recognition, facial movement and speech synthesis, and interface agents. **Research issues** include: merging of different (media) data streams, possibly non-synchronous and with different data rate, fusion of the different modalities (e.g., based on signal-to-noise ratio or confidence level estimation).
>
> **Mathematical methods**, including: Statistical modeling and statistical pattern classification, signal processing techniques, connectionist techniques, expert fusion, support vector machines.

These research dimensions will appear in the research groups and projects described below.

5.1 Speech Processing Group

The overall goals of the IDIAP speech processing group are to research and develop robust recognition and understanding techniques for realistic speaking styles and acoustic conditions, as well as robust speaker verification and identification techniques. This includes advanced research activities, maintenance of language resources for the training and testing of recognition systems, and development of real-time prototypes. The group has been involved in speech research projects for several years and is today at the leading edge of technology. The IDIAP Speech Processing group is also involved in numerous national and European collaborative projects, as well as industrial projects.

5.1.1 Research Themes

1. Automatic recognition of (isolated, continuous, or natural) speech based on phonetic (sub-word) modeling, using spectral-temporal profiles of speech, as well as articulatory.
2. Development and improvement of state-of-the-art speech recognition systems based on hidden Markov models (HMM).
3. Using discriminant artificial neural networks (ANN) to estimate a posteriori probabilities. In this regard, IDIAP (in collaboration with ICSI, Berkeley) is recognized as a leader in the use of hybrid HMM/ANN systems, exhibiting several advantages compared to standard HMM approaches.
4. Estimation of confidence levels, i.e., attaching a confidence score to each recognized word to indicate how likely the word is correctly recognized. In this context, the problem of detection out-of-vocabulary words is also investigated.
5. Multi-stream and multi-band speech recognition: improving robustness of state-of-the-art systems based on multiple feature streams. This includes the extraction of multiple features from a same input utterance, exhibiting different properties, such as multiple temporal resolutions and/or containing some new, novel, or robust type of information. As a particular case, multi-band speech recognition, combining multiple (HMM or HMM/ANN) recognizers, has been shown to significantly improve robustness to narrow band noise.
6. Multi-stream combination: Developing novel methods to combine information generated from multiple experts trained on multi-stream features to improve word recognition and increase robustness of the recognition to corrupting environmental conditions.
7. Acoustic change detection and clustering, as required when dealing with large audio and multimedia databases (such as broadcast news and sport videos). In this framework, different approaches are investigated towards automatic segmentation of (multimedia) sound tracks, including, among others, changes in acoustic environments, speaker change detection, speaker identification and tracking, and speech/music discrimination. This segmentation is also useful, e.g., towards automatic adaptation of the models, as well as for resetting time points for language models and topic extraction systems.
8. Pronunciation variants modeling: Automatic extraction and modeling of pronunciation variants based on various factors such as word context and speaking style (e.g., conversational speech, speaking rate).
9. Statistical language modelling: Extending current language models to better cope with natural speech, out-of-vocabulary word, and word classes.
10. Speaker adaptation: Improving recognition accuracy by automatically adapting (a subset of) the parameters of the recognition system.
11. Development and adaptation of efficient software for large vocabulary continuous speech recognition, on different computer platforms (mainly UNIX and Windows NT).
12. Development and testing of applications prototypes.

5.1.2 Application Examples

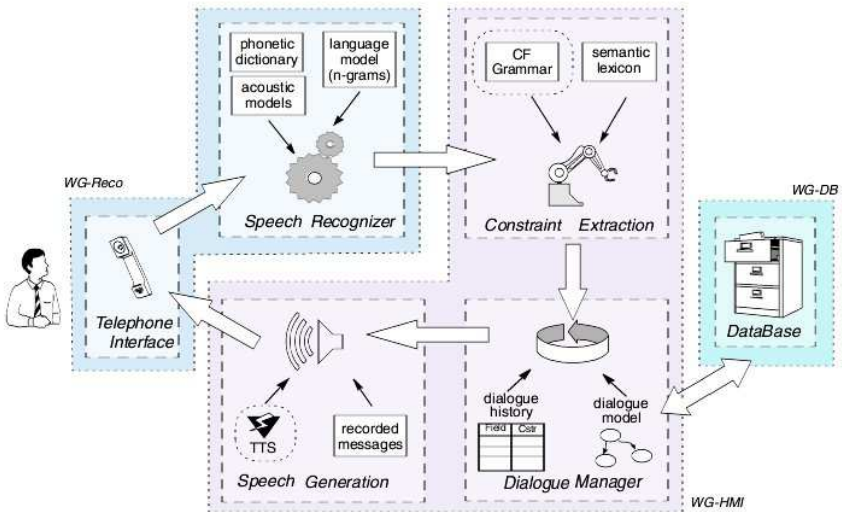


Figure 3: Block-diagram of the main modules involved in the InfoVox project, targeting an interactive voice server allowing users to get access to restaurants in Martigny through natural dialog.

1. Command and control systems, possibly used in noisy environments, e.g., to operate a speech enabled cellular phone in cars. See, e.g., the RESPITE and SPHEAR projects.
2. Speech enabled information systems: Building speech-enabled kiosks, desk tablets, and personal data assistants to enable users to find and display current information. See, e.g., the InfoVox project, also illustrated in Figure 3.
3. Information retrieval for audio documents: Using transcriptions automatically generated by a large-vocabulary speech recogniser to build indexes that can be queried by information retrieval engines for searchable audio archives. See, e.g., the ASSAVID and CIMWOS projects, allowing for:

   Automatic transcription of broadcast speech by an automatic speech recognition system

   Automatic indexing of the generated audio archives

   Content-based retrieval from typed or spoken input queries.
4. Automatic meeting manager: processing of multiple audio (microphone) streams for structuring, browsing and querying of an archive of automatically analysed meetings. See, e.g., the M4 project.

5.2 Computer Vision Group

The computer vision group studies problems in machine visual perception, such as media annotation, people detection and human gesture tracking and recognition. Research activities centre on multimodal interpretation of visual and multimedia data, and improvement of basic detection and classification measures and algorithms. This improvement may be achieved by enhancing and extending existing algorithms, or by creating new algorithms and measures. This frequently involves collaboration across research groups, as complementary expertise is brought to bear on a problem.

There is strong expertise within the vision group in areas of text processing from both documents and video, object tracking and recognition of gesture, and domain based video annotation. The group is active in all of these areas under a number of collaborative European and Swiss national projects.

5.2.1 Research Themes

1. Document Analysis and Recognition: Work in this area involves applying non-traditional methods to improve both preprocessing and recognition steps. In particular, methods from the speech processing area are being examined for use to improve recognition.

2. Improved modeling: statistical methods (PCA, ICA), Neural Networks and HMMs are used to improve the modeling of the handwritten data. Adaptive word normalisation algorithms allow independence with respect to a change of data set.

3. Sentence recognition: combination of word and language modeling allows the transcription of handwritten sentences. Pruning and computational weight reduction techniques make possible a fast decoding of the data.

4. Image and video annotation: Multimedia annotation involves both visual and audio data, and the deduction of higher level, semantic annotation which must be conducted under a machine learning framework. The work in this area thus brings together all three groups at IDIAP.

5. Accretive annotation for video: Accretive annotation uses low level feature information from a number of modalities, such as audio and video, to work towards semantic annotation. For this work domain specific information is used to provide a framework for interpreting the low level data, to direct progressively higher level processing for increasingly detailed annotation.

6. Text Detection and Recognition in Images and Videos: The vision group is involved in text detection and segmentation algorithms, and also examination of new paradigms in video text recognition. The goal of current research is to reduce false positive detection, and move away from explicit segmentation as a preprocessing step.

7. Face Algorithms: Face algorithm can be divided into four different areas.

   Face detection : The goal of face detection is to identify and locate human faces in images at different positions, scales, orientations and lighting conditions.

   Face localisation : Face localisation is a simplified face detection problem with the assumption that the image contain only one face.

   Face verification : Face verification is concerned with validating a claimed identity based on the image of its face, and either accepting or rejecting the identity claim.

   Face recognition : The goal of face recognition is to identify a person based on the image of its face. This face image has to be compared with all the registered persons. Therefore, face recognition is computationally expensive regarding the number of registered persons.

   During the last year, the problem of person verification has been addressed by researchers at IDIAP in the framework of security applications. Identity verification is a general task that has many real-life applications such as access control, transaction authentication (in telephone banking or remote credit card purchases for instance), voice mail, or secure teleworking. New methods proposed improves significantly the performance and achieves state-of-the-art results.

   Face detection is the fundamental step before the verification procedure. Its reliability and time-response have a major influence on the performance and usability of the whole face verification system. Therefore, we continue to investigate the problem of robust face detection in the general case of object detection (see Figure 4).

8. Gesture Recognition: Gestural interfaces based on the image is the most natural way for the construction of advanced man-machine interfaces. Thus, machines would be easier to use by associating the gestural command with the vocal command. It is necessary to distinguish two aspects of hand gestures:

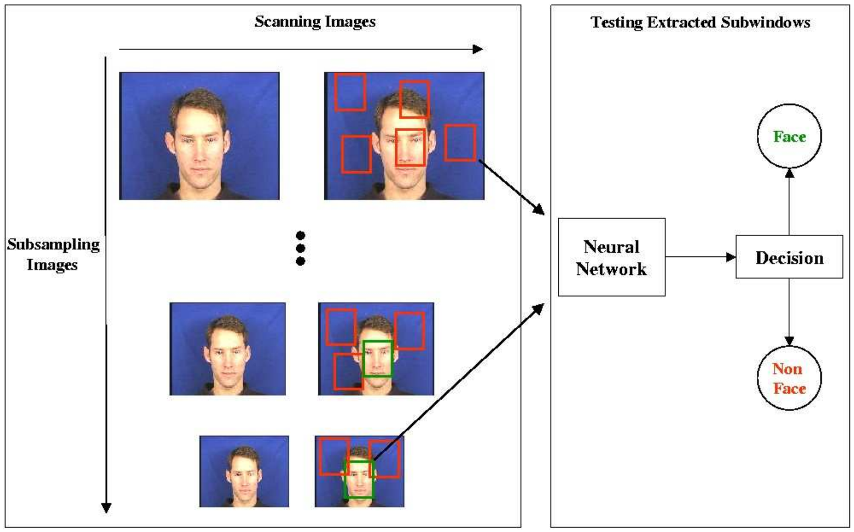


Figure 4: Face detection in an image using several subsampling stages.

the static aspect is, for instance, characterized by a posture of the hand in an image. Hand postures for example can be used to execute commands.

the dynamic aspect is defined either by the trajectory of the hand, or by the sequence of hand postures in a sequence of images.

Our work has as an ambition to develop techniques of statistical training for the recognition of hand gestures, and more generally for fixed images and animated images.

5.2.2 Application Examples

The purpose of image and video annotation is to provide access to the ever increasing digital archives of such data. Whether these archives are within a television station or publicly available web documents, the sheer volume of data being produced at any moment is beyond human ability to annotate. In addition there are large historical archives that contain priceless data recording important moments. Television stations will use such technology to provide a method of access to their archives, such as sports and news, and to access historical footage to enrich current programs, and for documentary pieces. Video and image text recognition is obviously a key part of this technology, as captions and in-vision text contain much useful information.

Hand drawn character and cursive writing recognition is useful for such tasks as automated address reading for postal services, and interface to such devices as PDAs. In addition, notes taken in meetings and during other discussions are predominantly handwritten. The ability to read such sources of information would be highly useful in many cases.

5.3 Machine Learning Group

The Machine Learning group at IDIAP is mainly interested in statistical machine learning, a research domain mostly related to statistical inference, artificial intelligence, and optimization. Its aim is to construct systems able to learn to solve tasks given a set of examples that were drawn from an unknown probability distribution, eventually given some prior knowledge of the task. Another important goal of statistical machine learning is to measure the expected performance of these systems on new examples drawn from the same probability distribution.

5.3.1 Research Themes

1. Large scale data analysis: most actual powerful machine learning algorithms have been used for medium scale datasets: less than one hundred features describing one example and less than ten thousand examples in the dataset. For instance, the now well-known Support Vector Machine algorithm needs resources that are quadratic in the number of examples, which forbid their use for problems with more than a few hundred thousands examples. Decomposition of the problem into sub-problems may lead to efficient solutions.

2. Ensemble models: One way to enhance generalization performance of machine learning algorithms is to combine the output of many algorithms instead of relying on only one algorithm. Many such methods are already known, such as AdaBoost, Bagging, Mixture of Experts.

3. Feature selection: Another way to enhance generalization performance of machine learning algorithms is to select and use only the input features that are well suited to solve a given problem.

4. Fusion of generative and discriminative models: two classes of machine learning algorithms are known and they have different advantages and disadvantages, depending on the problem to solve. We are interested in new algorithms that take advantages of both approaches.

5. Generalization performance analysis: As already stated, the goal of our group is not only to provide new and efficient machine learning algorithms but also to analyze and understand them in order to be able to compare them to other state-of-the-art algorithms.

6. Sequence modeling: most recent machine learning have been tailored for static problems. Given IDIAP's interest in speech processing, our group is also interested in developing and analyzing specific machine learning algorithm for sequence processing, including time series prediction and biological sequence analysis.

7. Spatial data analysis: We are specifically interested in building machine learning algorithms that would take into account spatial correlation between the input features and the target output in order to simultaneously enhance the prediction performance while preserving the spatial distribution of the dataset.

8. Multi-class classification: Many machine learning algorithms are in fact classification problems with multiple classes. One such problem in speech is the prediction of the phoneme (one out of 40 different phonemes) given the input features, at every time step.

9. Support to the Vision and Speech groups: the main role of the machine learning group is to support the research of the two other groups when machine learning is concerned.

5.3.2 Application Examples

The applications of Statistical Machine Learning are quite diverse. On top of all the applications related to speech and vision, which are best described by the two other groups, here is a sample of other interesting application domains:

Data Mining: how to extract interesting information from huge database wharehouses (for instance, churn detection, client modeling and prediction).

Finance and Economy: financial portfolio management, asset prediction, portfolio selection, auction analysis.

Pattern Recognition: handwritten character recognition, speech recognition, face detection.

Biological Sequence Analysis: classification of DNA or RNA sequences.

6 Current Projects

ADASEQ – AdaBoost and Other Ensemble Methods for Sequence Processing Problems

Funding: Swiss National Science Foundation

Duration: October 2001 - September 2002

Contact persons: Samy Bengio

Description: One of the main objectives of machine learning research is to develop algorithms that learn predictive relationships from data. This is a difficult task since inferring a function from data is in fact an "ill-posed" problem: many functions can often "fit" a given finite data set, but only some of them will behave adequately on new data drawn from the same distribution. Moreover it could happen that the function that fits best the given training data set will not behave as expected on new data. This is deeply related to the theory of statistical learning, which has been developed in the last years. Many approaches have been proposed recently to select the best function and to evaluate its expected performance on new data.

One approach to such problem is to select not only one function but many different functions and combine their outputs in order to produce a new solution. Nowadays, many machine learning algorithms are based on such technique, and are called ensemble methods. For instance, Bagging creates many functions, each of which being trained using a bootstrap of the data set (a new data set of the same size created by sampling independently from the original data set). The output of Bagging is then a simple average of the outputs of each function. This apparently simple method has been shown to significantly improve the performance on many tasks. More interestingly, AdaBoost also creates many functions, but each of them has been trained by putting more attention on the examples of the data set that produced the worst solutions using the previously trained function. A particular combination method is then applied which gives surprisingly good results over new data.

Most of these ensemble methods have been developed for classification (select a class among a fixed set of classes) or regression problems (predict a real-valued vector given another real-valued vector). On the other hand, some machine learning problems have their solution expressed as a sequence of output values. One such problem is the automatic speech recognition problem where the output is a sequence of words. This problem, as well as most of the sequence processing problems, is usually handled using hidden Markov models (HMMs), which are statistical models specifically designed for sequence processing problems and have given state of the art performance on many sequence problems. Unfortunately, there is currently not many ensemble algorithms specifically designed for HMMs or sequence problems.

The purpose of the ADASEQ project is thus to study, propose, develop and compare new ensemble methods tailored for sequence processing problems. As the current ensemble methods have usually bring good generalization performance on classification and regression problems, it is expected that it would also bring good performance for sequence processing problems. One of the main problems that will be addressed in the framework of this project will be the search for methods that efficiently combine sequences having a different size and a different confidence degree. Another research area will be to determine how the different models should be trained in order to give different yet complementary results. Finally, a theoretical analysis of these new ensemble techniques will also be done.

ARTIST – Articulatory Representation To Improve Speech Technologies

Funding: Swiss National Science Foundation

Duration: April 1999 - March 2001

Contact persons: Sacha Krstulovic (PhD, graduated in Dec. 2001), Hervé Bourlard

Description: Although speech and speaker recognition systems are now operational on small vocabularies and in noiseless conditions, their performance often degrades in real-life conditions. In the ARTIST project, we investigate the possibility to enhance current speech recognition systems by using articulatory features, or by using additional constraints inferred from those features. This can only be achieved through automatic acoustic-to-articulatory mapping, which is known to be a difficult (one-to-many) problem.

Among other models, the Distinctive Region Model (DRM) has been studied in detail, implemented and exploited, resulting also in a freely available software library. It has also been shown that integrating the DRM-derived constraints into the standard Linear Prediction Coding (LPC) modeling was bringing improvements to the modeling accuracy, with applications in speech synthesis, and to the speech recognition task.

ASSAVID – Automatic Segmentation and Semantic Annotation of Sports Videos

Funding: European project, 5th Framework Programme, Information Society Technology, supported by OFES

Duration: February 2000 – July 2002

Partners: Sony (UK), ACS (I), BBC (UK), University of Firenze (I), University of Surrey (UK)

Contact persons: Sebastien Marcel, Iain McCowan, Mark Barnard, Jitendra Ajmera, Datong Chen

Description: The most common method for accessing information today is still the textual query. Such technology is pervasive and well developed. Language is the dominant method we use to describe and communicate concepts, this is because we usually contend with semantics, that is the meaning of things. The explosion in availability of digital multimedia has led to a challenge in the way we describe and access information, in that much of the data presented is visual, either image or video, or multimodal. The challenges stem from the fact that it is extremely difficult to extract semantic information from such data, and that the commonly employed forms of access to this data are semantically shallow.

The main research issue in image and video that is confronted by the vision group at IDIAP is depth of annotation. Under the ASSAVID project an exploration is being continued into improvement of techniques for extraction of features from audio visual media, and the depth of annotation that may be achieved by fusing the multiple modes and features extracted. Part of the feature extraction and fusion work will be to examine new modalities of features which may be extracted and to determine their utility as a form of annotation. The fusion process will incorporate some domain knowledge to allow further deduction of semantic knowledge from the multimodal cues deduced from features. It is possible that new retrieval paradigms will suggest themselves in this process, due to the novel cues employed.

Recent work has produced improved methods for detection and segmentation of text from video. Importantly, the new detection method produces far fewer false detections than comparable systems. This not only reduces mis-recognitions, but also greatly reduces computation time spent on fruitless tasks. An improved segmentation algorithm based on a novel image feature, combined with the new detection algorithm, allows significantly higher recognition rates from OCR systems.

AudioSkim – Automatic Segmentation of Large Audio and Multimedia Documents

Funding: Swiss National Science Foundation

Duration: March 2001 - September 2002

Contact persons: Jitendra Ajmera, Iain McCowan, Hervé Bourlard

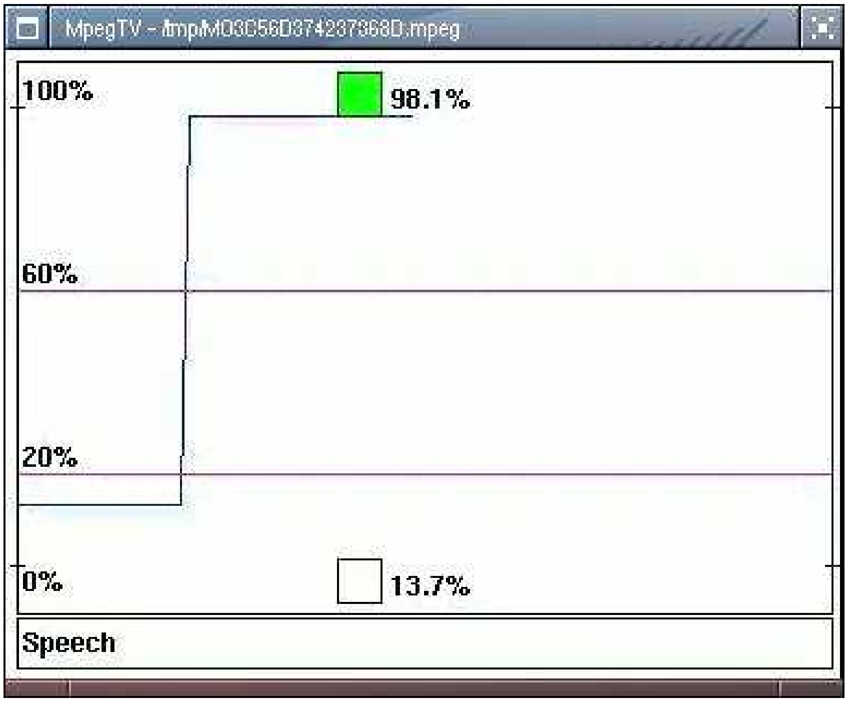


Figure 5: Automatic speech/non-speech segmentation: in realtime, this system can listen to an audio stream and display the “amount” of speech currently present at the signal.

Description: The problem of distinguishing speech signals from other audio signals (e.g., music) has become increasingly important as automatic speech recognition (ASR) systems are applied to more and more real-world multimedia domains. Furthermore, audio and speech segmentation will always be needed to break the continuous audio stream into manageable chunks applicable to the configuration of the ASR system. By using ASR acoustic models trained on particular acoustic conditions, such as wide bandwidth (high quality microphone input) versus telephone narrow bandwidth, male speaker versus female speaker, etc., overall performance can be significantly improved. Finally, this segmentation could also be designed to provide us with additional interesting information, such as the division into speaker turns and the speaker identities (allowing, e.g., for an automatic indexing and retrieval of all occurrences of a same speaker), as well as “syntactical information” (such as end of sentences, punctuation marks, etc). Although all those problems may sound simple, they are particularly challenging, and are just started to be systematically investigated.

In this project, we thus propose to investigate, implement, and test on large audio databases (possibly as part of multimedia databases) such as broadcast news and sport videos, different approaches towards automatic segmentation of (multimedia) sound tracks, including, among others, changes in acoustic environments, speaker change detection, speaker identification and tracking, and speech/music discrimination. After one year, and as illustrated in Figure 5, this project already resulted in an automatic system allowing for online segmentation of an audio signal into speech/non-speech segments and which is apparently outperforming other state-of-the-art approaches.

BANCA – Biometric Access Control for Networked and e-Commerce Applications

Funding: European project, 5th Framework Programme, Information Society Technology, supported by OFES

Duration: February 2000 – July 2002

Partners: IRISA (F), Banco Bilbao Vizcaya (E), EPFL (CH), Ibermatica S. A. (E), OSCARD S. A. (F), Thomson-CSF Communications (F), Université Catholique de Louvain (B), University of Surrey (UK)

Contact persons: Samy Bengio, Sebastien Marcel, Johnny Mariethoz

Description: The objectives of the project is to develop an implement a complete secured system with enhanced identification, authentication and access control schemes for applications over the Internet such as tele-working and Web-banking services. One of the major innovations of this project will be to obtain an enhanced security system by combining classical security protocols with robust multimodal verification schemes based on speech and image. The project includes the following objectives:

- development of scalable and robust multimodal verification algorithms
- development of scalable classifier combination techniques
- design and implementation of an overall secure architecture including security protocols adapted to biometrics
- development of three demonstrators: tele-working, home-banking, and ATM.

BN-ASR – Modeling the hidden dynamic structure of speech production in a unified framework for robust automatic speech recognition

Funding: Swiss National Science Foundation

Duration: March 1999 - September 2002

Contact persons: Todd Stephenson, Andrew Morris, Hervé Bourlard

Description: The main objective of this project is to develop new acoustic/phonetic models of speech for Automatic Speech Recognition (ASR). For years, Hidden Markov Models (HMM) have been the most successful technique in ASR. However, HMMs are rather general purpose stochastic models that only crudely reflect the nature of speech. This project will extend the hidden space of HMMs in various ways to better represent the hidden structure of speech production.

Bayesian Networks, relatively unknown in ASR, will serve as a framework for dynamic stochastic modeling. Thus the project will benefit from the past and current developments of the Bayesian networks theory. It is expected to contribute to this area as well.

This project will interact with other projects at IDIAP concerning the influence on speech production caused by prosody, speaker characteristics, and articulatory constraints. These information sources will be incorporated in the stochastic model in addition to the usual phonetic information.

CARTANN – Cartography by Artificial Neural Networks

Funding: Swiss National Science Foundation

Duration: January 1999 - January 2003

Partners: Lausanne University (prof. Michel Maignan)

Contact persons: Nicolas Gilardi, Mikhael Kanevski, Samy Bengio

Description: This work addresses a series of basic research items of spatial data analysis:

highly non stationary spatial processes,

cartography of distribution functions, as opposed to cartography of the mean value,

user and data-driven parameterization for the discrimination between a stochastic trend and auto-correlated residuals,

cartography of stochastic deviations related to advection-diffusion models.

Final solutions proposed for the resolution of geostatistical problems will mostly be hybrids involving ANNs and other learning methods (such as support vector machines and kernel ridge regression) to extract the general trends, together with classical approaches of geostatistics such as kriging estimations and simulations to estimate the residuals of the learning algorithm predictions if necessary.

CIMWOS – Combined IMages and WOrd Spotting

Funding: European project, 5th Framework Programme, Information Society Technology, supported by OFES

Duration: April 2001 - October 2003

Partners: Institute for Language and Speech Processing (ILSP, Greece), KULeuven (BE), ETHZ (CH), Sail-Labs (Austria), Canal+ (BE), and IDIAP

Contact persons: Iain McCowan, Jitendra Ajmera, Hervé Bourlard

Description: This project aims to facilitate common procedures of archiving and retrieval of audio-visual material. The objective of the project is to develop and integrate a robust unrestricted keyword spotting algorithm and an efficient image spotting algorithm specially designed for digital audio-visual content, leading to the implementation and demonstration of a practical system for efficient retrieval in multimedia databases. Specifically, a system will be developed to automatically retrieve images, video, and speech frames from an audio-visual database based on keywords entered by the user through keyboard or speech. Combined word and image spotting will be used and will provide an efficient mechanism enabling focused and precise searches with improved functionality and robustness. The CIMWOS system aims to become a valuable assistant in promoting the re-use of existing resources thus cutting down the budgets of new productions.

COST 275 – Biometrics-Based Recognition of People over the Internet

Funding: European project, 5th Framework Programme, COST, supported by OFES

Duration: June 2001 - May 2005

Countries involved: Belgium, Denmark, France, Ireland, Italy, Portugal, Spain, Slovenia, Sweden, Switzerland, Turkey, United Kingdom

Contact persons: Samy Bengio, Hervé Bourlard

Description: The main objective of the action is to investigate effective methods for the recognition of people over the Internet based on biometrics characteristics (principally voice and facial) in order to facilitate, protect, and promote various financial and other services over this growing telecommunication medium. In operational terms, the main objectives can be specified as follows:

1. To improve knowledge of the issues and problems involved.
2. To study the current techniques for voice and face recognition and to evaluate their performance in the medium considered.
3. To investigate methods for the fusion of the considered biometrics data and the interpretation of the results.
4. To analyze the implementation problems including user-interface issues and investigate effective solutions.
5. 5. To identify the potential applications and analyze the requirements of these.
6. 6. To develop standard methods and tools for the assessment of biometrics-based identification methods.

The secondary objectives are as follows:

1. To promote further research into (a) new and effective methods for voice and face recognition, and (b) novel techniques for data fusion.
2. To further research into multilingual interactive systems and their applications.
3. To standardize methods for the identification of individuals over the Internet.
4. To study the requirements and preferences of industry, and the attitude of the consumers.

As a partner of the COST 275 Action, IDIAP will be active in most of the research themes of the Action, with a particular emphasis on speaker recognition, face recognition, data fusion and assessment. However, thanks to the present project, these activities will take place in the framework of common efforts towards the research and development of a truly multi-modal (using voice and face characteristics) user authentication systems, with applications to internet transactions.

COST 278 – Spoken Language Interaction in Telecommunication

Funding: European project, 5th Framework Programme, COST, supported by OFES

Duration: June 2001 - May 2005

Countries involved: Belgium, Switzerland, Czech Republic, Germany, Spain, Finland, France, Greece, Hungary, Italy, The Netherlands, Norway, Portugal, Sweden, Slovenia, Slovakia, Turkey, United Kingdom

Contact persons: Hervé Bourlard, Sébastien Marcel

Description: The main objective of the proposed action is to "increase the knowledge of potentially useful applications and methodologies in deploying spoken language interaction in telecommunication. Emphasis is on achieving knowledge of speech and dialogue processing in multi-modal communication interfaces". Furthermore, the objective is to achieve knowledge of natural human-computer interaction through more cognitive, intuitive and robust interfaces, whether monolingual, multi-lingual or multi-modal. In operational terms, the main objectives can be specified as follows.

1. To improve the knowledge of the issues and problems involved in general in spoken language interaction in telecommunication.
2. To achieve knowledge of issues related to robustness and multi-linguality within spoken language processing.
3. To achieve knowledge of spoken language interaction in the context of multi-modal communication.

4. To achieve knowledge of human-computer dialogue theories, models and systems and associated tools for the establishment of such systems.
5. To achieve knowledge of and evaluate telecommunication applications that apply spoken language as one out of more input or output modalities.

As a partner of the COST 278 Action, IDIAP will mainly contribute to the Speech Input Processing and Multi-Modal Processing Working Groups. While these two research themes will address several open issues, the present project will also allow us to investigate these issues in the same general framework of robust multi-stream/multi-channel processing, as recently pioneered by IDIAP. The project will also allow further developments of related technologies in robust speech recognition and selected computer vision approaches such as the recognition of pointing gestures and face detection.

FNSNF Divide and Learn – Improved Learning for Large Classification Problems

Funding: Swiss National Science Foundation

Duration: October 2000 - September 2002

Partners: Swiss Federal Institute of Technology (EPFL)

Contact persons: Silvia Chiappa, Christos Dimitrakakis, Ronan Collobert, Samy Bengio

Description: The machine learning community has lately devoted considerable attention to the decomposition of large scale classification problems into a series of sub-problems and to the recombination of the learned models into a global model. Two major motivations underlie these approaches:

1. reducing the complexity of each single task, eventually by increasing the number of tasks,
2. improving the global accuracy by combining several classifiers.

These motivations are particularly relevant to the research themes covered by IDIAP (such as speech recognition and computer vision tasks), since the databases we are typically dealing with are of large size.

EDAM – Environmental data mining: Learning algorithms and statistical tools for monitoring and forecasting

Funding: European project, INTAS foundation

Duration: June 2000 – June 2002

Contact persons: Samy Bengio, Mikhail Kanevski

Description: To support the ongoing effort to develop indicators for environmentally sustainable development, there is a real need for research to enhance the development of technologies which contribute to the maintenance of environmental quality (water, air, soil). First step of a such research program consist in collecting and analysing data to provide useful tools for environmental monitoring and forecasting. Such tools would be also helpful for pollution prevention and compliance with environmental laws. Furthermore, if properly managed, they can be applied in environmental protection, for public information and lower operational costs in industry.

The main scientific objectives of the project are to develop a new methodology and tools inspired by artificial intelligence (AI), geostatistics and statistical learning theory to solve environmental problems. Specific scientific objectives to be reached for completion of the above are the following:

1. to develop environmental data mining methodology: structuring and development of framework,
2. to develop new statistical estimation algorithms for identification and prediction,
3. to develop and adapt statistical learning theory (Support Vector Machines) to spatio-temporal data,
4. to develop and adapt methods for detection, analysis, modelling and prediction of extreme and rare events in spatio-temporal environmental processes,
5. to develop tools for image and shape analysis of both descriptive input data and interpolated and simulated spatial and spatio-temporal data based on geostatistics, image analysis and mathematical morphology,
6. to develop new original technique for hazard estimation of natural disasters on the basis of recent achievements in statistics of extreme values and in the theory of heavy-tail distributions.

FaceX – Facial Expression Recognition through Temporal and Appearance Based Models

Funding: Swiss National Science Foundation

Duration: October 1998 – September 2002

Partners: Swiss Federal Institute of Technology, Zurich (ETHZ)

Contact person: Beat Fasel

Description: The goal of the project FacEx is to implement a robust, fully automatic facial expression analysis system. The results of this work are important in numerous domains: research and assessment of human emotion (psychiatry, neurology, experimental psychology), consumer-friendly human-computer interfaces, interactive video, and indexing and retrieval of image and video databases. The output of the project will also provide important but missing tools in related research areas such as face recognition, audio-visual speech recognition and animination of synthetic faces.

After having developed several baseline versions of algorithms allowing for facial expression classification, we recently started investigating convolutional neural networks applied for the task of both facial expression recognition and face identity recognition. They allowed us to obtain person-dependent facial expression analysis of previously seen faces and have the advantage of automatically extracting features relevant for a given task at hand, while not imposing complex object normalization procedures.

FGnet – Face and Gesture Recognition Working Group

Funding: European project, 5th Framework Programme, Information Society Technology, supported by OFES

Duration: 36 months, September 2001-August 2004

Partners: University of Manchester, Gerhard-Mercator-University Duisburg, Aalborg University, Institut National Polytechnique de Grenoble, Cyprus College

Contact person: Sebastien Marcel

Description: FGnet is a "Concerted Action and Thematic Network" on Face and Gesture Recognition. The use of shared resources and data sets to encourage the development of complex process and recognition systems has been very successful in the speech analysis and recognition field, and in the image analysis field in the specific cases where it has been applied. The aim of the project is thus to encourage the development of common databases, technological approaches, and evaluation standards in the area of face and gesture recognition, i.e.:

1. Providing focus and common grounds for researchers developing face and gesture recognition technology
2. Creating a set of foresight reports defining development roadmaps and future use scenarios for the technology in the medium (5-7 years) and long (10-20 years) term
3. Specifying, developing and supplying resources (e.g. image sets) supporting these scenarios. The resource generation activity will involve the specification of key data sets, evaluation protocols and reference architectures that will form the basis for technology development and sharing.
4. Encouraging the use of these resources to share and boost technology development.

HOARSE – Hearing Organisation and Recognition of Speech in Europe

Funding: European project, 5th Framework Programme, Training and Mobility of Researchers (TMR) programme, Research Network, supported by OFES

Duration: 48 months, September 2002-August 2006

Partners: Sheffield University (UK), Rurh- University Bochum (D), Daimler-Chrysler (D), Helsinki University (FIN), Keele University (UK), Patras University (G), IDIAP (CH)

Contact person: Hervé Bourlard

Description: As a follow-up of the SPHEAR TMR project (see below), the overall objectives of HOARSE are to gain a better understanding of speech production and hearing mechanisms and to use this understanding to explain the perceptual organisation of sound and improve speech technology. This project will thus involve several research themes, including:

1. Auditory Scene Analysis: Understanding how sound mixtures are perceptually organised into a coherent auditory scene, and how this organization can be used in speech recognition.
2. Dealing with Reverberant Conditions: Reverberant conditions are a big problem for speech recognition, and their processing in human hearing.
3. Speech Production Modelling: Understanding how speech is produced, how this relates to speech perception and cerebral speech processing, and how this knowledge can be integrated in state-of-the-art speech recognition systems.
4. Automatic Speech Recognition Methodologies: Generalisation of state-of-the-art automatic speech recognition algorithms to take advantage of the above. Specifically, we will focus on natural listening conditions, where the speech to be recognised is one of many sound sources (including noise and competing speech) which change unpredictably in space and time.

HMM2 – A New Framework for Robust and Adaptive Speech Recognition

Funding: Swiss National Science Foundation

Duration: October 2000 - September 2002

Contact persons: Ikbal Shajith, Hervé Bourlard

Description: The HMM2 project is directed towards extending the hidden Markov model (HMM) framework to simultaneously accommodate complex constraints in both the temporal and frequency domains. The generic idea of the approach investigated here, referred to as HMM2 for obvious reasons, is to associate with each (temporal) HMM-state a second, frequency based, HMM which will model the underlying probability density function. In other words, the multi-gaussians (or artificial neural network) typically used in standard HMMs will be replaced by a frequency-based HMM, responsible for estimating, through frequency-based latent variables, the "temporal" HMM emission probabilities and the correlation across the frequency bands.

Such an approach (for which standard multi-gaussians are a particular case) has many potential advantages, including: (1) in the case of multi-band speech recognition, dynamic definition and adaptation of the subbands, (2) automatic formant tracking, (3) nonlinear frequency warping, and (4) modeling of the correlation across frequency bands.

KTI CTI InfoVOX – Interactive Voice Servers for Advanced Computer Telephony Applications

Funding: Swiss Commission for Technology and Innovation (CTI)

Duration: April 1999 - April 2001

Partners: EPFL (DI/LIA), Swisscom, VOXAccess S.A., and Omedia S.A.

Contact persons: Frank Formaz, Hervé Bourlard

Description: The main objective of this project is to do further research and development in the field of interactive voice servers, with applications in the key area of call centers for computer telephony applications. This project also involves and supports VOXCom, an IDIAP spin-off company developing computer telephony applications.

More specifically, the generic goal of this project is to improve state-of-the-art automatic speech recognition and natural language processing technologies, and to integrate this technology in a specific speech enabled information system.

The targeted application mainly covers the development of Interactive Voice Response (IVR) systems (interactive vocal query systems) to access large and complex (possibly distributed) information databases. In the present project, and as a realtime testbed, we focused on the development of a natural voice interface to the restaurants in Martigny. This system, now available, can handle natural voice queries about restaurants and, through a natural dialog interface, provide the users with information about laction, style, cost, etc, of the restaurants in Martigny. Recently, a first alpha version of the complete integrated system has been released and was tested with volunteers.

FNSNF INSPECT – INtegrating Speech (acoustic and linguistic) ConsTraints for enhanced recognition systems

Funding: Swiss National Science Foundation

Duration: January 1999 - December 2000

Partners: EPFL (Dr Martin Rajman)

Contact persons: Martin Rajman (EPFL/DI/LIA), Hervé Bourlard, Alex Trutnev

Description: The main goal of the present project is to develop and assess new strategies for integrating state-of-the-art acoustic models and advanced language models (LM) into speech understanding systems, in view of improving dialog-based interactive voice response (IVR) systems.

Interfaces between continuous speech recognition systems and advanced language models are typically based on the rescoring of N-best hypotheses obtained from a maximum likelihood criterion. Unfortunately, the resulting hypotheses do not necessarily contain much semantic variability, and are not well suited for a post-processing that includes the higher level knowledge sources typically used in speech understanding systems. Consequently, the general research theme of the current project is to investigate new ways of generating N-best hypotheses that include more "semantic" variability, becoming therefore more appropriate for linguistic post-processing.

This research is done in the framework of a dialogue-based information system for Advanced Vocal Information Services.

KERNEL – Kernel Methods for Sequence Processing

Funding: Swiss National Science Foundation

Duration: February 2001 - February 2003

Contact persons: Quan Le, Samy Bengio

Description: *Hidden Markov Models* (HMMs) are one of the most powerful statistical tools developed in the last twenty years to model sequences of data such as time series, speech signals or biological sequences. One of their distinctive features lies on the fact that they can handle sequences of varying sizes, through the use of an internal state variable.

Unfortunately, it is well known that for classification problems, a better solution should in theory be to use a *discriminant* framework. In that case, instead of constructing a model independently for each class, one constructs a unique model that decides where the frontiers between classes are.

A series of recent papers have suggested some possible techniques that could be used to mix generative models such as HMMs (to handle the sequential aspects) and discriminant models such as Support Vector Machines.

The purpose of the present project is thus to study, experiment (on different kinds of sequential data), enhance, and adapt these new approaches of integrating discriminant models such as SVMs into generative models for sequence processing such as HMMs.

LAVA – Learning for Adaptable Visual Assistants

Funding: European project, 5th Framework Programme, Information Society Technology, supported by OFES

Duration: 36 months, March 2002-February 2005

Partners: Xerox Research Center Europe (UK and France), INRIA (F), University of London (UK), Lund University (S), Graz University of Technology (A), IDIAP (CH), Australian National University (AUS)

Contact person: Samy Bengio

Description: The overall objective of LAVA is to create fundamental enabling technologies for cognitive vision systems. The resulting widely transferable knowledge is to be thoroughly evaluated and widely disseminated. The new technologies that LAVA will provide will enable new tools for a wide range of applications including "ambient intelligence scenarios". The project includes the following objectives:

Robust and efficient categorisation and interpretation of large numbers of objects, scenes and events, in real settings

Automatic acquisition of knowledge of categories, for convenient construction or extension of applications.

M4 – MultiModal Meeting Manager

Funding: European project, 5th Framework Programme, Information Society Technology, supported by OFES

Duration: 36 months, March 2002-February 2005

Partners: Sheffield University (UK), München University (D), TNO/TPD (NL), University of Twente (I), EPFL/LTS (CH), UniGe (CH), IDIAP (CH), ICSI (Berkeley, CA).

Contact person: Hervé Bourlard, Jean-Marc Odobez, Daniel Gatica-Perez

Description: The overall aim of M4 is the construction of a demonstration system to enable structuring, browsing and querying of an archive of automatically analysed meetings. The archived meetings will have taken place in a room equipped with multimodal sensors. For each meeting, audio, video, textual, and (possibly) interaction information will be available. Audio information will come from close talking and distant microphones, as well as binaural recordings. Video information will come from multiple cameras. While the video and audio information will form several streams of data generated during the meeting, the textual information (the agenda, discussion papers, text of slides) will be pre-generated and will be used to guide the automatic structuring of the meeting. The interaction stream consists of any information that can help in analysing events within the meeting, for example, mouse tracking from a PC-based presentation or laser pointing information. The main research and development streams of M4 thus include::

1. Development of a "smart" meeting room, collection and annotation of a multimodal meetings database.
2. Automatic analysis and processing of the audio and video streams, including: robust conversational speech recognition, recognition of gestures and actions, multimodal identification of intent and emotion, multimodal person identification, source localization and tracking.
3. Integration and structuring using the output of the various recognizers and analyses, including: specification of a flexible intelligent information management framework, models for the integration of multimodal stream, summarization of a meeting, or a meeting segment, multimodal information extraction and cross-lingual retrieval/browsing across the archive.
4. Construction of a demonstrator system for browsing and accessing information from an archive of processed meetings.

MULTICHAN – A new approach to exploiting dependencies with hidden Markov models

Funding: Swiss National Science Foundation

Duration: January 2000 - December 2001

Contact persons: Katrin Weber, Hervé Bourlard

Description: All state-of-the-art speech recognition systems today are using hidden Markov models (HMM), which are well suited to deal with the temporal aspects of the speech signal, and which can now also be extended to deal with multiple data streams. However, these HMMs require the calculation of local "emission probabilities", which usually require strong assumptions regarding the distribution of the data and the correlation between the different components.

The main goal of this project is to develop new techniques towards speech recognition based on multiple data streams (e.g., representing different time scales), with multi-band speech recognition as a particular case. In this framework, one of the important open issues being investigated to properly model the (relevant) correlation between the different components (or streams) of the signal with a reasonable number of parameters. Although several solutions to this problem have already been proposed, we here investigate a drastically different approach, which seems to (1) be particularly promising and (2) fit well into the multi-channel speech recognition formalism.

PROMO – PROnunciation MOdelling in Automatic Speech Recognition Systems

Funding: Swiss National Science Foundation

Duration: August 2000 - July 2002

Contact persons: Mathew Magimai Doss, Hervé Bourlard

Description: Natural speech and casual human conversation exhibit a large amount of nonstandard variability in pronunciation. Phonological studies of the way a word is pronounced in different lexical contexts by native speakers of a language in clearly articulated speech lead to more than one acceptable pronunciation for many words. This results in a mismatch between the baseline phonetic transcriptions given in the lexicon and the actual pronunciation of the words, seriously hindering the recognition performance.

The mismatch between the dictionary representation of words and their actual realization may be reduced using an improved pronunciation model. In state-of-the-art speech recognition systems, this is often achieved simply by adding many pronunciation alternatives for each word, or by automatically inferring pronunciation variants from multiple utterances of each word.

The main motivation of this project is thus to investigate new techniques towards robust modelling of pronunciation variants in the context of continuous speech recognition, and more particularly in the case of natural speech recognition. On top of further investigating standard approaches (such as the automatic generation of pronunciation variants based on a maximum likelihood criterion), this project will focus on (1) dynamic pronunciation modelling, and (2) discriminant training of pronunciation models.

RESPITE – REcognition of Speech by Partial Information TEchniques

Funding: European project, 4th Framework Programme, Long Term Research (now Information Society Technology), supported by OFES

Duration: January 1999 - September b2002

Partners: Sheffield University (UK), Daimler Chrysler (D), BaBel (B), FPMs (Polytechnic University of Mons) (B), University of Grenoble (F), ICSI (USA)

Contact persons: Hervé Bourlard, Andrew Morris

Description: This project aims at developing techniques for automatic speech recognition that are truly robust to unanticipated noise and corruption. These techniques are based on a combination of emergent theories of decision-making from multiple, incomplete evidence sources and of human speech perception. More specifically, new recognition paradigms based on multi-stream processing and the missing data theory are currently investigated here.

The resulting algorithms are being tested and deployed in two application areas, i.e., cellular phones related applications and recognition in cars. The expected results of this project are: (1) The extension of the range of conditions under which ASR can be used, and specifically the extension to cellular phones related applications and recognition in cars, and (2) advances in adjacent recent fields, such as the handling of multiple temporal resolutions and the processing of multi-modal information (e.g., audio-visual fusion).

SCRIPT – Cursive Handwriting Recognition

Funding: Swiss National Science Foundation

Duration: October 1999 - September 2002

Contact person: Alessandro Vinciarelli

Description: The recognition of cursive handwritten words when only the image of the data is available is called Off-Line Cursive Script Recognition (CSR). The great variability of handwriting styles and the fact that the letters are connected are the major difficulties of the problem.
A system for single word recognition was developed. It presents an original normalisation method (based on statistics) that improved significantly the performance with respect to traditional normalisation methods.
We are extending now the recognition problem to the automatic reading of sentences.

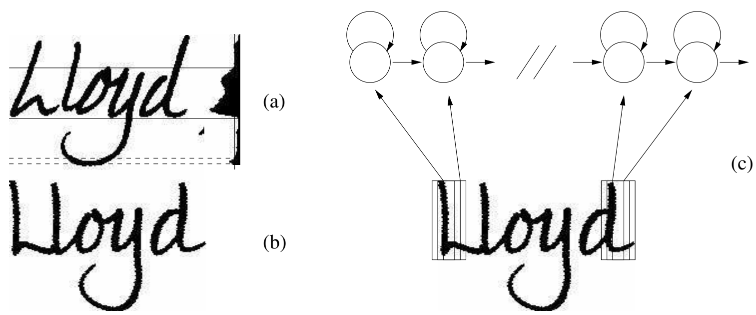


Figure 6: Single word recognition. The original image (a) is normalized (b) and modeled with HMMs (c). A HMM is created for every word in a list of possible interpretations of the data. The most likely model is assumed as transcription of the data.

Language modeling Unlike the case of single word recognition, it is possible to apply language modeling techniques to improve the performance. The n-gram models, the current state of the art, will be extensively applied in order to verify their effectiveness in the handwriting problem. Furthermore, language models only partially successful in speech domains (i.e. stochastic grammars), can be probably more helpful when applied to the written communication that is, in general, more formal than the oral one.

Search Technique The recognition of the handwritten data consists in measuring the matching between the observations (the vectors extracted from the data images) and the sentence models (HMM concatenations). This is done by finding the optimal path (in terms of some specified criterion) in a properly structured search space. This must involve both local (single letter level) and global (language model level) constraints. Besides, pruning techniques must be studied and applied in order to limit as much as possible the number of hypotheses considered (without reducing the overall recognition performance).

Hidden Markov Modeling Several parameters require to be set in Hidden Markov Models: number of states, topology, number of Gaussians in mixtures. Accurate experiments will be performed in order to find their optimal values. Moreover, an approach successfully applied in speech recognition will be applied, the hybrid HMM/ANN architecture.

SOCRATES – European Masters in Language and Speech

Funding: European project, 4th Framework Programme, Socrates/Erasmus, supported by OFES

Duration: September 1997 - September 2001

Partners: Univ. of Saarlandes (D), Aalborg Univ (DK), Univ. of Sheffield (UK), Univ. of Essex (UK), Univ. of Edimburgh (UK), Univ. of Brighton (UK), Univ. of Athens (GR), Univ. of Patras (GR), Univ. of Nijmegen (NL), Univ. of Utrecht (NL), Univ. of Lisbon (P), IDIAP-IKB (CH), EPFL (CH)

Contact persons: Hervé Bourlard

Description: The purpose of this project is to organize an advanced course (recognized as a European Masters degree) allowing students to qualify for multidisciplinary team-working in the language industries. Besides in depth knowledge of Speech Science, Natural Language Processing or Computer Science, provided by undergraduate studies, the student will obtain contextual knowledge from the fields that were not part of his/her specialization. At the European level, this would cover well-defined common courses, based on a common curriculum, and taught in every participating country. See http://www.cstr.ed.ac.uk/EuroMaster for more information.

IDIAP has the objective to create a center of excellence in the field of speech and language processing for graduated students. This center is expected to become part of a large European teaching network. At the Swiss level, this SOCRATES program thus resulted in the implementation of an EPFL Postgraduate Cycle in Speech and Language Engineering. This Postgraduate Cycle thus covers the multi-disciplinary curriculum of the resulting European Masters and include: theoretical linguistics, phonetics and phonology, cognitive models for speech and language processing, natural language processing, speech signal processing, statistical pattern recognition, and language engineering applications. In addition, the student is expected to spend a few (typically three) months on a project work, if possible abroad and/or as part of a traineeship in industry (through the contacts provided through the European consortium).

SPHEAR – SPeech, HEAring and Recognition

Funding: European project, 4th Framework Programme, Research Network, supported by OFES

Duration: March 1998 - September 2002

Partners: IDIAP, Ruhr-Universitat Bochum (Germany), Mercedes Benz (Ulm, Germany), Institut National Polytechnique de Grenoble (F), University of Keele (UK), University of Patras (GR), University of Sheffield (Sheffield, UK).

Contact persons: Astrid Hagen, Hervé Bourlard, Andrew Morris

Description: The twin goals of this research network are to achieve better understanding of auditory processing and to deploy this understanding in automatic speech recognition in adverse conditions. This project has several themes, including computational scene analysis, sound-source segregation and new recognition techniques based on multi-band and multi-stream processing.

In this project, IDIAP is mainly involved in multistream recognition techniques, where the objective is to extend current recognition paradigms, which are based on a single data stream, to multiple data streams which function in a natural auditory scene. The effectiveness of these techniques are being assessed for cellular phones and in-car applications, in collaboration with Daimler-Chrysler.

SV-UCP – Speaker Verification based on User-Customized Password

Funding: Swiss National Science Foundation

Duration: January 1999 - September 2002

Contact persons: Mohamed Benzeghiba, Hervé Bourlard

Description: The general objective of the present project is to further improve state-of-the-art speaker verification systems, where IDIAP has a recognized leading position. More specifically, the aim of this project is to investigate new alternatives to speaker verification systems, based on user-customized password (allowing the user to choose his/her password, just by pronouncing it a few times).

In the context of this project, automatic HMM inference approaches and fast speaker adaptation techniques will be investigated. This research is carried out in the framework of standard HMM, as well as in the context of hybrid HMM/ANN systems. Particular attention is however paid to the use of HMM/ANN systems since ANN have been shown to yield significantly better phonetic classification performance, which should potentially benefit to the precision of the automatically inferred HMMs (from a few pronunications of the password). On the basis of that inferred HMM, different speaker adaptation techniques are also being studied, and the resulting speaker verification performance is assessed on the Polyvar reference database.

VOCR - Text Recognition for Video Retrieval

Funding: Swiss National Science Foundation

Duration: December 1999 - September 2002

Contact persons: Datong Chen, Daniel Gatica-Perez, Jean-Marc Odobez

Description: The objective of this project is the investigation and development of algorithms for the detection, segmentation, and recognition of text in images and videos to be used for indexing and retrieval. Different image properties will be investigated including colour, texture, geometry, and character shape. In addition, the analysis of videos will exploit temporal characteristics of both the scene and the text. An important topic of the project will deal with the combination of evidence acquired my the different modules to perform detection and segmentation. Whereas previous research has treated detection, segmentation, and recognition as three separate problems, that often lead to individual errors, this work will investigate integration methods for all three processes to draw a joint decision driven by the result of the text recognition module.

A text detection algorithm has been developed which shows a false detection rate much superior to other current methods. In addition an algorithm has been developed to enhance edges in images before segmentation. This allows a much cleaner image to be used for OCR, giving improved results for text recognition.

7 Educational Activities

7.1 Current PhD Theses

The list of current IDIAP PhD students, together with their PhD projects and funding sources, is summarized in the table on next page. For a brief description of their research projects, we refer to Section 6.

7.2 PhD Defenses

Ph.D. candidate: Miguel Moreira

Supervisor: Prof. A Hertz (EPFL)

Examiners: Dr. S Bengio (IDIAP), Prof. G Coray (EPFL), Dr. E Mayoraz (Motorola), Prof. R C Dalang (EPFL).

University: EPFL, Lausanne

Title: The Use of Boolean Concepts in General Classification Contexts

Ph.D. candidate: Astrid Hagen

Supervisor: Prof. H Bourlard

Examiners: Prof. P Green (Sheffield University, UK), Dr. A Morris (IDIAP), Prof. G Rigoll (Duisburg University, D), Prof. P Vandergheynst (EPFL).

University: EPFL, Lausanne

Title: Robust Speech Recognition Based on Multi-stream Processing

Ph.D. candidate: Sacha Krstulovic

Supervisor: Prof. M Hasler (EPFL)

Examiners: Prof. H. Bourlard (IDIAP/EPFL), Dr. F. Bimbot (IRISA, F), Prof. G. Kubin (TU Graz, Austria)

University: EPFL, Lausanne

Title: Speech Analysis with Production Constraints

7.3 Participation in PhD Thesis Committees

Ph.D. candidate: Johan de Veth

Committee member: Hervé Bourlard

University: Nijmegen University, NL

Date: April 19

Title: On speech sound model accuracy

Ph.D. candidate: Hervé Glotin

Committee member: Hervé Bourlard

University: ICP-INPG, Grenoble, France

Date: June 13

Title: Elaboration et étude comparative de systèmes adaptatifs multi-flux de reconnaissance robuste de la parole

	PhD Students	Funding	Project	Expected PhD	At IDIAP since	PhD Status	Thesis Supervisor	Thesis Director
				SNSF FUNDING				
1	Jitendra AJMERA	FN 2100-65067.01	Audioskim	2004	01.01.01	2nd year, Speech	H. Bourlard	Prof. H. Bourlard, EPFL
2	Mohamed BENZEGHIBA	FN 2000-63721.00	SV-UCP	2004	01.08.00	3rd year, Speech	H. Bourlard	Prof. H. Bourlard, EPFL
3	Datong CHEN	FN 2000-65051.01	VOCR	2003	01.11.99	3rd year, Vision	J.M. Odobez	Dr. J.-P. Thiran, EPFL
4	Silvia CHIAPPA	FN 2100-61243.00	Divide & Learn II	2005	01.11.01	1st year, Learning	S. Bengio	Not decided yet
5	Christos DIMITRIKAKIS	FN 2100-61243.00	Divide & Learn II	2005	01.10.01	1st year, Learning	S. Bengio	Not decided yet
6	Beat FASEL	FN 2051-61877.00	FACEX	2002	01.10.98	4th year, Vision	D. Gatica-Perez	Prof. Van Gool, ETHZ
7	Nicolas GILARDI	FN 2100-54115.98	CARTANN	2002	01.01.99	4th year, Learning	S. Bengio	Prof. Maignan, UNIL
8	Shajith IKBAL	FN 2100-61325.00	HMM2	2004	01.05.00	2nd year, Speech	H. Bourlard	Prof. H. Bourlard, EPFL
9	Sacha KRSTULOVIC	FN 2000-55634.98	ARTIST 2	2001	01.04.96	Accepted, Speech	H. Bourlard	Prof. M. Hasler, EPFL
10	Quan LE	FN 2100-61245.00	KERNEL	2004	01.02.01	2nd year, Learning	S. Bengio	Not decided yet
11	Mathew MAGIMAI DOSS	FN 2100-57245.99	PROMO	2004	25.10.99	3rd year, Speech	H. Bourlard	Prof. H. Bourlard, EPFL
12	Todd STEPHENSON	FN 2000-64172.00	BNS 4 ASR	2003	01.03.99	4th year, Speech	A. Morris	Prof. H. Bourlard, EPFL
13	Alex TRUTNEV	FN 2100-54100.98	INSPECT	2004	01.08.00	2nd year, Speech	M. Rajman	Prof. H. Bourlard, EPFL
14	Vivek TYAGI	FN 2000-65077.01	SCRIPT	2005	01.06.01	2nd year, Speech	H. Bourlard	Prof. H. Bourlard, EPFL
15	Alessandro VINCIARELLI	FN 2000-65077.01	SCRIPT	2003	01.10.99	3rd year, Vision	S. Bengio	Prof. H. Bunke, Univ. Bern
16	Katrin WEBER	FN 2000-59169.99	CORREL	2002	01.01.98	4th year, Speech	H. Bourlard	Prof. H. Bourlard, EPFL
				OFES FUNDING				
17	Mark BARNARD	OFES 99.0562	ASSAVID	2005	15.03.01	1st year, Speech	H. Bourlard	Prof. H. Bourlard, EPFL
18	Fabien CARDINAUX	OFES 99.0229-2	CIMWOS	2005	01.10.01	1st year, Vision	S. Bengio	Not decided yet
19	Astrid HAGEN	OFES 97.0288	SPHEAR	2001	01.11.97	Accepted, Speech	H. Bourlard	Prof. H. Bourlard, EPFL
20	Maja POPOVIC	OFES 99.0562	ASSAVID	resigned	01.03.00			
21	Hemant MISRA	OFES 98.0086	RESPITE	2005	24.07.01	1st year, Speech	H. Bourlard	Prof. H. Bourlard, EPFL

7.4 Courses

Title: Speech and Language Engineering

Lecturer and Director of the course: Prof. H Bourlard

School: EPFL, Postgraduate

Title: Decision, estimation and statistical pattern recognition: Application to speech recognition

Lecturer: Prof. H Bourlard

School: EPFL, DI/DSC predoctoral school

Title: Speech Processing

Lecturer: Prof. H Bourlard

School: EPFL, Undergraduate (2nd cycle)

7.5 Other student projects

Trainee: Mario Zimmermann

Committee member: Frank Formaz, Johnny Mariethoz

University/School: HEVs

Date: March 2001 - July 2001 and October 2001 - January 2002

Title: Linux Based Voice Dialer

8 Scientific Activities

8.1 Editorship

Prof. Hervé Bourlard is

Editor-in-Chief of Speech Communication

Action Editor of Neural Network

Member of the Editorial Board of Futur(e)

8.2 Scientific and Technical Committees

Prof. Hervé Bourlard is

Member of the Board of Trustees, Intl. Computer Science Institute, Berkeley, CA, USA.

Member of the Advisory Board of the European Speech Technology Network

Member of the Board of Trustees of the Swiss Network for Innovation

Member of the Advisory Council of ISCA (International Speech Communication Association)

Member of the IEEE Technical Committee on Neural Network Signal Processing

Member of the Program Committee, ISCA workshop of the Adaptation Methods for Speech Recognition, Sophia-Antipolis, F, August 2001.

Member of the Scientific Committee, Odyssey Speaker Verification Workshop 2001

Co-General Chairman, IEEE MultiMedia Signal Processing (MMSP) workshop, Nice, 2001

European Liaison, and member of the Scientific Committee, IEEE Neural Network for Signal Processing workshop, 2001

General Chairman, IEEE Neural Network for Signal Processing workshop, Martigny, 2002

General Chairman, Eurospeech 2003, Geneva

Co-Technical Chairman, IEEE Intl. Conference of Acoustics, Speech, and Signal Processing (ICASSP), Orlando, May 2002

Member of the Scientific Committee, European Symposium of Artificial Neural Networks (ESANN), 2001 and 2002

Member of the Scientific Committee, International Association for Cybernetics

Member of the Administration Committee, European Association for Signal Processing (EURASIP)

Invited expert for the review of the BLISS European Project on Blind Source Separation, Granada, Spain (12.06.2001)

Invited panel moderator at ISCA Workshop on "Adaptation methods for speech recognition" (29.08.2001)

Session chairman on "Audio-visual speech recognition", Eurospeech'01, Aalborg, Denmark (04.09.2001)

Member of the international review panel of European "Training and Mobility of Researchers" programme, September 2001.

8.3 Short Term Visits

Location: AT&T Labs-Research, Florham Park, New Jersey, U.S.A.

Visitor: Todd Stephenson

Date: 03.01.2001 – 29.03.2001

Location: Visits, technical talks, and talks about IDIAP in India:

Visitor: H. Bourlard

Dates: Indian Institute of Technology (IIT), Bombay, 10.01.2001 — Tata Infotech Ltd, Bombay, 12.01.2001 — IIT Madras, 13.01.2001 — SSTIL Ltd, Madras, 15.01.2001 — IIT Delhi, 16.01.2001 — IIT Kanpur, 18.01.2001.

Location: ISPJAE, Facultad Eléctrica, Departamento Automática y Computación, Habana, Cuba

Visitor: A.C. Morris

Date: 01.06.2001

Location: Dept. of Medical Informatics, Institute of Biomedical Engineering, TU Graz, Austria

Visitor: A.C. Morris

Date: 18–25.06.2001

Location: IBM T.J. Watson Research Center, Yorktown Heights (NY), USA.

Visitor: Alessandro Vinciarelli

Date: 23.07.2001 – 26.10.2001

Location: Institute of Earthquake Prediction and Mathematical Geophysics, Moscow

Visitor: M. Kanevski

Date: 25–30.09.2001

Location: IRISA Rennes France

Visitor: Johnny Mariéthoz

Date: 25.11.2001 – 15.12.2001

Location: Université de Montréal, Montréal, Canada

Visitor: Samy Bengio

Date: 26–30.11.2001

Location: Sail Labs, Vienne, Austria

Visitor: Thierry Collado

Date: 13–14.12.2001

8.4 Scientific Presentations (other than conferences)

In this section, we briefly list the scientific events and external (e.g., invited) talks, other than conferences, and which did not necessarily result in a publication.

Event: Invited talk at Nijmegen University (NL), April 19

Speaker: H. Bourlard

Title: Non-stationary multi-stream processing

Event: Credit Suisse Perspectives 2002, Lausanne, September 26

Speaker: H. Bourlard

Title: Les technologies de l'information au 21ième siècle

Event: Invited speaker, Martigny, SSCCom conference, September 28

Speaker: H. Bourlard

Title: Intelligence artificielle: applications utiles pour le laboratoire d'analyses médicales

Event: Invited speaker, Rotary Club, October 29

Speaker: H. Bourlard

Title: Les technologies de l'information au 21ième siècle

Event: Invited lecturer at the "European School of Medical Physics", European Scientific Institute, Archamps, France, November 27

Speaker: H. Bourlard

Title: Statistical pattern recognition and multimodal interaction

Event: Seminar given at Ecole Polytechnique de Paris, April 23

Speaker: Samy Bengio

Title: Modélisation de données discrètes en haute dimension à l'aide de réseaux de neurones artificiels

Event: Workshop on Machines That Learn, Snowbird, Utah, April 2001

Speaker: Ronan Collobert and Samy Bengio

Title: A Parallel Mixture of SVMs for Very Large Scale Problems

Event: Containment & Remediation, International conference , Orlando

Speaker: M. Kanevski

Title: Spatial Data Analysis and Modelling of Radioactively-Contaminated Territories

Title: Characterization of Hydrogeologic Systems with Machine Learning Algorithms and Geostatistical Models

Event: Geneva Research Collaboration seminar, Geneva

Speaker: M. Kanevski

Title: MIRAGE: Exploratory data analysis of mortgage interest rate time series

Event: First Tuesday, November 16, Martigny, Switzerland

Speaker: S. Marcel

Title: Vision par Ordinateur: Detection et Reconnaissance des formes

Event: Forum de l'Economie Rhodanienne, Martigny, Switzerland

Speaker: S. Marcel

Title: Applications Multimodales de l'Intelligence Artificielle

9 Publications (2000 and 2001)

9.1 Books and Book Chapters

Note: **Every publication can also be accessed as an IDIAP Research Report from our web page at `http://www.idiap.ch/publicationsNF.html`.**

[1] F. BEAUFAYS, H. BOURLARD, H. FRANCO, AND N. MORGAN, *Neural networks in automatic speech recognition*, in to be published in The Handbook of Brain Theory and Neural Networks, M. A. Arbib, ed., Bradford Books, The MIT Press, 2000.

[2] R. BOITE, H. BOURLARD, T. DUTOIT, J. HANCQ, AND H. LEICH, *Traitement de la Parole*, Presses Polytechniques Universitaires Romandes, 2000.

[3] H. BOURLARD AND S. BENGIO, *Hidden markov models and other finite state automata for sequence processing*, to be published in The Handbook of Brain Theory and Neural Networks: The Second Edition, M. A. Arbib, ed., The MIT Press, 2002.

[4] H. BOURLARD, S. BENGIO, AND K. WEBER, *Towards robust and adaptive speech recognition models*, to be published in Mathematical Foundations of Speech Processing and Recognition, M. Ostendorf, S. Khudanpur, and R. Rosenfeld, eds., Institute for Mathematics and its Applications (IMA) Series, Springer-Verlag, 2002.

[5] N. MORGAN, H. BOURLARD, AND H. HERMANSKY, *Automatic speech recognition: an auditory perspective*, to be published in Speech Processing in the Auditory System, S. Greenberg, W. Ainsworth, A. Popper, and R. Fay, eds., Springer Verlag, New York, 2002.

9.2 Articles in International Journals

Note: **Every publication can also be accessed as an IDIAP Research Report from our web page at `http://www.idiap.ch/publicationsNF.html`.**

[1] S. BENGIO AND Y. BENGIO, *Taking on the curse of dimensionality in joint distributions using neural networks*, IEEE Transaction on Neural Networks special issue on data mining and knowledge discovery, (2000), pp. 550–557.

[2] F. CAMASTRA AND A. VINCIARELLI, *Cursive character recognition by learning vector quantization*, Pattern Recognition Letters, (2001).

[3] F. CAMASTRA AND A. VINCIARELLI, *Intrinsic dimension estimation of data: an approach based on Grassberger-Procaccia's algorithm*, Neural Processing Letters, 14 (2001).

[4] R. COLLOBERT AND S. BENGIO, *SVMTorch: Support vector machines for large-scale regression problems*, Journal of Machine Learning Research, 1 (2001), pp. 143–160.

[5] R. COLLOBERT, S. BENGIO, AND Y. BENGIO, *A parallel mixture of SVMs for very large scale problems*, Neural Computation, 14 (2002).

[6] S. DUPONT AND J. LUETTIN, *Audio-visual speech modelling for continuous speech recognition*, IEEE Transactions on Multimedia, (2000).

[7] N. GILARDI AND S. BENGIO, *Local machine learning models for spatial data analysis*, Journal of Geographic Information and Decision Analysis, 4 (2000), pp. 11–28.

[8] S. MOELLER AND H. BOURLARD, *Analytic assessment of telephone transmission impact on ASR performance using a simulation model*, accepted for publication in Speech Communication (2002).

[9] A. MORRIS, A. HAGEN, H. GLOTIN, AND H. BOURLARD, *Multi-stream adaptive evidence combination for noise robust ASR*, Speech Communication, 14 (2001), pp. 25-40.

[10] K. SHEARER, H. BUNKE, AND S. VENKATESH, *Video indexing and similarity retrieval by largest common subgraph detection using decision trees*, Pattern Recognition, 34 (2000).

[11] K. SHEARER, K. D. WONG, AND S. VENKATESH, *Combining multiple tracking algorithms for improved general performance*, Pattern Recognition, 34 (2000).

[12] A. VINCIARELLI, *A survey on off-line cursive word recognition*, Pattern Recognition, (2002).

[13] A. VINCIARELLI AND S. BENGIO, *Writer adaptation techniques in HMM based off-line cursive script recognition*, in Pattern Recognition Letters, 23/8 (2002), pp. 905-916.

[14] A. VINCIARELLI AND J. LUETTIN, *A new normalization technique for cursive handwritten words*, Pattern Recognition Letters, 22/9 (2001), pp. 1043-1050.

9.3 Articles in Conference Proceedings (refereed)

Note: **Every publication can also be accessed as an IDIAP Research Report from our web page at `http://www.idiap.ch/publicationsNF.html`.**

[1] J. AJMERA, I. MCCOWAN, AND H. BOURLARD, *Robust HMM-based speech/music segmentation*, accepted for publication in Proceedings of IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASS), Orlando, May 2002.

[2] S. BENGIO AND J. MARIÉTHOZ, *Learning the decision function for speaker verification*, in IEEE International Conference on Acoustic, Speech, and Signal Processing (ICASSP), 2001.

[3] F. BERTHOMMIER AND H. GLOTIN, *Reconnaissance de la parole dans le bruit après renforcement fondé sur l'harmonicité*, in Proceedings of JEP'2000, Aussois, 2000.

[4] F. BERTHOMMIER, H. GLOTIN, AND E. TESSIER, *A front-end using the harmonicity cue for speech enhancement in loud noise*, in Int. Conf. on Spoken Language Processing (ICSLP), Benjing, 2000.

[5] H. BOURLARD, S. BENGIO, AND K. WEBER, *New approaches towards robust and adaptive speech recognition*, in Advances in Neural Information Processing Systems 13, T. Leen, T. Dietterich, and V. Tresp, eds., MIT Press, 2001.

[6] D. CHEN AND J. LUETTIN, *Multiple hypotheses video OCR*, in Proceedings of the 4th International Workshop on Document Analysis System.

[7] S. CHOI, H. HONG, H. GLOTIN, AND F. BERTHOMMIER, *Multichannel signal separation for cocktail party speech recognition: a dynamic recurrent network*, in Int. Conf. on Spoken Language Processing (ICSLP), 2000.

[8] R. COLLOBERT, S. BENGIO, AND Y. BENGIO, *A parallel mixture of SVMs for very large scale problems*, in Advances in Neural Information Processing Systems, NIPS 14, T. Dietterich, S. Becker, and Z. Ghahramani, eds., MIT Press, 2002.

[9] D. CHEN, H. BOURLARD, *Text identification in complex background using SVM*, in Proceedings of the Int. Conf. on computer vision and pattern recognition, 2001.

[10] D. CHEN, K. SHEARER AND H. BOURLARD, *Text enhancement with asymmetric filter for video ocr*, in Proceedings of the 11th International Conference on Image Analysis and Processing, 2001.

[11] D. CHEN, K. SCHEARER AND H. BOURLARD, *Video OCR for sport video annotation and retrieval*, in Proceedings of the 8th IEEE International Conference on Mechatronics and Machine Vision in Practice, 2001.

[12] V. DEMYANOV, M. KANEVSKI, M. MAIGNAN, E. SAVELIEVA, V. TIMONIN, S. CHERNOV, AND G. PILLER, *Indoor radon risk assessment with geostatistics and artificial neural networks*, in Geostatistical congress, 2000.

[13] V. DEMYANOV, M. KANEVSKI, E. SAVELIEVA, V. TIMONIN, AND S. CHERNOV, *Neural network residual stochastic co-simulation for environmental data analysis*, in Neural Computation, 2000.

[14] B. FASEL AND J. LUETTIN, *Recognition of asymmetric facial action unit activities and intensities*, in Proceedings of International Conference on Pattern Recognition (ICPR 2000), Barcelona, Spain, 2000.

[15] C. FREDOUILLE, J. MARIÉTHOZ, C. JABOULET, J. HENNEBERT, C. MOKBEL, AND F. BIMBOT, *Behavior of a Bayesian adaptation method for incremental enrollment in speaker verification*, in Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Istanbul, Turkey, June 5–9 2000.

[16] N. GILARDI, M. KANEVSKI, M. MAIGNAN, AND E. MAYORAZ, *Environmental and pollution spatial data classification with support vector machines and geostatistics*, in Geostatistical congress, 2000.

[17] N. GILARDI, T. MELLUISH, AND M. MAIGNAN, *Confidence evaluation for risk prediction*, 2001 Annual Conference of the IAMG.

[18] H. GLOTIN AND F. BERTHOMMIER, *Test of several external posterior weighting functions for multiband full combination ASR*, in Int. Conf. on Spoken Language Processing (ICSLP), Beijing-China, Oct 2000.

[19] A. HAGEN AND H. BOURLARD, *Using multiple time scales in the framework of multi-stream speech recognition*, in Int. Conf. on Spoken Language Processing (ICSLP), 2000.

[20] A. HAGEN AND H. BOURLARD, *Error correcting posterior combination for robust multi-band speech recognition*, Proc. of Eurospeech'2001, Aalborg, pp. 591-594, 2001.

[21] A. HAGEN, H. BOURLARD, AND A. MORRIS, *Adaptive weighting in multi-band recombination of gaussian mixture ASR*, in Proc. of IEEE Intl. Conf. on Acoustics, Speech and signal Processing, (ICASSP), vol. 1, 2001.

[22] A. HAGEN AND H. GLOTIN, *Etudes comparatives des robustesses au bruit de l'approche "full combination" et de son approximation*, in Journée d'Etudes sur la Parole, Aussois, Aussois, France, Juin 2000.

[23] A. HAGEN AND A. MORRIS, *Comparison of HMM experts with MLP experts in the full combination multi-band approach to robust ASR*, in ICSLP, 2000.

[24] A. HAGEN, A. MORRIS, AND H. BOURLARD, *From multi-band full combination to multi-stream full combination processing in robust ASR*, in Proc. of ISCA ITRW ASR, 2000.

[25] H. HONG, S. CHOI, H. GLOTIN, AND F. BERTHOMMIER, *Blind acoustic source separation for cocktail party speech recognition*, in ICONIP, 7th IEEE Int. Conf. on Neural Information Processing, IEEE, ed., Korea, November 2000.

[26] S. KRSTULOVIĆ, *LPC modeling with speech production constraints*, in Proc. 5th Speech Production Seminar, 2000.

[27] S. KRSTULOVIĆ, *Relating LPC modeling to a factor-based articulatory model*, in Proc. ICSLP 2000, 2000.

[28] S. KRSTULOVIĆ AND F. BIMBOT, *Inverse lattice filtering of speech with adapted non-uniform delays*, in Proc. ICSLP 2000, 2000.

[29] S. KRSTULOVIĆ AND F. BIMBOT, *Signal modeling with Non Uniform Topology lattice filters*, in Proc. ICASSP 2001, 2001.

[30] M. KURIMO, *Fast latent semantic indexing of spoken documents by using self-organizing maps*, in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP'2000, Istanbul, Turkey, June 2000.

[31] M. KURIMO, *Indexing spoken audio by LSA and SOMs*, in Proceedings of the European Signal Processing Conference EUSIPCO'2000, Tampere, Finland, September 2000.

[32] J. MARIÉTHOZ AND F. BIMBOT, *Adaptation robuste de modèles HMM pour la vérification du locuteur dépendante du texte*, in Journee d'Etudes sur la Parole, Aussois, Aussois, France, Juin 2000.

[33] J. MARIÉTHOZ, J. LINDBERG, AND F. BIMBOT, *A MAP approach, with synchronous decoding and unit-based normalization for text-dependent speaker verification*, in ICSLP, 2000.

[34] A. MORRIS, *EEG pattern recognition through multi-stream evidence combination*, in Proc. World Congress on Neuroinformatics, Vienna University of Technology, Austria, September 24-29 2001.

[35] A. MORRIS, A. HAGEN, H. BOURLARD, *MAP combination of multi-stream HMM or HMM/ANN experts*, in Proc. Eurospeech, pp. 225-228, Aalborg, Denmark, September 3-7 2001.

[36] A. MORRIS, J. BARKER, AND H. BOURLARD, *From missing data to maybe useful data: soft data modelling for noise robust ASR*, in Proc. WISP, no. 06, Stratford-upon-Avon, England, April 2-3 2001.

[37] A. MORRIS, *Some applications of a priori knowledge in multi-stream HMM and HMM/ANN based ASR*, in Phonus No.5,Dec.2000, ISSN 0949-1791, Proc. Workshop on Phonetics and Phonology in ASR.

[38] A. MORRIS, *Data utility modelling for mismatch reduction*, in Proc. ISCA Workshop on Consistent & Reliable Acoustic Cues for sound analysis, Aalborg, Denmark, September 2 2001.

[39] A. MORRIS, L. JOSIFOVSKI, H. BOURLARD, M. COOKE, AND P. GREEN, *A neural network for classification with incomplete data: application to robust ASR*, in Proc. ICSLP.

[40] C. NETI, G. POTAMIANOS, J. LUETTIN, I. MATTHEWS, H. GLOTIN, D. VERGYRI, J. SISON, AND A. MASHARI, *Audio visual speech recognition*, Johns Hopkins University-CLSP, 2000.

[41] V. POLISHCHUK AND M. KANEVSKI, *Comparison of unsupervised and supervised training of RBF neural networks. Case study: Mapping of contamination data*, in Neural Computation, 2000.

[42] K. SHEARER, C. DORAI, AND S. VENKATESH, *Incorporating domain knowledge with video and voice data analysis in news broadcasts*, in Proceedings of the Sixth ACM International Conference on Knowledge Discovery and Data Mining, ACM, 2000.

[43] M.-C. SILAGHI AND H. BOURLARD, *Iterative posterior-based keyword spotting without filler models*, in Proceedings of the IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing, Istanbul, Turkey, 2000.

[44] T. A. STEPHENSON, H. BOURLARD, S. BENGIO, AND A. C. MORRIS, *Automatic speech recognition using dynamic Bayesian networks with both acoustic and articulatory variables*, in 6th International Conference on Spoken Language Processing: ICSLP 2000 (Interspeech 2000), Beijing, October 2000, pp. II:951–954.

[45] T. A. STEPHENSON, M. MATHEW, AND H. BOURLARD, *Modeling auxiliary information in Bayesian network based ASR*, in 7th European Conference on Speech Communication and Technology (Eurospeech 2001), vol. 4, Aalborg, Denmark, September 2001, pp. 2765–2768.

[46] A. VINCIARELLI AND J. LUETTIN, *Off-line cursive script recognition based on continuous density HMM*, in Proceedings of 7th International Workshop on Frontiers in Handwriting Recognition, 2000.

[47] K. WEBER, *Multiple timescale feature combination towards robust speech recognition*, in KONVENS 2000 / Sprachkommunikation, 2000.

[48] K. WEBER, S. BENGIO, AND H. BOURLARD, *HMM2- a novel approach to HMM emission probability estimation*, in Proc. of Intl. Conf. on Spoken Language Processing (ICSLP), 2000.

[49] K. WEBER, S. BENGIO, AND H. BOURLARD, *HMM2- extraction of formant features and their use for robust ASR*, in Proc. of Eurospeech, 2001.

[50] K. WEBER, S. BENGIO, AND H. BOURLARD, *Speech recognition using advanced HMM2 features*, in IEEE Automatic Speech Recognition and Understanding (ASRU) workshop, 2001.

[51] K. WEBER, S. BENGIO, AND H. BOURLARD, *Increasing speech recognition noise robustness with HMM2*, to be published in Proc. of IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP), 2002.

9.4 IDIAP Research Reports

NOTE: **Excluding all the above publications. All the reports below can be accessed from our web page at `http://www.idiap.ch/publicationsNF.html`.**

[1] J. AJMERA, I. MCCOWAN, AND H. BOURLARD, *Speech/music discrimination using entropy and dynamism features in a HMM classification framework*, submitted to *Speech Communication*, IDIAP-RR 01-26, IDIAP, Martigny, Switzerland, 2001.

[2] S. BENGIO, H. BOURLARD, AND K. WEBER, *An EM algorithm for HMMs with emission distributions represented by HMMs*, IDIAP-RR 01-11, IDIAP, 2000.

[3] S. BENGIO, C. MARCEL, S. MARCEL, AND J. MARIÉTHOZ, *Confidence measures for multimodal identity verification*, IDIAP-RR 01-38, IDIAP, 2001.

[4] S. BENGIO AND J. MARIÉTHOZ, *Comparison of client model adaptation schemes*, IDIAP-RR 01-25, IDIAP, 2001.

[5] S. BENGIO, J. MARIÉTHOZ, AND S. MARCEL, *Evaluation of biometric technology on XM2VTS*, IDIAP-RR 01-21, IDIAP, 2001.

[6] M. F. BENZEGHIBA AND H. BOURLARD, *User Customized HMM/GMM based Speaker Verification*, IDIAP-RR 01 32, IDIAP, 2001.

[7] M. F. BENZEGHIBA, H. BOURLARD, AND J. MARIÉTHOZ, *Speaker verification based on user-customized password*, IDIAP-RR 01 13, IDIAP, 2001.

[8] H. BOURLARD, *Auto-association by multilayer perceptrons and singular value decomposition*, IDIAP-RR 00-16, IDIAP, 2000.

[9] F. CAMASTRA AND A. VINCIARELLI, *Estimating the intrinsic dimension of data with a fractal-based method*, IDIAP-RR 02-02, IDIAP, 2002

[10] F. CAMASTRA AND A. VINCIARELLI, *Combining neural gas and learning vector quantization for cursive character recognition*, IDIAP-RR 01-18, IDIAP, 2001.

[11] D. CHEN, J.-M. ODOBEZ, AND H. BOURLARD, *Text recognition in complex background based on markov random field*, IDIAP-RR 01 47, IDIAP, 2001.

[12] D. CHEN AND K. SHEARER, *Asymmetric filter for text recognition in video*, IDIAP-RR 00-37, IDIAP, 2000.

[13] R. COLLOBERT AND S. BENGIO, *On the convergence of SVMtorch, an algorithm for large-scale regression problems*, IDIAP-RR 00-24, IDIAP, 2000.

[14] R. COLLOBERT AND S. BENGIO, *Support vector machines for large-scale regression problems*, IDIAP-RR 00-17, IDIAP, 2000.

[15] R. COLLOBERT, S. BENGIO, AND Y. BENGIO, *A parallel mixture of SVMs for very large scale problems*, IDIAP-RR 01-12, IDIAP, 2001.

[16] D. CHEN AND K. SHEARER, *A survey of text detection and recognition in images and videos*, IDIAP-RR 00-38, IDIAP, 2000.

[17] M. M. DOSS AND H. BOURLARD, *Pronunciation models and their evaluation using confidence measures*, IDIAP-RR 01-29, IDIAP, 2001.

[18] B. FASEL, *Robust Face Analysis using Convolutional Neural Networks*, IDIAP-RR 01-48, IDIAP, 2001.

[19] B. FASEL, *Use of Convolutional and Time-Delay Neural Networks for Facial Expression Analysis*, IDIAP-RR 01-49, IDIAP, 2001.

[20] C. FREDOUILLE, J. MARIÉTHOZ, C. JABOULET, J. HENNEBERT, C. MOKBEL, AND F. BIMBOT, *Behavior of a bayesian adaptation method for incremental enrollment in speaker verification*, IDIAP-RR 00-02, IDIAP, 2000.

[21] H. GLOTIN, *Robust multi-stream speech recognition based on the combined reliabilities of the speech signal and phonemes estimates*, IDIAP-RR 00-36, IDIAP, 2000.

[22] H. GLOTIN, D. VERGYRI, C. NETI, G. POTAMIANOS, AND J. LUETTIN, *Weighting schemes for audio-visual fusion in speech recognition*, IDIAP-RR 00-44, IDIAP, 2000.

[23] E. GRAND, *Handwritten digits recognition*, IDIAP-RR 00-07, IDIAP, 2000.

[24] A. HAGEN, *Robust speech recognition based on multi-stream processing*, IDIAP-RR 00-41, PhD Thesis, Ecole Polytechnique Fédérale de Lausanne, Switzerland, December 2001.

[25] M. KANEVSKI, *Evaluation of SVM binary classification with nonparametric stochastic simulations*, IDIAP-RR 01-07, 2001.

[26] M. KANEVSKI AND S. CANU, *Spatial data mapping with support vector regression*, IDIAP-RR 00-09, IDIAP, 2000.

[27] M. KURIMO, *Thematic indexing of spoken documents by using self-organizing maps*, IDIAP-RR 00-05, IDIAP, 2000.

[28] S. MARCEL, *Approches génératives pour le traitement de séquences d'images: application à la reconnaissance dynamique des gestes de la main*, IDIAP-RR 00-45, IDIAP, 2000.

[29] S. MARCEL AND S. BENGIO, *Improving face verification using skin color information*, IDIAP-RR 01-44, IDIAP, 2001.

[30] J. MARIÉTHOZ AND S. BENGIO, *A comparison of adaptation methods for speaker verification*, IDIAP-RR 01-34, IDIAP, 2001.

[31] S. C. M. KANEVSKI, P. WONG, *Environmental data mapping with support vector regression and geostatistics*, IDIAP-RR 00-10, IDIAP, 2000.

[32] P. MOERLAND, *Mixtures of latent variable models for density estimation and classification*, IDIAP-RR 00-25, IDIAP, 2000.

[33] M. MOREIRA, *The use of boolean concepts in general classification contexts*, IDIAP-RR 00-46, IDIAP, Martigny, Switzerland, December 2000.

[34] B. NEDIC AND H. BOURLARD, *Recent developments in speaker verification at IDIAP*, IDIAP-RR 00-26, IDIAP, 2000.

[35] M. POPOVIĆ, *Using posterior probabilities for speech/music discrimination*, IDIAP-RR 01-08, IDIAP, Martigny, Switzerland, 2001.

[36] S. S. IKBAL, H. BOURLARD AND K. WEBER, *IDIAP HMM/HMM2 system: Theoretical basis and software specifications*, IDIAP-RR 01-27, IDIAP, Martigny, Switzerland, 2001.

[37] K. SHEARER, C. DORAI, AND S. VENKATESH, *Detection of narrative structure for annotation of news broadcasts*, IDIAP-RR 01-03, IDIAP, 2001.

[38] K. SHEARER AND S. VENKATESH, *Artifacts of the colour coherence vector and an alternative similarity measure*, IDIAP-RR 01-02, IDIAP, 2001.

[39] K. SHEARER, S. VENKATESH, AND H. BUNKE, *Video sequence matching via decision tree path following*, IDIAP-RR 00-12, IDIAP, 2000.

[40] T. A. STEPHENSON, *An introduction to Bayesian network theory and usage*, IDIAP-RR 00-03, IDIAP, 2000.

[41] T. A. STEPHENSON, M. MAGIMAI DOSS, AND H. BOURLARD, *Automatic speech recognition using pitch information in dynamic Bayesian networks*, IDIAP-RR 00-41, IDIAP, 2000.

[42] T. A. STEPHENSON, M. MAGIMAI-DOSS, AND H. BOURLARD, *Mixed Bayesian networks with auxiliary variables for automatic speech recognition*, IDIAP-RR 01-45, IDIAP, 2001.

[43] A. VINCIARELLI AND S. BENGIO, *Offline cursive word recognition using continuous density hidden markov models trained with pca or ica features*, IDIAP-RR 01-46, IDIAP, 2001.

[44] A. VINCIARELLI AND S. BENGIO, *Writer adaptation techniques in HMM based off-line cursive script recognition*, IDIAP-RR 01-15, IDIAP, 2001.

[45] K. WEBER, S. BENGIO, AND H. BOURLARD, *A pragmatic view of the application of HMM2 for ASR*, IDIAP-RR 01-23, IDIAP, Martigny, Switzerland, 2001.

[46] , K. WEBER, S. IKBAL, S. BENGIO AND H. BOURLARD, *Robust Speech Recognition and Feature Extraction Using HMM2*, IDIAP-RR 01-42, IDIAP, Martigny, Switzerland, 2001.

9.5 IDIAP Communications

[1] F. BRESSOUD AND H. WANG, *Personal voice dialing over PC*, IDIAP-COM 00-05, IDIAP, 2000.

[2] TH. COLLADO, *Développement d'un système de demande interactif via le téléphone (INFOVOX)* IDIAP-COM 01-08, IDIAP, 2001.

[3] R. COLLOBERT, *Support vector machines, théorie et application*, IDIAP-COM 00-03, IDIAP, 2000.

[4] F. FORMAZ, M. GOYAL, AND O. BORNET, *Development of a DTW based Speech Recognition System over the telephone line*, IDIAP-COM 01-05, IDIAP, 2001.

[5] H. GLOTIN, *Various adaptive weighting schemes for large vocabulary robust audio-visual ASR, with particular reference to the cocktail party effect*, IDIAP-COM 00-04, IDIAP, 2000.

[6] IDIAP, *Activity report 1999*, IDIAP-COM 00-01, IDIAP, 2000.

[7] IDIAP, *Activity report 2000*, IDIAP-COM 01-01, IDIAP, 2000.

[8] S. KRSTULOVIĆ, *Epfl lab session 1/2: Introduction to Gaussian statistics and pattern recognition*, IDIAP-COM 01-06, IDIAP, 2001.

[9] S. KRSTULOVIĆ, *Epfl lab session 2/2: Introduction to hidden markov models*, IDIAP-COM 01-07, IDIAP, 2001.

[10] V. SIIVOLA, *Language modeling based on neural clustering of words*, IDIAP-COM 00-02, IDIAP, 2000.

[11] H. WANG, *Rebuilding speech recognition on windows*, IDIAP-COM 01-09, IDIAP, 2001.

[12] H. WANG, *Speech recognition engine for interactive voice response application on windows*, IDIAP-COM 01-10, IDIAP, 2001.

9.6 Other Documents

[1] A. HAGEN, *Robust speech recognition based on multi-stream processing*, PhD thesis, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, December 2001.

[2] S. KRSTULOVIĆ, *PhD Thesis: Speech Analysis with Production Constraints*, PhD thesis, École Polytechnique Fédérale de Lausanne, 2001.

[3] P. MOERLAND, *Mixture Models for Unsupervised and Supervised Learning*, PhD thesis, École Polytechnique Fédérale de Lausanne, Computer Science Department, Lausanne, Switzerland, June 2000.

[4] M. MOREIRA, *The use of Boolean concepts in general classification contexts*, PhD thesis, Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland, December 2000.

PROJET DE PÔLE NATIONAL DE RECHERCHE (PNR)

(NCCR : National Centre of Competence in Research)

Interactive Multimodal Information Management
(IM)2

Director: Prof. Hervé Bourlard, IDIAP (Martigny) and EPFL (Lausanne)

S O M M A I R E

1. Le contexte et historique du projet (IM)2 ____ 3
   1.1 "Des programmes nationaux de recherche" aux "Pôles de recherche nationaux" ____ 3
   1.5 8.12.1999 : 1ère évaluation scientifique internationale ____ 4
   1.6 15.3.2000 : Présentation d'une proposition complète ____ 4
   1.7 17.5.2000 : 2ième évaluation scientifique internationale ____ 4
   1.8 17.12.2000 (10:30) : Décision finale ____ 4
   1.8 13.6.2001 : Le Parlement accepte le financement de 4 nouveaux PRN ____ 5

2. Les partenaires du projet ____ 6
   2.1 Des universités et Hautes écoles ____ 6
   2.3 Des institutions publiques ____ 6

3. Un Bref descriptif du projet ____ 8
   3.1 Définition ____ 8
   3.2 Activités concernées dans les domaines de la recherche et de l'économie ____ 8
   3.3 Une technologie d'avenir ____ 8
   3.4 Thèmes de recherche : perspectives multiples d'applications concrètes ____ 9

4. Conclusion ____ 11

Annexe

A1. L'Institut Dalle Molle d'Intelligence Artificielle Perceptive (IDIAP) ____ 12
   A1.1 Historique ____ 12
   A1.2 Financement ____ 12
   A1.3 Activités ____ 12

1. Le contexte et historique du projet (IM)2

1.1 "Des programmes nationaux de recherche" aux "Pôles de recherche nationaux (PRN)"

En août 1998 (adopté par le Conseil Fédéral lors de sa séance du 25 novembre 1998), le Fonds National Suisse mettait en place le concept de «réseaux nationaux de recherche» en vue de prendre le relais des anciens «programmes prioritaires». Ces réseaux portent le nom de «Pôles de Recherche Nationaux (PRN)». Le but de ces projets est de cibler quelques domaines de recherche clés autour desquels plusieurs institutions de recherche collaboreraient et bénéficieraient d'un soutien financier important sur une durée maximale de 10 ans. Les domaines couverts concernent les sciences de la vie, les sciences humaines et sociales, les sciences et technologies de l'information (auquel appartient le projet (IM)2), l'environnement. Chacun de ces réseaux serait dirigé par une des institutions membres, le réseau lui-même étant reconnu comme «centre de compétence».

L'objectif premier de ces Pôles de Recherche Nationaux est de renforcer durablement la place scientifique suisse dans les domaines stratégiquement importants pour l'économie et la société. Un total d'environ 20 pôles PRN doit servir de base à réaliser ces objectifs, dont une première série de quelques huit pôles qui devaient démarrer en janvier 2001.

1.2 Janvier 1999 : Appel d'offre du Fonds National

En janvier 1999, une mise au concours et un appel à propositions était lancé en vue d'une première sélection, auquel les réseaux potentiellement intéressés sont invités à répondre par une déclaration d'intention.

1.3 31.3.1999 : Réponse de l'IDIAP à l'appel d'offre du Fonds National

Le 31.3.1999, 230 déclarations d'intentions ont été reçues, dont une (IM2) émanant de l'IDIAP (un institut de recherche semi-privé, situé à Martigny et affilié à l'EPFL et l'Université de Genève – voir brève description en annexe), ainsi qu'une autre du Prof. Murat Kunt du Département d'Electricité de l'EPFL (Protecting rights and privacy in the information society).

1.4 31.7.1999 : Remises d'avant-projets

Après une première sélection, la remise d'avant-projets assez détaillés a alors été requise pour le 31.7.1999. 82 avant-projets ont été envoyés au Fonds National dont celui de l'IDIAP qui, entre-temps, s'était également vu renforcé par l'intégration du projet initialement proposé par le Professeur Murat Kunt.

1.5 8.12.1999 : 1ère évaluation scientifique internationale

Le 8 décembre 1999, et sur la base d'une évaluation scientifique internationale rigoureuse, les 82 esquisses soumises ont alors été classées en trois catégories :

- Catégorie «chances de succès intactes» : 27 esquisses qui remplissaient entièrement les exigences scientifiques et pour lesquelles leurs auteurs ont été invités à soumettre une requête complète pour le 15 mars 2000.
- Catégorie «chances de succès incertaines» : 22 esquisses dans une zone intermédiaire.
- Catégorie «chances de succès faibles» : 33 esquisses ne remplissant pas les exigences scientifiques ou autres.

Le projet (IM)2 de l'IDIAP se trouvait classé en première catégorie (chances de succès intactes) et une requête complète a été préparée et envoyée pour le 15 mars 2000.

1.6 15.3.2000 : Présentation d'une proposition complète

Le 15 mars 2000, 34 propositions complètes ont été reçues par le Fonds National, pour une nouvelle évaluation scientifique.

1.7 17.5.2000 : 2ième évaluation scientifique internationale

Le 17 mai 2000, le Professeur Hervé Bourlard (Directeur de IM2, et Directeur de l'IDIAP) était invité à défendre le projet devant une commission scientifique (comprenant des représentants nationaux et internationaux).

Ceci a permis au Fonds National de sélectionner 18 projets qui ont alors envoyés, vers la mi-juillet, au Conseil Fédéral (Département de l'Intérieur) pour la choix final (sur base de critères de la politique scientifique) des projets financés.

Le projet (IM)2 fait partie de ces 18 projets.

1.8 17.12.2000 (10:30) : Décision finale

Madame la Conseillère Fédérale Ruth Dreyfus nous fait parvenir la liste officielle des projets retenus :

1. 10 projets PRN sont retenus pour financement
2. 4 requêtes (dont celle de l'IDIAP) sont retenues mais nécessitent des ressources supplémentaires. Ces ressources supplémentaires seront soumises pour approbation par le Parlement, au travers d'un message du Conseil Fédéral aux Chambres.
3. 4 requêtes sont rejetées.

1.8 13.6.2001 : Le Parlement accepte le financement de 4 nouveaux PRN

Le 16 juin 2001, l'IDIAP salue l'approbation par les Chambres d'un crédit de 35 millions de francs destiné à quatre nouveaux Pôles de recherche nationaux, dont celui proposé par l'IDIAP qui se réjouit d'être désigné comme leader du Pôle de recherche en gestion interactive et multimodale de l'information. L'IDIAP voit dans la décision du Parlement la reconnaissance de l'excellence de la recherche menée dans ses murs.

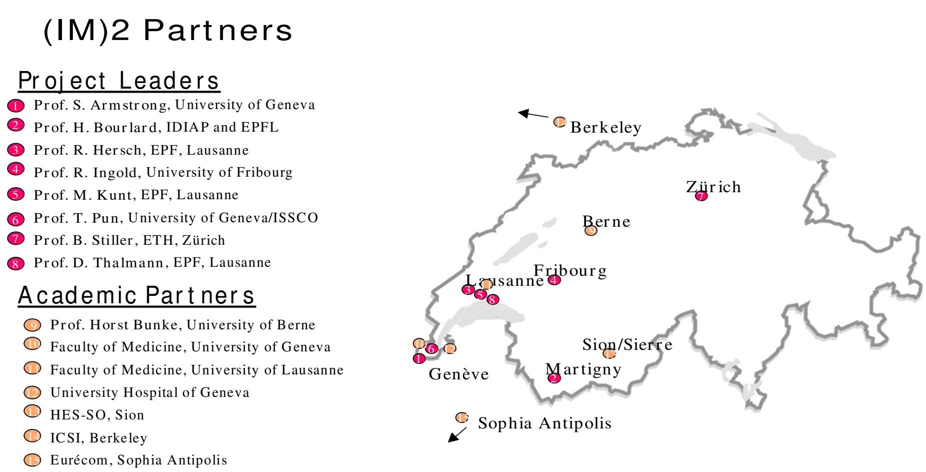
1.9 01.01.2002 : Démarrage officiel du Pôle de Recherche National IM2

Après adaptation des budgets, mise en place de l'administration interne, et signature des contrats avec le Fonds National, le PRN IM2 démarre officiellement le 1er janvier 2002, avec un financement garanti du Fonds National pour les quatre prochaines années de CHF 15'400'00, complété de fonds propres et de tiers s'élevant à CHF 16'220'000.

2. LES PARTENAIRES DU PROJET

2.1 Des universités et Hautes écoles

Comme illustré par la carte ci-dessous, (IM)2 représente une opportunité unique de collaboration entre de nombreuses institutions universitaires et de recherche suisses. Toutes ces institutions sont particulièrement actives dans les domaines relatifs aux interfaces multimodaux, mais ne bénéficient aujourd'hui d'aucune action concertée.



2.2 Des centres de recherche et des partenaires industriels

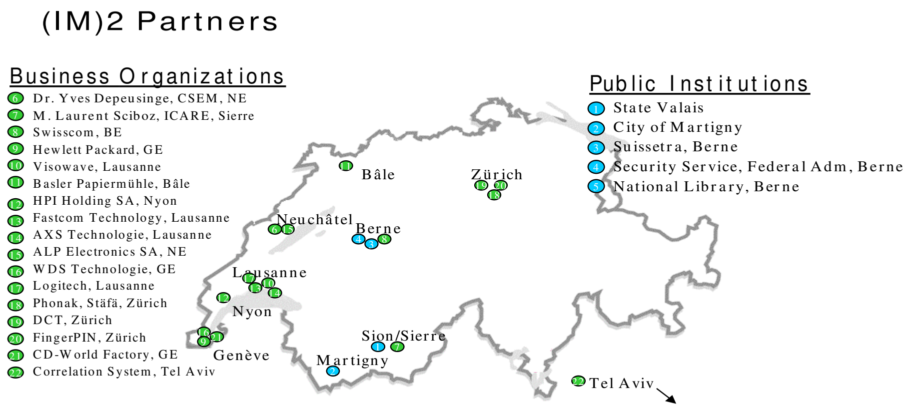
A cela vient s'ajouter la collaboration directe ou le support de nombreux partenaires industriels également répartis dans toute la Suisse. C'est le cas, par exemple, du CSEM de Neuchâtel et de ICARE/HES de Sierre pour les problèmes d'intégration logicielle. Ces soutiens démontrent le très grand intérêt dans les thèmes de développement et de recherche abordés dans (IM)2 (cf. deuxième carte ci-dessous).

2.3 Des institutions publiques

(IM)2 bénéficie également d'un support important de nombreuses institutions publiques, dont notamment :

- L'Etat du Valais et la Ville de Martigny, qui se sont engagés à renforcer significativement le financement de l'IDIAP (et donc du réseau) dans le cas de l'acceptation du projet, au travers d'un apport supplémentaire de 500'000.-/an.
- SUISSETRA, projet initié et financé par la Chancellerie Fédérale et touchant à certains domaines de (IM)2, et qui viendra se joindre aux efforts du réseau, notamment à travers l'ISSCO.

Finalement, d'autres institutions publiques telles que le Service de Sécurité de l'Administration Fédérale et la Bibliothèque Nationale ont exprimé un intérêt des plus vifs dans les thèmes abordés par (IM)2.



3. UN BREF DESCRIPTIF DU PROJET

3.1 Définition

"Interactive Multimodal Information Management", en court (IM)2, est la technologie qui coordonne des modes d'entrée naturelles (telles que parole, image, crayon, toucher, gestes de la main, mouvements de la tête et/ou du corps et même des capteurs physiologiques) avec des sorties de systèmes multimédia comme parole, sons, image, graphique 3D et animation.

Ces systèmes représentent une nouvelle direction pour les technologies de l'information. Ces interfaces multimodaux doivent accommoder, d'une manière souple, une grande variété d'utilisateurs, de tâches et d'environnements pour lesquels une quelconque modalité ne suffirait jamais. Les interfaces idéales doivent être capables, en premier lieu, de manipuler des données plus complètes et réalistes, y compris des données mixtes comme l'audio et la vidéo.

3.2 Activités concernées dans les domaines de la recherche et de l'économie

Le domaine des interactions multimodales couvre une gamme très large d'activités et d'applications comprenant la reconnaissance et l'interprétation de langages parlés, écrits et gesticulés, spécialement pour les systèmes d'informations multimédia. Les autres thèmes importants sous-jacents sont la protection du contenu informationnel, le contrôle d'accès et la structuration, la récupération et la présentation de l'information multimédia. Tout ceci nécessite une compréhension profonde des technologies de traitement des signaux complexes et bénéficiera énormément des approches différentes mais complémentaires, utilisées dans les composantes techniques d'importance stratégique.

Ainsi, le développement de cette technologie est obligatoirement multidisciplinaire et nécessite des contributions collaboratives d'experts en ingénierie, informatique, linguistique et psychologie. Toutefois, le facteur clé et fédérateur pour l'élaboration des systèmes multimodaux est que les technologies complémentaires sont souvent basées sur des outils mathématiques similaires qui, à leur tour s'appuient sur des formalismes mathématiques et statistiques bien définis.

Les progrès dans les systèmes multimodaux peuvent déclencher une tendance forte dans la qualité, l'utilité et l'accessibilité de l'informatique moderne, tout en créant de nouvelles opportunités d'industrialisation. Ces systèmes ont le potentiel de supporter des interactions homme-machine plus souples, faciles à apprendre et productives. Facilitant les interactions homme-machine en général, ces systèmes conduiront aussi, en particulier, à des interfaces améliorées facilitant leur utilisation par des usagers mobiles.

3.3 Une technologie d'avenir

Ces développements sont d'une importance croissante dans nos sociétés. Ils ont été identifiés comme des domaines de recherche clé dans le 5ème Programme-

Cadre de la Commission Européenne et par le programme américain DARPA et le US National Science Foundation.

Le management des systèmes d'information multimédia est un domaine très large et important de recherche qui inclut non seulement les interactions multimodales décrites ci-dessus mais également l'analyse, l'indexation et la récupération de documents multimédia. En conséquence, le pôle proposé ne pourra pas s'occuper de tous les aspects d'une façon exhaustive. Toutefois, étant donnée la nature fondamentale des systèmes d'information, les différentes composantes des interfaces multimodales s'appuieront sur des technologies similaires.

Par exemple, les mêmes technologies de base de reconnaissance de parole sont souvent utilisées en traitement des signaux pour beaucoup d'autres applications, y compris l'interaction avec un ordinateur et l'indexation de la vidéo. Une approche globale dans ce domaine peut tirer parti des nombreux points communs entre les composantes et générer des résultats intermédiaires utiles. L'expertise multidisciplinaire s'appliquera à plusieurs technologies, ainsi qu'à leur intégration.

3.4 Thèmes de recherche : perspectives multiples d'applications concrètes

Le pôle (IM)2 s'occupera donc de plusieurs thèmes de recherche importants, comprenant les entrées multimodales, les sorties multimodales, les coordinations entre les modalités, la sécurité, l'indexation et la récupération de l'information multimodale, l'évaluation objective des technologies, et leur intégration.

On peut notamment prévoir le développement des thèmes suivants avec des synergies optimales entre les différents laboratoires du réseau (IM)2 :

Technologies d'entrée parlée: couvrant le traitement du signal parole et la reconnaissance multilingue de parole. Les thèmes de recherche courants incluent : amélioration de la robustesse, portabilité sur d'autres applications, modélisation des langages, adaptation automatique, mesure de confiance, mots hors vocabulaire, parole spontanée, prosodie, adaptation dynamique des modèles, indexation d'information multimédia.

Exemple d'une application concrète : accès vocal à un service d'information multimédia peu structurée (ou service web).

Technologies d'entrée écrite: couvrant l'analyse de documents; OCR (imprimé et manuscrit, reconnaissance différée); écriture comme interface (reconnaissance en ligne). Les thèmes de recherche courants incluent: analyse de documents complexes, reconnaissance d'écriture dégradée, reconnaissance d'écriture courante. Les grands défis de recherche importants en reconnaissance manuscrite sont les séparations de mots et de ligne, la reconnaissance de grande lexiques et la modélisation de langage.

Exemple d'une application concrète : traitement automatisé de chèques ou autres formulaires manuscripts.

Technologies d'entrée visuelle: suivi de formes (comprenant le suivi du mouvement des lèvres, de visages); reconnaissance de gestes, d'expressions faciales (émotions) et d'images (p.ex., esquisses dessinées à la main, signatures, photos). Les thèmes de recherche courants incluent notamment: robustesse des algorithmes; combinaison de la couleur, du mouvement de la texture et de la forme dans l'analyse, analyse basée sur modèle plus précis, complexité de calcul.

Exemple d'une application concrète : reconnaissance et suivi de visages dans des documents vidéo, en vue de leur indexation automatique.

Analyse et compréhension d'entrée, couvrant la segmentation et l'analyse sémantique/syntaxique et la modélisation. Les thèmes de recherche courants incluent: spécification et formalisme des contraintes d'analyse sémantique/syntaxique unimodales et multimodales, utilisation de ces contraintes dans le traitement unimodal ou multimodal des entrées et le mixage des modalités en utilisant une grammaire multimodale.

Exemple d'une application concrète : accès à un système d'information multimédia au travers d'un dialogue naturel homme-machine.

Technologies de sortie parlée: codage et synthèse de parole ainsi que conversion texte à parole multilingue et sortie parlée. Les thèmes de recherche courants incluent: amélioration de la technologie de parole de base, modèles de calcul de la variabilité (pour éviter la monotonie), intégration de la synthèse et la génération de langage, adaptation et méthodes d'évaluation.

Exemple d'une application concrète : *accès aux messages email et informations sur le web au travers d'un assistant digital portable.*

Technologies de sortie visuelle: structuration de l'information et conversion de données (du graphique à la parole par exemple). Les thèmes de recherche courants incluent: visualisation d'information complexe de dimension élevée et conversion de média.

Exemple d'une application concrète : visualisation d'espaces virtuels (réalité virtuelle) et « télé-présence ».

Contrôle d'accès, comprenant : vérification du locuteur, reconnaissance de signature, reconnaissance de visages, identification multimodale d'utilisateurs (par exemple, en utilisant conjointement la parole et la vision). Les thèmes de recherche courants incluent: amélioration de la robustesse d'identification de l'utilisateur, identification multimodale (par expertise accumulée).

Exemple d'une application concrète: contrôle d'accès (bâtiment, internet, compte bancaire) par vérification de l'empreinte vocale et validation du visage.

Contrôle de contenu, nécessitant le développement de dispositifs renforçant des droits accessibles aux machines et des moyens sûrs de distribuer de l'information audio et vidéo. Les thèmes de recherche courants incluent: protection de contenu multimedia; étude de faisabilité, capacité, robustesse des techniques de marquage.

Exemple d'une application concrète : protection des droits d'auteurs sur les documents audio et video mis à disposition sur le web.

4. CONCLUSION

Compte tenu des résultats obtenus pour la première esquisse ainsi que l'importance et la qualité du réseau de partenaires mis en place pour la requête présentée en mars 2000, on peut affirmer que le projet présenté par l'IDIAP répond totalement aux exigences scientifiques et au concept des PRN. Le projet (IM)2 aborde en effet des thèmes hautement novateurs, mettant en œuvre une synergie importante entre de nombreuses institutions suisses à haut potentiel et bénéficiant d'une renommée internationale largement reconnue. Ce projet revêt donc une importance capitale pour toutes les institutions de recherche impliquées dans (IM)2.

Les décisions définitives concernant les huit à dix projets qui seront retenus reposeront maintenant davantage sur des critères d'ordre politique et de représentativité nationale des PRN. Dans ce contexte, l'esprit des PRN, qui prévoit justement un fonctionnement en réseau, devrait offrir une chance à un centre périphérique comme l'IDIAP d'être étroitement relié au monde de la recherche et de la science en Suisse et sur le plan international. De plus, la réputation internationale déjà reconnue de toutes les institutions participant à (IM)2 garantit une portée scientifique et économique du projet qui dépassera largement le cadre national.

A1. L'INSTITUT DALLE MOLLE D'INTELLIGENCE ARTIFICIELLE PERCEPTIVE (IDIAP)

A1.1 Historique

L'Institut Dalle Molle d'Intelligence Artificielle Perceptive (IDIAP, http://www.idiap.ch) est un institut de recherche semi-privé à but non lucratif situé à Martigny, Valais. Il a été créé en 1991 pour célébrer le 20ème anniversaire de la Fondation Dalle Molle, et représente le quatrième centre de recherche initié par cette fondation, après l'ISSCO à Genève (http://www.issco.ch), l'IDSIA à Lugano (http://www.idsia.ch) et MEDIPLANT à Conthey.

En novembre 1996, et comme convenu lors de sa création, l'IDIAP a acquis le statut de fondation de recherche (Fondation IDIAP), désormais indépendante de la Fondation Dalle Molle, et dont les fondateurs sont la Ville de Martigny, l'État du Valais, l'École Polytechnique Fédérale de Lausanne (EPFL), l'Université de Genève et Swisscom.

A1.2 Financement

Aujourd'hui, l'IDIAP est principalement financé par un support à long terme de la Confédération suisse (Office Fédéral de l'Éducation et de la Science, OFES selon la Loi fédérale sur la recherche), de l'État du Valais et de la Ville de Martigny. La Loterie Romande soutient également ces efforts de recherche au travers d'un subside annuel. En plus de ces subsides de base, l'IDIAP bénéficie de nombreux projets financés par le Fonds National Suisse de la Recherche Scientifique (FNSRS) pour de la recherche fondamentale (couvrant essentiellement des étudiants doctorants), de la Commission pour la technologie et l'innovation (CTI) ainsi que de l'OFES dans le cadre de nombreux projets européens.

A1.3 Activités

Aujourd'hui, l'IDIAP emploie entre 30 et 35 scientifiques, composés essentiellement de personnel permanent, de chercheurs post-doctorat, d'ingénieurs doctorants, de visiteurs à court ou moyen terme, et est active dans trois domaines de recherche importants et complémentaires, à savoir :

- la reconnaissance automatique de la parole et du locuteur
- la vision par ordinateur
- l'apprentissage automatique (par exemple, reconnaissance des formes et traitement de séries temporelles).

Les activités de l'IDIAP peuvent se répartir selon différentes catégories : les activités de recherche et développement, la participation à de nombreux projets de recherche européens et nationaux, les collaborations avec diverses organisations et sociétés et les activités d'enseignement et de formation. La mission de l'IDIAP consiste donc en :

- La poursuite d'activités de recherche fondamentale et appliquée, dans le but de transfert technologique à court, moyen et long terme.
- L'enseignement et la formation.

Durant ces dernières années, les activités de l'IDIAP ont été des plus florissantes. Le nombre de projets nationaux et internationaux ainsi que le partenariat avec les institutions académiques n'ont cessé de s'accroître. De plus, grâce au soutien continu de nos institutions et aux compétences élevées de notre personnel motivé par un travail d'excellente qualité, l'IDIAP est maintenant reconnu comme un partenaire de haut niveau dans ses domaines de compétence (traitement automatique de la parole, vision par ordinateur, apprentissage automatique). L'IDIAP tend maintenant à poursuivre et à renforcer ses activités de recherche et développement dans ses domaines d'expertise tout en favorisant le transfert technologique au travers de partenariats industriels (y compris avec VOXCom, une société «spin-off» de l'IDIAP démarrée en juillet 1998).