



Martigny - Valais - Suisse

## ACTIVITY REPORT 2000

IDIAP-COM 2001-01

FEBRUARY 2001

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](mailto:secretariat@idiap.ch)  
internet <http://www.idiap.ch>



## Institut Dalle Molle d'Intelligence Artificielle Perceptive

### MEMBERS

Supporting:

- Swiss Confederation, Federal Office for Education and Science
- State of Valais
- City of Martigny
- Loterie Romande

Affiliated:

- Swiss Federal Institute of Technology at Lausanne (EPFL)
- University of Geneva

### FOUNDATION COUNCIL

Pierre Crittin (Chairman, President of the City of Martigny), Jean-Pierre Rausis (Secretary, Director of BERSY), Hervé Bourlard (Director of IDIAP, Professor at EPFL), Pierre Dal Pont (Director of NOFIDA), Daniel Forchelet (Swisscom, Skill Family Manager), Gilbert Fournier (State of Valais), Jurg Hérold (Director of CIMO SA), Nicolas Markwalder (Attorney at Law, Delegate of the Economic Commission, Bern), Jérôme Sierro (University of Geneva), Dominique de Werra (Professor, Vice-President of EPFL).

### BOARD OF DIRECTORS

Jean-Pierre Rausis (Chairman, Director of BERSY), Pierre Dal Pont (Secretary, Director of NOFIDA), Hervé Bourlard (Director of IDIAP, Professor at EPFL), Daniel Forchelet (Swisscom, Skill Family Manager), Gilbert Fournier (State of Valais), Jurg Hérold (Director CIMO SA), Nicolas Markwalder (Attorney at Law, Delegate of the Economic Commission, Bern), Christian Pellegrini (Professor, University of Geneva), Léopold Pflug (Professor, EPFL).

### SCIENTIFIC COMMITTEE

Prof. Christian Pellegrini (Chairman, University of Geneva, CH), Prof. Hervé Bourlard (Director IDIAP, Professor EPFL), Dr. Robin Breckenridge (F. Hofmann-La Roche Ltd, CH), Prof. Giovanni Coray (EPFL, CH), Dr. J. Cywinsky (Institute of Medical Technology, CH), Prof. Wulfram Gerstner (EPFL, CH), Prof. Martin Hasler (EPFL, CH), Prof. Jean-Paul Haton (CRIN/INRIA, F), Prof. Beat Hirsbrunner (University of Fribourg, CH), Prof. Rolf Ingold (University of Fribourg, CH), Prof. Eric Keller (University of Lausanne, CH), Prof. Nelson Morgan (ICSI and UCB, Berkeley, USA), Prof. Beat Pfister (ETH, CH), Prof. Thierry Pun (University of Geneva, CH), Prof. Ian Smith (EPFL, CH), Mr. Robert Van Kommer (Swisscom, CH), Prof. Eric Vittoz (CSEM and EPFL, CH), Prof. Christian Wellekens (EURECOM, F).



## Table of Contents

<b>1</b>	<b>Introduction (in English)</b>	<b>1</b>
<b>2</b>	<b>Introduction (en français)</b>	<b>3</b>
<b>3</b>	<b>Staff</b>	<b>5</b>
3.1	Scientific Staff . . . . .	5
3.2	Visitors . . . . .	7
3.3	Students . . . . .	7
3.4	Administrative Staff . . . . .	7
<b>4</b>	<b>Research Activities</b>	<b>9</b>
4.1	Speech Processing Group . . . . .	10
4.1.1	Research Themes . . . . .	10
4.1.2	Application Examples . . . . .	11
4.2	Computer Vision Group . . . . .	12
4.2.1	Research Themes . . . . .	12
4.2.2	Application Examples . . . . .	14
4.3	Machine Learning Group . . . . .	14
4.3.1	Research Themes . . . . .	14
4.3.2	Application Examples . . . . .	15
<b>5</b>	<b>Current Projects</b>	<b>15</b>
<b>6</b>	<b>Educational Activities</b>	<b>30</b>
6.1	Current PhD Theses . . . . .	30
6.2	PhD Defenses . . . . .	30
6.3	Participations to PhD Thesis Committees . . . . .	30
<b>7</b>	<b>Scientific Activities</b>	<b>33</b>
7.1	Editorship . . . . .	33
7.2	Scientific and Technical Committees . . . . .	33
7.3	Short Term Visits . . . . .	34
7.4	Scientific Presentations (other than conferences) . . . . .	35
<b>8</b>	<b>Publications (1999 and 2000)</b>	<b>38</b>
8.1	Books and Book Chapters . . . . .	38
8.2	Articles in International Journals . . . . .	38
8.3	Articles in Conference Proceedings . . . . .	39
8.4	IDIAP Research Reports . . . . .	43
8.5	IDIAP Communications . . . . .	45
8.6	Other Documents . . . . .	46



## 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 2000, 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 2000, IDIAP numbered around 30-35 scientists, including permanent staff, postdoctoral fellows, PhD students (around 18), 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 2000, 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 exemple, the number of **national and international projects** has significantly increased, and many new projects were granted or started in 2000. As of this writing, there are about 15 SNSF and 10 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, VOXCom 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, 11 journal papers, 67 international conference papers, and 42 unpublished (or not yet published) internal research reports.

Finally, the **partnerships with academic institutions** have also significantly been strengthened. In this framework, an important success for IDIAP in 2000 was its final selection as “Leading House” of one of the potential **National Centres of Competence in Research** (NCCR), which should start during the Summer 2001. Following a very strict and competitive selection process, the Centre on “Interactive Multimodal Information Management (IM)2”, proposed by IDIAP and centered on many of its research activities, has indeed been selected as one potential NCCR. 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 would also confirm the leading role of IDIAP at the national level in the targeted research areas.

Thanks to the continued support of our authorities, and to our most competent personnel, motivated to the highest level, IDIAP is 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 2000, 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 2000, l’IDIAP employait environ 30-35 scientifiques, composés essentiellement de personnel permanent, de chercheurs post-doctoraux, d’ingénieurs doctorants (environ 18), et d’ingénieurs 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 2000, 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 2000. A ce jour, environ 15 projets du Fonds National Suisse de la Recherche Scientifique et 10 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, 11 articles dans des revues internationales, 67 articles dans des conférences internationales, et 42 rapports scientifiques internes non publiés (ou pas encore publiés).

Finalement, les **collaborations avec les institutions académiques** se sont également fortement renforcées. Dans ce contexte, un succès important pour l’IDIAP en 2000 a été sa sélection finale comme “Leading House” d’un des **Pôles de Recherche Nationaux** (PRN) potentiels qui devraient démarrer dans le courant de l’été 2001. Après un long processus de sélection, strict et particulièrement sélectif, le pôle “Interactive Multimodal Information Management (IM)2” proposé par l’IDIAP, et centré sur de nombreuses activités de notre institut, a effectivement été sélectionné comme PRN potentiel. 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 renforçant 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.

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 Staff

**Mail:** IDIAP — Institut Dalle Molle d'Intelligence Artificielle Perceptive  
 CP 592  
 CH-1920 Martigny (VS)  
 Switzerland

**Phone:** +41 – 27 – 721 77 11

**Fax:** +41 – 27 – 721 77 12

**Internet:** <http://www.idiap.ch>

#### 3.1 Scientific Staff

Persons at IDIAP in 2000 or as of this writing:

Dr.	Samy BENGIO	Machine Learning group leader
	<a href="mailto:Samy.Bengio@idiap.ch">Samy.Bengio@idiap.ch</a>	+41 – 27 – 721 77 39
Mr.	Mohamed BENZEGHIBA	research assistant
	<a href="mailto:Mohamed.Benzeghiba@idiap.ch">Mohamed.Benzeghiba@idiap.ch</a>	+41 – 27 – 721 77 41
Ms.	Giulia BERNARDIS	research assistant
	<a href="mailto:Giulia.Bernardis@idiap.ch">Giulia.Bernardis@idiap.ch</a>	until June 2000
Prof.	Hervé BOURLARD	Director, Professor EPFL
	<a href="mailto:Herve.Bourlard@idiap.ch">Herve.Bourlard@idiap.ch</a>	+41 – 27 – 721 77 20
Mr.	Datong CHEN	research assistant
	<a href="mailto:Datong.Chen@idiap.ch">Datong.Chen@idiap.ch</a>	+41 – 27 – 721 77 56
Mr.	Thierry COLLADO	development engineer
	<a href="mailto:Thierry.Collado@idiap.ch">Thierry.Collado@idiap.ch</a>	+41 – 27 – 721 77 42
Mr.	Ronan COLLOBERT	research assistant
	<a href="mailto:Ronan.Collobert@idiap.ch">Ronan.Collobert@idiap.ch</a>	+41 – 27 – 721 77 31
Mr.	Beat FASEL	research assistant
	<a href="mailto:Beat.Fasel@idiap.ch">Beat.Fasel@idiap.ch</a>	+41 – 27 – 721 77 23
Mr.	Frank FORMAZ	System Management group leader
	<a href="mailto:Frank.Formaz@idiap.ch">Frank.Formaz@idiap.ch</a>	+41 – 27 – 721 77 28
Mr.	Nicolas GILARDI	research assistant
	<a href="mailto:Nicolas.Gilardi@idiap.ch">Nicolas.Gilardi@idiap.ch</a>	+41 – 27 – 721 77 47
Mr.	Hervé GLOTIN	research assistant
	<a href="mailto:Herve.Glotin@idiap.ch">Herve.Glotin@idiap.ch</a>	+41 – 27 – 721 77 33
Mr.	Eric GRAND	development engineer
	<a href="mailto:Eric.Grand@idiap.ch">Eric.Grand@idiap.ch</a>	until July 2000
Ms.	Astrid HAGEN	research assistant
	<a href="mailto:Astrid.Hagen@idiap.ch">Astrid.Hagen@idiap.ch</a>	+41 – 27 – 721 77 34
Mr.	Shajith IKBAL	research assistant
	<a href="mailto:Shajith.Ikbal@idiap.ch">Shajith.Ikbal@idiap.ch</a>	+41 – 27 – 721 77 46

Prof.	Mikhael KANEVSKI <i>Mikhael.Kanevski@idiap.ch</i>	research scientist +41 - 27 - 721 77 49
Mr.	Sacha KRSTULOVIC <i>Sacha.Krstulovic@idiap.ch</i>	research assistant +41 - 27 - 721 77 43
Dr.	Mikko KURIMO <i>Mikko.Kurimo@idiap.ch</i>	research scientist until May 2000
Mr.	Bertrand LIARDON <i>Bertrand.Liardon@idiap.ch</i>	development engineer until September 2000
Dr.	Jürgen LÜTTIN <i>Juergen.Luettin@idiap.ch</i>	Computer Vision group leader until September 2000
Mr.	Mathew MAGIMAI DOSS <i>Magimaidoss.Mathew@idiap.ch</i>	research assistant +41 - 27 - 721 77 51
Dr.	Sebastien MARCEL <i>Sebastien.Marcel@idiap.ch</i>	research scientist +41 - 27 - 721 77 27
Mr.	Johnny MARIÉTHOZ <i>Johnny.Mariethoz@idiap.ch</i>	research engineer +41 - 27 - 721 77 44
Mr.	Perry MOERLAND <i>Perry.Moerland@idiap.ch</i>	research assistant until May 2000
Mr.	Miguel MOREIRA <i>Miguel.Moreira@idiap.ch</i>	research assistant until December 2000
Dr.	Andrew MORRIS <i>Andrew.Morris@idiap.ch</i>	research scientist +41 - 27 - 721 77 35
Mr.	Bojan NEDIĆ <i>Bojan.Nedic@idiap.ch</i>	research assistant until September 2000
Ms.	Maja POPOVIĆ <i>Maja.Popovic@idiap.ch</i>	research assistant +41 - 27 - 721 77 53
Dr.	Kimberly SHEARER <i>Kim.Shearer@idiap.ch</i>	research scientist +41 - 27 - 721 77 26
Mr.	Todd STEPHENSON <i>Todd.Stephenson@idiap.ch</i>	research assistant +41 - 27 - 721 77 52
Mr.	Alex TRUTNEV <i>Alex.Trutnev@idiap.ch</i>	research assistant +41 - 27 - 721 77 38
Mr.	Alessandro VINCIARELLI <i>Alessandro.Vinciarelli@idiap.ch</i>	research assistant +41 - 27 - 721 77 24
Mrs.	Haiyan WANG <i>Haiyan.Wang@idiap.ch</i>	development engineer +41 - 27 - 721 77 54
Mrs.	Katrin WEBER <i>Katrin.Weber@idiap.ch</i>	research assistant +41 - 27 - 721 77 37

### 3.2 Visitors

Mr.	Joerg BUCHHOLZ	University of Bochum, Germany
Mr.	Sebastian MOELLER <i>Sebastian.Moeller@idiap.ch</i>	University of Bochum, Germany
Ms.	Susagna POL FONT <i>Susagna.Pol.Font@idiap.ch</i>	European Masters in Speech and Language
Mr.	Pere PUJOL <i>Pere.Pujol@idiap.ch</i>	European Masters in Speech and Language
Mr.	Vesa SIIVOLA <i>Vesa.Siivola@idiap.ch</i>	University of Technology, Helsinki, Finland
Mr.	Vivek TYAGI <i>Vivek.Tyagi@idiap.ch</i>	Indian Institute of Technology, Kanpur, India

### 3.3 Students

Mr.	Christopher BOISSET <i>Christopher.Boisset@idiap.ch</i>	from July 2000 to August 2000
Mr.	Santiago CRUZ <i>Santiago.Cruz@idiap.ch</i>	from May 2000 to July 2000
Mr.	Olivier GRANGES <i>Olivier.Granges@idiap.ch</i>	from July 2000 to August 2000

### 3.4 Administrative Staff

Mrs.	Sylvie MILLIUS <i>Sylvie.Millius@idiap.ch</i>	secretary +41 - 27 - 721 77 21
Mrs.	Nadine ROUSSEAU <i>Nadine.Rousseau@idiap.ch</i>	secretary +41 - 27 - 721 77 22



## 4 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.

## 4.1 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.

### 4.1.1 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.

#### 4.1.2 Application Examples

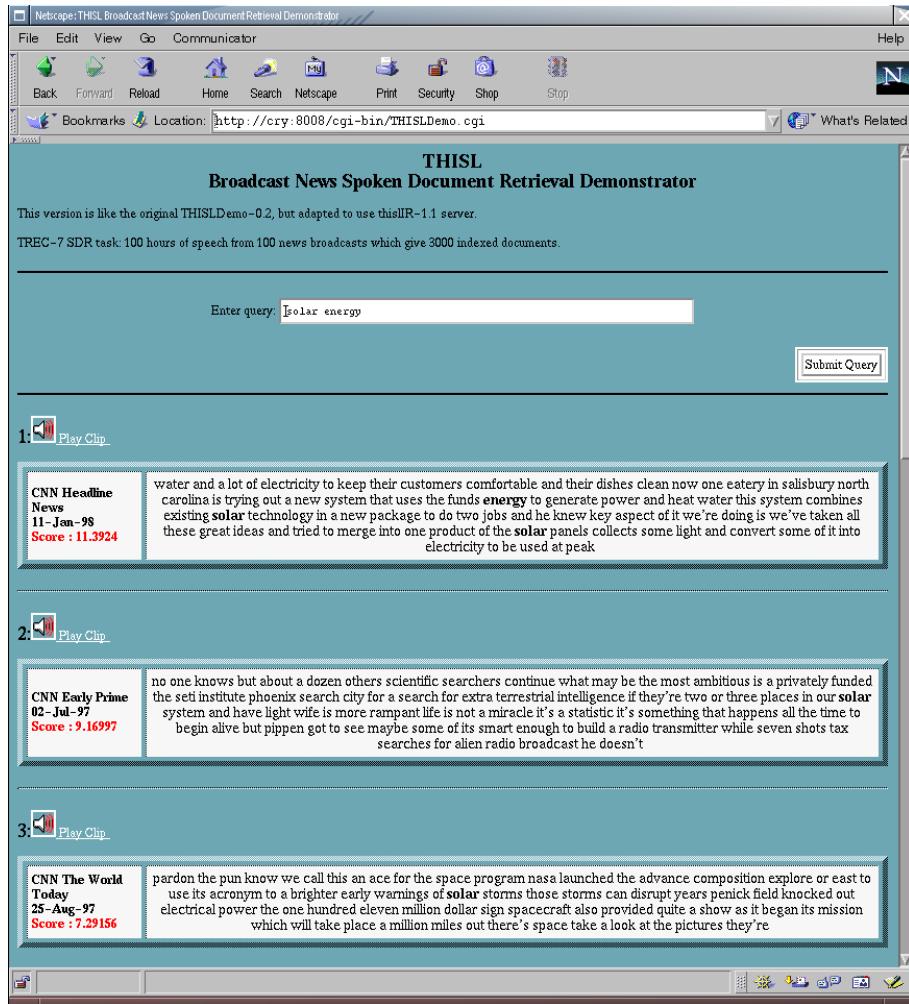


Figure 1: Interface of the THISL audio indexing and retrieval system.

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 project.
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.
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 THISL, ASSAVID and CIMWOS projects.

As illustrated in Figure 1, and further discussed in Section 5 (project description), THISL is a real-time prototype system for navigating in the sound-track of a TV news broadcast, including:

- 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.

## 4.2 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.

### 4.2.1 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 pre-processing: Statistical methods and HMM techniques are applied to improve the normalisation and word modeling process, continuing improvements in the preprocessing step for hand written and cursive text.
3. Sentence recognition: Speech recognition methods such as language models are being applied to improving sentence recognition in script recognition. This is made possible by the statistical word modeling approach taken in earlier processing.
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 Detection and Gesture Recognition: Persons detection and analysis is a challenging problem in computer vision for human computer interaction. Real-time computer vision systems, which detects and tracks a face in a sequence of video images coming from a camera, exist. In such systems, hand postures can be used to execute commands and hand gesture recognition can be used to detect the intention of the user to execute a command.

- **Image Processing for Face and Hand Segmentation:** We are interested in face detection and hand gesture recognition. Consequently, we must segment faces and hands from the image. We filter the image using a fast look-up indexing table of skin color pixels in YUV color space. After filtering, skin color pixels are gathered into blobs, which are statistical objects based on the location and the colorimetry of the skin color pixels in order to determine homogeneous areas.
- **Face Detection:** The task of face detection consists in the analysis of an entire image aiming to detect all faces that appear in the image. We investigate this problem in the general case of object detection.
- **Hand Posture Recognition:** We propose to use a neural network model already applied to face detection: the constrained generative model (CGM). The goal of the constrained generative learning is to closely fit the probability distribution of the set of hands using a non-linear compression neural network and non-hand examples. After learning, the Euclidean distance between inputs and outputs of the neural network is smaller for hand examples than for non-hands examples. Using this distance, and given a distance threshold estimated on a generalisation set, the classification of the input image (Figure 2) is possible.

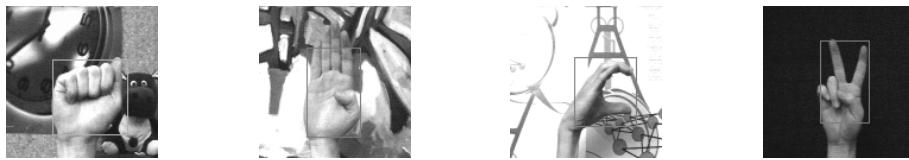


Figure 2: Examples of hand posture from Jochen Triesch Gallery used for testing our CGM based detector.

- **Hand Gesture Recognition:** Our goal is to recognize two classes of gestures: deictic and symbolic gestures. Deictic gestures are pointing movements towards the left (right) of the body-face space and symbolic gestures are intended to execute commands (grasp, clic, rotate) on the left (right) of shoulders. We propose to use Input-Output Hidden Markov Models (IOHMM) which have HMM properties and NN discrimination efficiency.

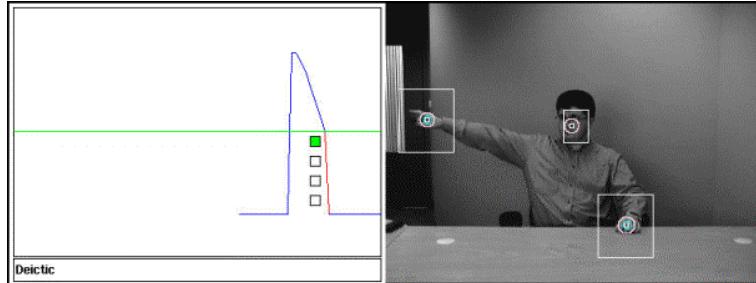


Figure 3: Recognition of a deictic gesture using a IOHMM. The Likelihood (on the left) indicate the end of the gesture and the marker identify it.

First, blobs are tracked in the image. Second, the 2D trajectory of the hand-blob<sup>1</sup> during a gesture is extracted. Finally, the extracted trajectory is given to a trained IOHMM which classifies the gesture (Figure 3).

<sup>1</sup>center of gravity of the blob corresponding to the hand

#### 4.2.2 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.

### 4.3 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.

#### 4.3.1 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.

#### 4.3.2 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 warehouses (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.

## 5 Current Projects

### ◊ ARTIST – Articulatory Representation To Improve Speech Technologies

**Funding:** Swiss National Science Foundation

**Duration:** April 1999 - March 2001

**Contact persons:** Sacha Krstulovic, 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, Mark Barnard

**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.



**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.

◊  **B**N-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 - February 2001

**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.

◊  **C**ARTANN – Cartography by Artificial Neural Networks

**Funding:** Swiss National Science Foundation

**Duration:** January 1999 - January 2001

**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.

◊  **C**IMWOS – Combined IMages and WOrd Spotting

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

**Duration:** 30 months (expected to start in April 2001)

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

**Contact persons:** 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 278 – Spoken Language Interaction in Telecommunication

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

**Duration:** 4 years (expected to start in June 2001)

**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

**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.

◊ **D**ivide and Learn

**Funding:** Swiss National Science Foundation

**Duration:** April 1996 - March 2000

**Partners:** Swiss Federal Institute of Technology (EPFL)

**Contact persons:** Perry Moerland, Samy Bengio

**Description:** The aim of this project is the study and the extension of the mixture of experts model. This model adaptively partitions the input space (using a gating network) and attributes local experts to these regions. This model has shown to be a powerful tool for dealing with classification and regression problems. The goal of this project was to better understand the influence of the choice of the gating network and develop new methods for learning the parameters of “mixture of experts”-like models.



**Funding:** Swiss National Science Foundation

**Duration:** October 2000 - September 2002

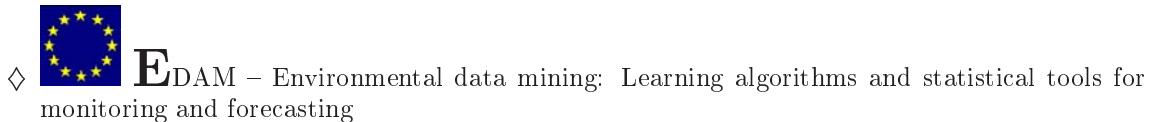
**Partners:** Swiss Federal Institute of Technology (EPFL)

**Contact persons:** 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.



**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.

◊  **F**aceX – 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 animation of synthetic faces.

In the second year of the project FacEx, preliminary experiments have been carried out in order to evaluate the effectiveness of appearance-based, data-driven feature extraction and representation methods such as PCA, ICA and Kernel PCA. For the first time Kernel PCA has been applied for the analysis of facial expressions yeilding even better recognition results than with ICA, which is probably due to a non-linear mapping of the input space, allowing for a more accurate separation of overlapping facial actions signals. So far, we have relied on difference-image based facial motion extraction which reduces physiognomical dependencies and concentrates on facial motions only. It works well, as long as faces do not move and a neutral reference frame is available. This is, however, highly problematic in natural environments. Therefore, other facial feature extraction methods were examined and Gabor filters found to be an interesting alternative. They allow a reduction in the influence of lighting changes, while focusing on line and edges, which are important indicators for facial expressions. This results in a reduced size of the sub-space of facial expressions, similarly to the afore mentioned difference-image approach. A first automatic facial expression recognition system has been built, that can deal with asymmetric facial expressions. It does not rely on hand-crafted rules or any other artificial means such as markers. We have obtained recognition performance comparable to a human expert. Finally, FACS showed to be an adequate coding scheme, as it allows to separate facial expression recognition from facial expression interpretation. Individual facial activities can thus be evaluated more reliably, without being limited to a single category of facial signals such as emotions.



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

**Duration:** 30 months, to be started in 2001

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

**Contact person:** Sébastien Marcel

**Description:** FGnet is a Concerted Action and Thematic Network on Face and Gesture Recognition. The aim of this project is to encourage technology development in the area of face and gesture recognition. The precise goals are:

1. to act as a focus for the workers developing face and gesture recognition technology
2. to create 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. to specify, develop and supply resources (e.g. image sets) supporting these scenarios
4. to use these resources to encourage technology development.

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 basis of project, is that when properly defined and collected, such resources would also be of benefit in the development of wider problems in face and gesture recognition.

Foresight workshops will attempt to describe a set of possible futures enabled by intelligent methods in face and gesture recognition. Participants will include members of the consortium, as well as invited speakers from active EU projects across action lines, and national programs. The output will be a white paper indicating development roadmaps and defining scenarios and opportunities for the technology in the medium and long term. The report will be circulated within the community for comment.

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. This will be based on the scenarios defined in the visioning workshops. They will form baseline systems upon which the community could build. This activity will consist of a feasibility stage, a specification stage to define the size and content of the data to be collected, a collection and annotation stage and an archiving stage. The collection of data will be spread amongst the partners to exploit the various possibilities (diverse racial and cultural types). Speech will be collected in parallel where appropriate.

The resources will be made freely available to the community, via a website. Resource workshops will be held at major conferences and industrial events will diffuse information on the data and showcase baseline results. Additional complementary data sets and resources will be placed on the website where possible.



**Funding:** Swiss National Science Foundation

**Duration:** November 1998 - October 2000

**Partners:** Swiss Federal Institute of Technology (EPFL)

**Contact persons:** Miguel Moreira, Samy Bengio

**Description:** This project is about the generalization of Logical Analysis of Data (LAD) into a method capable of handling classification problems with large databases. LAD has been shown to be a very efficient machine learning technique for several types of databases. However, so far it is limited to classification problems with two classes only. Moreover, the algorithms available for the resolution of each step of the method scale up very badly with the size of the database (number of data items and number of attributes). In particular, the method designed to solve the first phase of the process (the binarization phase) is quadratic in the number of data, and thus is not usable for problems with more than a couple of hundreds of data items.

In this project, several solutions to generalize LAD to  $K$ -class classification problems are proposed. New algorithms to make the whole process suitable for large scale problems are developed. For example, a new algorithm solving the binarization for a problem of  $n$  data in  $O(n \log(n))$  is designed.

◊  **H**MM2 – 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.

◊  **I**nfoVOX – 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, VOXCom 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 focus on the development of a natural voice interface to the restaurants in Martigny.

◊  **I**NSPECT – 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.

◊  **K**ERNEL – 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.

◊  **M**ULTICHAN – 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 is 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.

◊  **P**ICASSO – PIoneering Caller Authentication for Secure Service Operation

**Funding:** European project, 4th Framework Programme, Telematics, supported by OFES

**Duration:** March 1998 - September 2000

**Partners:** IDIAP, Ubilab (CH), Swisscom (CH), ENST (F), IRISA (F), PTT-Telecom (NL), KPN Research (NL), KUN (NL), Fortis (NL), KTH (SE), Telia (SE), and Vocalis (UK)

**Contact persons:** Hervé Bourlard, Johnny Mariethoz

**Description:** The main goals of PICASSO are the improvement of speaker verification systems, their testing in real life conditions, and the development of banking and telecom applications. More specifically, the project aims at integrating speaker verification with automatic speech recognition (ASR) to develop a new generation of telephone enquiry systems, combining high-accuracy customer verification with easy-to-use speech recognition interfaces, with applications to telephone calling cards/accounts, messaging services (“voice mail”) and retail banking services.

On top of scientific achievements, IDIAP collected the Polyvar database, particularly designed to speaker verification research, and distributed it to the PICASSO partners. Through their regular participation to the international NIST (National Institute of Standards and Technology, USA) evaluation, IDIAP (in collaboration with their partners) has always shown competitive performance.

◊  **P**ROMO – PROnunciation MOdelling in Automatic Speech Recognition Systems

**Funding:** Swiss National Science Foundation

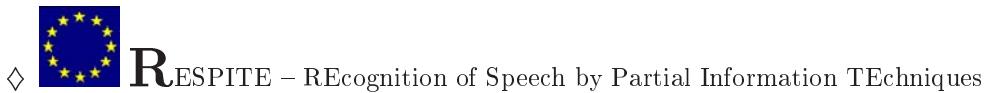
**Duration:** August 2000 - July 2002

**Contact persons:** Mathew Magimai Doss, Hervé Bourlard

**Description:** Natural speech and casual human conversation exhibit a large amount of non-standard 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.



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

**Duration:** January 1999 - December 2001

**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).



**Funding:** Swiss National Science Foundation

**Duration:** October 1999 - September 2001

**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.

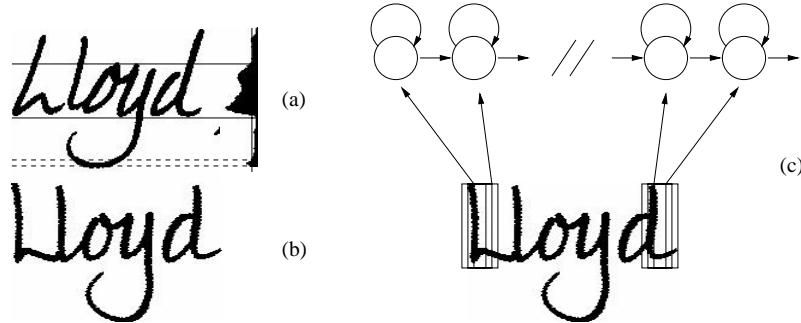


Figure 4: Single word recognition. The original image (a) is normalized (b) and modeled with HMMs (c). A HMM is created for every word in a list of possible interpretations of the data. The most likely model is assumed as transcription of the data.

**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.



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

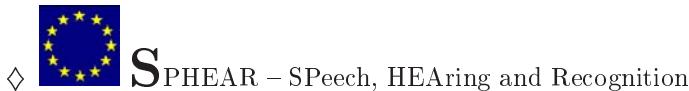
**Duration:** September 1997 - September 2001

**Partners:** Univ. of Saarlandes (D), Aalborg Univ (DK), Univ. of Sheffield (UK), Univ. of Essex (UK), Univ. of Edinburgh (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/EuroMasters/> 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).



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

**Duration:** March 1998 - February 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.



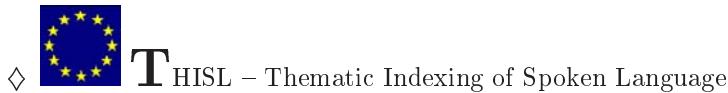
**Funding:** Swiss National Science Foundation

**Duration:** January 1999 - December 2000

**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 pronunciations 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.



**Funding:** European project, 4th Framework Programme, Long Term Research, supported by OFES

**Duration:** February 1997 - January 2000

**Partners:** Sheffield University (UK), BBC (UK), FPMs (B), SoftSound (UK), Thomson-CSF (F), IDIAP (CH), and ICSI (USA, Subcontractor)

**Contact persons:** Hervé Bourlard

**Description:** The overall objective of the THISL project was to produce a demonstrator system for content-based indexing and retrieval of TV and news broadcasts. This has been achieved: at the project end a demonstrator containing 1800 hours of automatically transcribed and indexed BBC news output (updated daily) had been installed and made available on the BBC intranet. Evaluation both by users within the BBC and through an international evaluation programme conducted by the US National Institute of Standards and Technology (NIST) have indicated that the THISL system is at the state-of-the-art technologically, and offers a valuable new service to archive users, with the promise of a much broader domain of applicability.

The THISL system is based on a connectionist/statistical speech recogniser for radio and TV broadcasts, which produces a word-level transcription. Indexing and probabilistic retrieval techniques are then used, to enable users to search for new items of interest to them. The interface of the demonstrator is similar to that of a web search engine; an alternative spoken query interface was also developed.

The principal technical achievements of the projects are as follows:

1. Development of a broadcast news retrieval demonstrator, installed and in daily use at the BBC.
2. Development of an alternative spoken query interface, using natural language processing technology to recover from recognition errors.
3. Development of a speech recognition system for British English broadcast speech.

4. Investigation of novel information retrieval strategies based on latent semantic analysis: this was one of the main contributions of IDIAP.
5. Development and implementation of algorithms for tracking speakers in radio and TV broadcasts.
6. Development and implementation of a novel decoding algorithm for large vocabulary speech recognition to enable real-time recognition of broadcast speech with a relatively low memory overhead.
7. Development and implementation of query expansion and segmentation algorithms for information retrieval.
8. Development and implementation of new probabilistic confidence measures on speech recogniser output: contributions from IDIAP.
9. Adaptation of the basic THISL demonstrator to French language broadcast news: contributions from IDIAP.
10. Investigation of novel approaches to speech/music discrimination.
11. Development of a keyword spotting algorithm based on a new technique referred to as Iterative Viterbi Decoding: developed by IDIAP.

Each of these areas resulted in (a) software modules that were available for incorporation in the THISL system, and (b) scientific papers published in leading journals and conferences.



**Funding:** Swiss National Science Foundation

**Duration:** December 1999 - November 2001

**Person involved:** Datong Chen

**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 by 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.

In the past year work on the project has included collection of a representative database of video containing text. This includes both caption text of various forms and “in vision” text. 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.

## 6 Educational Activities

### 6.1 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 5.

### 6.2 PhD Defenses

- **Ph.D. candidate:** Perry Moerland

**Supervisor:** Prof. W Gerstner (EPFL)

**Examiners:** Prof. B Faltings (EPFL), Dr. E Mayoraz (Motorola), Prof. C Pellegrini (Uni Geneva), Dr. J Schmidhuber (IDSIA).

**University:** EPFL, Lausanne

**Title:** Mixture Models for Unsupervised and Supervised Learning

- **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

### 6.3 Participations to PhD Thesis Committees

- **Ph.D. candidate:** Stéphane Dupont

**Committee member:** Hervé Bourlard

**University:** Faculté Polytechnique of Mons, Mons, Belgium

**Date:** June 23

**Title:** Etude et développement d'architectures multi-bandes et multi-modales pour la reconnaissance robuste de la parole

- **Ph.D. candidate:** Sébastien Marcel

**Committee member:** Samy Bengio

**University:** Université de Rennes I, France

**Date:** October 4

**Title:** Une approche générative neuro-markovienne du traitement de séquences d'images: Application à la reconnaissance statique et dynamique des gestes de la main

- **Ph.D. candidate:** Miguel Moreira

**Committee member:** Samy Bengio

**University:** Ecole Polytechnique Fédérale de Lausanne, Switzerland

**Date:** December 6

**Title:** The use of Boolean concepts in general classification contexts

	PhD Students	Funding	Project	Expected PhD	Start Date at IDIAP	PhD Status	Thesis Supervisor	Thesis Director
<i>SNSF FUNDING</i>								
1	Mohamed BENZEGHIBA	FN 21.00-05.40/18.98/1	SV-UCP	2004	01.08.00	1st year, Speech	H. Bourlard	Prof. H. Bourlard, EPFL
2	Datong CHEN	FN 21.00-05.7231.99/1	VOCR	2003	01.11.99	2nd year, Vision	K. Shearer	Not decided yet
3	Ronan COLLOBERT	FN 21.00-06.1243.00/1	Divide & Learn II	2004	01.04.00	1st year, Learning	S. Bengio	Not decided yet
4	Beat FAESL	FN 21.51-54.000.98	FACEX	2002	01.10.98	3rd year, Vision	S. Marcel	Prof. Van Gool, ETHZ
5	Nicolas GILARDI	FN 21.00-05.4115.98	CARTANN	2002	01.01.99	3rd year, Learning	S. Bengio	Prof. Maignan, UNIL
6	Shajith IKBAL	FN 21.00-06.1325.00/1	HMM2	2004	01.05.00	1st year, Speech	H. Bourlard	Prof. H. Bourlard, EPFL
7	Sacha KRSTULOVIC	FN 2000-05.5634.98/1	ARTIST 2	2001	01.04.96	4th year, Speech	H. Bourlard	Prof. M. Hasler, EPFL
8	Quan LE	FN 21.00-06.1245.00/1	KERNEL	2004	01.02.01	1st year, Learning	S. Bengio	Not decided yet
9	Mathew MAGIMAI DOSS	FN 21.00-05.7245.99/1	PROMO	2004	25.10.99	1st year, Speech	H. Bourlard	Prof. H. Bourlard, EPFL
10	Perry MOERLAND	FN 21.00-06.1243.00/1	Divide & Learn	2000	30.04.94	Accepted, Learning	S. Bengio	Prof. W. Gerstner, EPFL
11	Miguel MOREIRA	FN 2000-05.3902.98/1	GLAD	2000	01.11.96	Accepted, Learning	S. Bengio	Prof. A. Hertz, EPFL
12	Todd STEPHENSON	FN 21.00-05.3960.98/1	BNS 4 ASR	2003	01.03.99	3rd year, Speech	A. Morris	Prof. H. Bourlard, EPFL
13	Alex TRUTINOV	FN 21.00-05.4100.98/1	INSPECT	2004	01.08.00	1st year, Speech	M. Rajman	Prof. H. Bourlard, EPFL
14	Alessandro VINCIARELLI	FN 21.00-05.5733.98/1	SCRIPT	2003	01.10.99	2nd year, Vision	S. Bengio	Prof. H. Bunke, Univ. Bern
15	Katrin WEBER	FN 2000-59.169.99/1	CORREL	2001	01.01.98	4th year, Speech	H. Bourlard	Prof. H. Bourlard, EPFL
<i>OFES FUNDING</i>								
16	Jitendra AJMERA	OFES 98.0086	RESPITE	2004	01.01.01	1st year, Speech	H. Bourlard	Not decided yet
17	Astrid HAGEN	OFES 97.0288	SPHEAR	2001	01.11.97	4th year, Speech	H. Bourlard	Prof. H. Bourlard, EPFL
18	Maja POPOVIC	OFES 99.0562	ASSAVID	2004	01.03.00	1st year, Speech	H. Bourlard	Not decided yet
19	Mark BARNARD	OFES 99.0562	ASSAVID	2005	15.03.01	1st year, Speech	H. Bourlard	Not decided yet
20	Hervé GLOTIN	OFES 97.0288	SPHEAR	2001	01.04.97	4th year, Speech	H. Bourlard	Prof. H. Bourlard and Berthommier

- **Ph.D. candidate:** Joseph Rynkiewicz
- Committee member:** Hervé Bourlard
- University:** Université Paris I, Panthéon-Sorbonne, France
- Date:** December 18
- Title:** Modèles hybrides intégrant des réseaux de neurones artificiels à des modèles de chaînes de Markov cachées: application à la prédiction de séries temporelles

## 7 Scientific Activities

### 7.1 Editorship

- **Name:** Prof. Hervé Bourlard  
**Function:** Editor-in-Chief  
**Journal:** Speech Communication
- **Name:** Prof. Hervé Bourlard  
**Function:** Action Editor  
**Journal:** Neural Network
- **Name:** Prof. Hervé Bourlard  
**Function:** Member of the Editorial Board  
**Journal:** Futur(e)

### 7.2 Scientific and Technical Committees

- **Name:** Prof. Hervé Bourlard  
**Function:** Member of the Advisory Board of the European Speech Technology Network
- **Name:** Prof. Hervé Bourlard  
**Function:** Member of the Board of Trustees of the Swiss Network for Innovation
- **Name:** Prof. Hervé Bourlard  
**Function:** Vontobel Bank, member of the Technical Advisory Board
- **Name:** Prof. Hervé Bourlard  
**Function:** Member of the Advisory Council of ISCA (International Speech Communication Association)
- **Name:** Prof. Hervé Bourlard  
**Function:** Member of the IEEE Technical Committee on Neural Network Signal Processing
- **Name:** Prof. Hervé Bourlard  
**Function:** Member of the Program Committee  
**Conference:** ISCA workshop of the Adaptation Methods for Speech Recognition
- **Name:** Prof. Hervé Bourlard  
**Function:** Member of the Scientific Committee  
**Conference:** Odyssey Speaker Verification Workshop 2001
- **Name:** Prof. Hervé Bourlard  
**Function:** Co-General Chairman  
**Conference:** IEEE MultiMedia Signal Processing (MMSP) workshop, Nice, 2001
- **Name:** Prof. Hervé Bourlard  
**Function:** European Liaison, and member of the Scientific Committee  
**Conference:** IEEE Neural Network for Signal Processing workshop, 2001
- **Name:** Prof. Hervé Bourlard

**Function:** General Chairman

**Conference:** IEEE Neural Network for Signal Processing workshop, Martigny, 2002

- **Name:** Prof. Hervé Bourlard

**Function:** Technical Chairman

**Conference:** IEEE Intl. Conference of Acoustics, Speech, and Signal Processing (ICASSP), 2002

- **Name:** Prof. Hervé Bourlard

**Function:** Member of the Scientific Committee

**Conference:** European Symposium of Artificial Neural Networks (ESANN)

- **Name:** Prof. Hervé Bourlard

**Function:** Member of the Scientific Committee

**Society:** International Association for Cybernetics

- **Name:** Prof. Hervé Bourlard

**Function:** Member of the Administration Committee

**Conference:** European Association for Signal Processing (EURASIP)

### 7.3 Short Term Visits

- **Location:** Université de Montréal, Montréal, Québec, Canada

**Visitor:** Samy Bengio

**Date:** from April 10 to April 14

**Location:** MacKay Institute of Communication & Neuroscience, School of Life Sciences, Keele University, U.K.

**Visitor:** Astrid Hagen

**Date:** from May 2 to May 19

- **Location:** Université de Montréal, Montréal, Québec, Canada

**Visitor:** Ronan Collobert

**Date:** from September 24 to December 23

- **Location:** ICP Grenoble, France.

**Visitor:** Hervé Glotin

**Date:** frequent visits, for a total of almost 6 months.

- **Location:** Johns Hopkins University, Research Workshop on Language Engineering, Baltimore, USA

**Visitor:** Juergen Luettin

**Date:** from July 17 to August 25

- **Location:** Johns Hopkins University, Research Workshop on Language Engineering, Baltimore, USA

**Visitor:** Hervé Glotin

**Date:** from July 17 to August 25

-

## 7.4 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:** Visit Seminar, IDSIA, Lugano, Switzerland, April 28, 2000  
**Speaker:** Hervé Bourlard  
**Title:** Research in interactive multimodal information management at IDIAP
- **Event:** AVIOS (Automatic Voice Input Output Systems) workshop, San Jose (CA, USA), May 2-24, invited keynote speaker  
**Speaker:** Hervé Bourlard  
**Title:** Speech Activities in Europe—Recent Trends
- **Event:** Visit Seminar, Bern, Switzerland, June 9, 2000  
**Speaker:** Datong Chen  
**Title:** Text Detection and Recognition In Image and Video
- **Event:** Visit Seminar, Zurich, Switzerland, June 29, 2000  
**Speaker:** Datong Chen  
**Title:** Text Detection and Recognition In Image and Video
- **Event:** Workshop on Digital Libraries, VTLS (Virginia Tech Library System), Martigny, 24 August, Plenary Talk,  
**Speaker:** Hervé Bourlard  
**Title:** Will the spoken words be used by libraries and computers?
- **Event:** Closing Day Presentation of the workshop of the Center for Language and Speech Processing, The Johns Hopkins University, Baltimore, USA - 25th of August 2000.  
**Speaker:** Hervé Glotin  
**Title:** Weight estimation for audio visual speech recognition.
- **Event:** EPFL/UNIL/IDIAP 2 days workshops, Lausanne, Switzerland, May 18-19, 2000  
**Speaker:** Mikhail Kanevski  
**Title:** Operational Geostatistics: Methods and Practice with Geostat Office
- **Event:** EPFL/UNIL/IDIAP 2 days workshops, Lausanne, Switzerland, May 25-26, 2000  
**Speaker:** Mikhail Kanevski  
**Title:** Advanced Spatial Data Analysis: Artificial Neural Networks, Support Vector Machines
- **Event:** INSA seminar, Rouen, France, June 15, 2000  
**Speaker:** Mikhail Kanevski  
**Title:** Probabilistic and Risk Mapping
- **Event:** UNIL lectures, Lausanne, Switzerland, June 28-29, 2000  
**Speaker:** Mikhail Kanevski  
**Title:** Geostatistics and Geographical Information Systems
- **Event:** MITPAN Institute seminar, Moscow, Russia, November 3, 2000  
**Speaker:** Mikhail Kanevski

- Title:** Conditional Stochastic Simulations of Spatial Data
- **Event:** IRISA Seminar, Rennes, France, March 22-24, 2000  
**Speaker:** Sacha Krstulović
- Title:** Articulatory paradigms in Automatic Speech Processing
- **Event:** IBM T. J. Watson Research Center, USA, May 19, 2000  
**Speaker:** Juergen Luettin
- Title:** Research in Multimodal Recognition at IDIAP
- **Event:** IBM Seminar, Herrenberg, Germany, July 5, 2000  
**Speaker:** Juergen Luettin
- Title:** Audio-Visuelle Spracherkennung - Computer lernen Lippenlesen
- **Event:** EUSIPCO Tutorial, Tampere, Finland, September 4, 2000  
**Speaker:** Juergen Luettin
- Title:** Audio-Visual Speech Recognition
- **Event:** University of Minnesota (MI, USA), Institute for Mathematics and its Applications, workshop on mathematical foundations of speech processing and recognition, invited speaker  
**Speaker:** Hervé Bourlard
- Title:** Hard problems in automatic speech recognition
- **Event:** IBM T. J. Watson Research Center, USA, September 20, 2000  
**Speaker:** Juergen Luettin
- Title:** Asynchronous Stream Modelling for LVCSR
- **Event:** ICP/INPG - Journee Analyses des Scenes Audiovisuelles, Grenoble, France, September 27, 2000  
**Speaker:** Juergen Luettin
- Title:** Audio-Visual Speech and Speaker Recognition
- **Event:** EURO XVII - 17th European Conference on Operational Research, Budapest, Hungary, July 16-19, 2000  
**Speaker:** Miguel Moreira
- Title:** Single pattern generation using Tabu search
- **Event:** Analyse, compression et synthese de scenes audiovisuelles, ICP/INPG, Grenoble, France, September 27, 2000  
**Speaker:** Kim Shearer
- Title:** Video annotation based on image, text and audio
- **Event:** Forum International d'Urbistique, CREM, Martigny, 2 November  
**Speaker:** Hervé Bourlard
- Title:** Les Technologies de l'Information au 21ième siècle
- **Event:** Johns Hopkins University, MA, USA, 6 November, invited talk  
**Speaker:** Hervé Bourlard

**Title:** Non-Stationary Multi-Stream Processing Towards Robust and Adaptive Speech Recognition

- **Event:** AT&T Shannon Laboratory, NJ, USA, 7 November, invited talk

**Speaker:** Hervé Bourlard

**Title:** Non-Stationary Multi-Stream Processing Towards Robust and Adaptive Speech Recognition

- **Event:** Neural Information Processing Systems, Denver (CO, USA), Nov. 27-29, invited keynote speaker

**Speaker:** Hervé Bourlard

**Title:** Avoiding tunnel vision in speech recognition research (lessons from the past and new opportunities)

- **Event:** International Computer Science Institute, Berkeley (CA, USA), 1 December

**Speaker:** Hervé Bourlard

**Title:** Non-Stationary Multi-Stream Processing Towards Robust and Adaptive Speech Recognition

- **Event:** Delos Network of Excellence on Digital Libraries, Zurich, Switzerland, 11-12 December 2000, invited keynote speaker

**Speaker:** Hervé Bourlard

**Title:** Will the spoken words be back to libraries?

## 8 Publications (1999 and 2000)

### 8.1 Books and Book Chapters

- [1] F. BEAUFAYS, H. BOURLARD, H. FRANCO, AND N. MORGAN, *Neural networks in automatic speech recognition*, 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, *Traitements de la Parole*, Presses Polytechniques Universitaires Romandes, 2000.
- [3] M. KURIMO, *Indexing audio documents by using latent semantic analysis and SOM*, in Kohonen Maps, E. Oja and S. Kaski, eds., Elsevier, 1999, pp. 363–374.
- [4] J. LUETTIN, *Speech reading*, in Modern Interface Technology: The Leading Edge, J. Noyes and M. Cooke, eds., Research Studies Press Ltd., 1999, pp. 97–121.
- [5] N. MORGAN, H. BOURLARD, AND H. HERMANSKY, *Automatic speech recognition: an auditory perspective*, in Speech Processing in the Auditory System, S. Greenberg, W. Ainsworth, A. Popper, and R. Fay, eds., Springer Verlag, New York, 2000.

### 8.2 Articles in International Journals

- [1] S. BEN-YACOUB, Y. ABDELJAOUED, AND E. MAYORAZ, *Fusion of face and speech data for person identity verification*, IEEE Transactions on Neural Networks, 10 (1999), pp. 1065–1074.
- [2] 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.
- [3] F. CAMASTRA AND A. VINCIARELLI, *Cursive character recognition by learning vector quantization*, Pattern Recognition Letters, (2001).
- [4] F. CAMASTRA AND A. VINCIARELLI, *Intrinsic dimension estimation of data: an approach based on Grassberger-Procaccia's algorithm*, Neural Processing Letters, 14 (2001).
- [5] S. DUPONT AND J. LUETTIN, *Audio-visual speech modelling for continuous speech recognition*, IEEE Transactions on Multimedia, (2000).
- [6] 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.
- [7] E. MAYORAZ, *On the complexity of recognizing regions computable by two-layered perceptrons*, Annals Mathematics and Artificial Intelligence, (1999).
- [8] A. MORRIS, A. HAGEN, H. GLOTIN, AND H. BOURLARD, *Multi-stream adaptive evidence combination for noise robust ASR*, Speech Communication, (2001).
- [9] B. NEDIC, F. BIMBOT, R. BLOUET, J.-F. BONASTRE, G. CALOZ, J. CERNOCKY, G. CHOLLET, G. DUROU, C. FREDOUILLE, D. GENOUD, G. GRAVIER, J. HENNEBERT, J. KHARROUBI, I. MAGRIN-CHAGNOLLEAU, T. MERLIN, C. MOKBEL, D. PETROVSKA, S. PIGEON, M. SECK, P. VERLINDE, AND M. ZOUHAL, *The ELISA systems for the NIST'99 evaluation in speaker detection and tracking*, DSP Journal (Special Issue on the NIST Speaker Recognition Workshop), (1999).

- [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).

### 8.3 Articles in Conference Proceedings

- [1] S. BEN-YACOUB, *Multi-modal data fusion for person authentication using SVM*, in Proc. Second International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA'99), 1999, pp. 25–30.
- [2] S. BEN-YACOUB, B. FAESL, AND J. LUETTIN, *Fast face detection using MLP and FFT*, in Proc. Second International Conference on Audio and Video-based Biometric Person Authentication (AVBPA'99), 1999, pp. 31–36.
- [3] S. BEN-YACOUB, J. LUETTIN, K. JONSSON, J. MATAS, AND J. KITTLER, *Audio-visual person verification*, in Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June 1999, Fort Collins, USA, 1999.
- [4] S. BENGIO AND J. MARIÉTHOZ, *Learning the decision function for speaker verification*, in IEEE International Conference on Acoustic, Speech, and Signal Processing, 2001.
- [5] F. BERTHOMMIER AND H. GLOTIN, *A measure of speech and pitch reliability from voicing*, in Proc. Int. Joint Conf. on Artificial Intelligence (IJCAI), F. Klassner, ed., Computational Auditory Scene Analysis (CASA) workshop, Stockholm, July 1999, Scandinavian AI Society, pp. 61–70.
- [6] F. BERTHOMMIER AND H. GLOTIN, *A new SNR-feature mapping for robust multistream speech recognition*, in Proc. Int. Congress on Phonetic Sciences (ICPhS), B. University Of California, ed., vol. 1 of XIV, San Francisco, August 1999, pp. 711–715.
- [7] 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, no IDIAP RR, see RESPITE www.
- [8] 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), 2000.
- [9] L. BESACIER, J. LUETTIN, G. MAITRE, AND E. MEURVILLE, *Experimental evaluation of text-dependent speaker verification on laboratory and field test databases in the M2VTS project*, in Proceedings of the European Conference on Speech Communication and Technology, 1999, pp. 751–754.
- [10] F. BIMBOT, M. BLOMBERG, L. BOVES, G. CHOLLET, C. JABOULET, B. JACOB, J. KHARROUBI, J. KOOLOWAAIJ, J. LINDBERG, J. MARIÉTHOZ, C. MOKBEL, AND H. MOKBEL, *An overview of the PICASSO project research activities in speaker verification for telephone applications*, in 6th european conference on speech communication and technology — eurospeech'99, vol. 5, Budapest, Hungary, September 5–10 1999, pp. 1963–1966.
- [11] H. BOURLARD, *Non-stationary multi-channel (multi-stream) processing towards robust and adaptive ASR*, in Proc. of the ESCA Workshop on Robust Methods for Speech Recognition in Adverse Conditions, 1999.
- [12] 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.

- [13] 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.
- [14] S. CHOI, Y. LYU, F. BERTHOMMIER, H. GLOTIN, AND A. CICHOCKI, *Blind separation of delayed and superimposed acoustic sources: learning algorithms an experimental study*, in Proc. IEEE Int. Conference on Speech Processing (ICSP), Seoul, September 1999, IEEE.
- [15] V. DEMYANOV, N. GILARDI, M. KANEVSKI, M. MAIGNAN, AND V. POLISHCHUK, *Decision-oriented environmental mapping with radial basis function neural networks*, in Intelligent techniques for Spatio-Temporal Data Analysis in Environmental Applications. Workshop W07, 1999, pp. 33–42.
- [16] 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, 2000.
- [17] 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, 2000.
- [18] 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.
- [19] 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 ICASSP2000 - IEEE International Conference on Acoustics, Speech, and Signal Processing, Istanbul, Turkey, June 5–9 2000.
- [20] D. GENOUD AND G. CHOLLET, *Deliberate imposture: a challenge for automatic speaker verification systems*, in Proceedings of the European Conference on Speech Communication and Technology, 1999.
- [21] N. GILARDI, M. KANEVSKI, M. MAIGNAN, AND E. MAYORAZ, *Environmental and pollution spatial data classification with support vector machines and geostatistics*, in Intelligent techniques for Spatio-Temporal Data Analysis in Environmental Applications. Workshop W07, 1999, pp. 43–51.
- [22] 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, 2000.
- [23] 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.
- [24] H. GLOTIN, F. BERTHOMMIER, AND E. TESSIER, *A CASA-labelling model using the localisation cue for robust cocktail-party speech recognition*, in Proc. European Conf. on Speech Communication and Technology (EUROSPEECH), vol. 5, september 1999, pp. 2351–2354.
- [25] A. HAGEN AND H. BOURLARD, *Using multiple time scales in the framework of multi-stream speech recognition*, in ICSLP, 2000.
- [26] A. HAGEN AND H. GLOTIN, *Etudes comparatives des robustesses au bruit de l'approche 'full combination' et de son approximation*, in Journées d'Etudes sur la Parole, Aussois, Aussois, France, Juin 2000.

- [27] 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.
- [28] A. HAGEN, A. MORRIS, AND H. BOURLARD, *Different weighting schemes in the full combination subbands approach for noise robust ASR*, in Robust Methods for Speech Recognition in Adverse Conditions, Tampere, Finland, May 1999.
- [29] A. HAGEN, A. MORRIS, AND H. BOURLARD, *From multi-band full combination to multi-stream full combination processing in robust ASR*, in ISCA ITRW ASR2000, 2000.
- [30] 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.
- [31] C. KERMORVANT AND C. MOKBEL, *Towards introducing long-term statistics in muse for robust speech recognition*, in Automatic Speech Recognition and Understanding (ASRU) workshop, Keystone, Colorado, USA, December 1999.
- [32] C. KERMORVANT AND A. MORRIS, *A comparison of two strategies for ASR in additive noise: Missing data and spectral subtraction*, in 6th European Conference on Speech Communication and Technology — Eurospeech'99, Budapest, Hungary, September, 5–10 1999.
- [33] S. KRSTULOVIC, *LPC-based inversion of the DRM articulatory model*, in Proc. Eurospeech'99, 1999.
- [34] S. KRSTULOVIC, *LPC modeling with speech production constraints*, in Proc. 5th Speech Production Seminar, 2000.
- [35] S. KRSTULOVIC, *Relating LPC modeling to a factor-based articulatory model*, in Proc. ICSLP 2000, 2000.
- [36] S. KRSTULOVIC AND F. BIMBOT, *Inverse lattice filtering of speech with adapted non-uniform delays*, in Proc. ICSLP 2000, 2000.
- [37] 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.
- [38] M. KURIMO, *Indexing spoken audio by LSA and SOMs*, in Proceedings of the European Signal Processing Conference EUSIPCO'2000, Tampere, Finland, September 2000.
- [39] M. KURIMO AND C. MOKBEL, *Latent semantic indexing by self-organizing map*, in ESCA ETRW workshop on Accessing Information in Spoken Audio, Cambridge, UK, April 1999, pp. 25–30.
- [40] J. LUETTIN AND S. BEN-YACOUB, *Robust person verification based on speech and facial images*, in Proceedings of the European Conference on Speech Communication and Technology, 1999, pp. 991–994.
- [41] 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.
- [42] J. MARIÉTHOZ, D. GENOUD, F. BIMBOT, AND C. MOKBEL, *Client / world model synchronous alignment for speaker verification*, in 6th European Conference on Speech Communication and Technology — Eurospeech'99, Budapest, Hungary, September 5–10 1999.
- [43] 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.

- [44] E. MAYORAZ AND M. MOREIRA, *Combinatorial approach for data binarization*, in Principles of Data Mining and Knowledge Discovery: third european conference; proceedings / PKDD'99, J. Zytkow and J. Rauch, eds., vol. 1704 of Lecture Notes in Artificial Intelligence, Springer, 1999, pp. 442–447.
- [45] K. MESSER, J. MATAS, J. KITTNER, J. LUETTIN, AND G. MAITRE, *XM2VTSDB: The extended M2VTS database*, in Proc. Second International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA'99), 1999.
- [46] P. MOERLAND, *Classification using localized mixtures of experts*, in Proceedings of the International Conference on Artificial Neural Networks (ICANN'99), vol. 2, London: IEE, 1999, pp. 838–843.
- [47] P. MOERLAND, *A comparison of mixture models for density estimation*, in Proceedings of the International Conference on Artificial Neural Networks (ICANN'99), vol. 1, London: IEE, 1999, pp. 25–30.
- [48] C. MOKBEL AND O. COLLIN, *Incremental enrollment of speech recognizers*, in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'99), Phoenix, Arizona, USA, 1999.
- [49] M. MOREIRA, A. HERTZ, AND E. MAYORAZ, *Data binarization by discriminant elimination*, in Proceedings of the ICML-99 Workshop: From Machine Learning to Knowledge Discovery in Databases, I. Bruha and M. Bohanec, eds., 1999, pp. 51–60.
- [50] A. MORRIS, A. HAGEN, AND H. BOURLARD, *The full combination sub-bands approach to noise robust HMM/ANN based ASR*, in 6th European Conference on Speech Communication and Technology — Eurospeech'99, Budapest, Hungary, September 5–10 1999.
- [51] 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 2000 (in press).
- [52] B. NEDIC, G. GRAVIER, J. KHARROUBI, G. CHOLLET, D. PETROVSKA, G. DUROU, F. BIMBOT, R. BLOUET, M. SECK, J.-F. BONASTRE, C. FREDOUILLE, T. MERLIN, I. MAGRIN-CHAGNOLLEAU, S. PIGEON, P. VERLINDE, AND J. CERNOCKY, *The ELISA'99 speaker recognition and tracking systems*, in IEEE Workshop on Automatic Advanced Technologies, 1999.
- [53] 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.
- [54] 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, 2000.
- [55] G. RICHARD, Y. MENGUY, I. GUI, N. SUAUDEAU, J. BOUDY, P. LOCKWOOD, C. FERNNDEZ, F. FERNNDEZ, D. GARCIA-PLAZA, C. KOTROPOULOS, A. TEFAS, I. PITAS, R. HEIMGARTNER, P. RYSER, C. BEUMIER, P. VERLINDE, S. PIGEON, G. MATAS, J. KITTNER, J. BIGÜN, Y. ABDELJAOUED, E. MEURVILLE, L. BESACIER, M. ANSORGE, G. MAITRE, J. LUETTIN, S. BEN-YACOUB, B. RUIZ, J. CORTÉS, AND K. ALDAMA, *Multi modal verification for teleservices and security applications*, in IEEE International Conference on Multimedia Computing and Systems, 1999.
- [56] 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.

- [57] M.-C. SILAGHI AND H. BOURLARD, *Iterative posterior-based keyword spotting without filler models*, in Proceedings of the IEEE Automatic Speech Recognition and Understanding (ASRU'99) Workshop, 1999.
- [58] 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.
- [59] 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 ICSLP 2000 Proceedings, Beijing, October 2000.
- [60] E. TESSIER, F. BERTHOMMIER, H. GLOTIN, AND S. CHOI, *A CASA front-end using the localisation cue for segregation and then cocktail-party speech recognition*, in Proc. IEEE Int. Conference on Speech Processing (ICSP), Seoul, September 1999, IEEE.
- [61] G. THIMM, *Tracking articulators in X-ray movies of the vocal tract*, in 8th Int. Conf. Computer Analysis of Images and Patterns, Lecture Notes in Computer Science, Springer Verlag, 1999, pp. 126–133.
- [62] G. THIMM, S. BEN-YACOUB, AND J. LUETTIN, *Evaluating the complexity of databases for person identification and verification*, in 8th Int. Conf. Computer Analysis of Images and Patterns, Lecture Notes in Computer Science, Springer Verlag, 1999, pp. 49–56.
- [63] G. THIMM AND J. LUETTIN, *Extraction of articulators in X-ray image sequences*, in Proceedings of the European Conference on Speech Communication and Technology, 1999, pp. 157–160.
- [64] G. THIMM AND J. LUETTIN, *Illumination-robust pattern matching using distorted color histograms*, in Pattern Recognition and Image Understanding, B. Radig, H. Niemann, Y. Zhuravlev, I. Gourevitch, and I. Laptev, eds., Sankt Augustin, 1999, Infix, pp. 259–266.
- [65] 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.
- [66] K. WEBER, *Multiple timescale feature combination towards robust speech recognition*, in KONVENS 2000 / Sprachkommunikation, 2000.
- [67] K. WEBER, S. BENGIO, AND H. BOURLARD, *HMM2- a novel approach to HMM emission probability estimation*, in ICSLP, 2000.

## 8.4 IDIAP Research Reports

- [1] S. BENGIO, H. BOURLARD, AND K. WEBER, *An EM algorithm for HMMs with emission distributions represented by HMMs*, IDIAP-RR 11, IDIAP, 2000.
- [2] S. BENGIO AND J. MARIÉTHOZ, *Learning the decision function for speaker verification*, IDIAP-RR 40, IDIAP, 2000.
- [3] G. BERNARDIS, H. BOURLARD, M. RAJMAN, AND J.-C. CHAPPELIER, *INtegrating SPEech acoustic and linguistic Constraints: Baseline System Development*, IDIAP-RR 21, IDIAP, 1999.
- [4] F. BIMBOT, M. BLOMBERG, L. BOVES, G. CHOLLET, C. JABOULET, B. JACOB, J. KHARROUBI, J. KOOLWAIIJ, J. LINDBERG, J. MARIÉTHOZ, C. MOKBEL, AND H. MOKBEL, *An overview of the PICASSO project research activities in speaker verification for telephone applications*, IDIAP-RR 24, IDIAP, 1999.

- [5] H. BOURLARD, *Auto-association by multilayer perceptrons and singular value decomposition*, IDIAP-RR 16, IDIAP, 2000.
- [6] R. COLLOBERT AND S. BENGIO, *On the convergence of SVM-Torch, an algorithm for large-scale regression problems*, IDIAP-RR 24, IDIAP, 2000.
- [7] R. COLLOBERT AND S. BENGIO, *Support vector machines for large-scale regression problems*, IDIAP-RR 17, IDIAP, 2000.
- [8] B. FASEL AND J. LUETTIN, *Automatic facial expression analysis: A survey*, IDIAP-RR 19, IDIAP, 1999.
- [9] 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 02, IDIAP, 2000.
- [10] H. GLOTIN, *Robust multi-stream speech recognition based on the combined reliabilities of the speech signal and phonemes estimates*, IDIAP-RR 36, IDIAP, 2000.
- [11] H. GLOTIN, D. VERGYRI, C. NETI, G. POTAMIANOS, AND J. LUETTIN, *Weighting schemes for audio-visual fusion in speech recognition*, IDIAP-RR 44, IDIAP, 2000.
- [12] E. GRAND, *Handwritten digits recognition*, IDIAP-RR 07, IDIAP, 2000.
- [13] M. KANEVSKI AND S. CANU, *Spatial data mapping with support vector regression*, IDIAP-RR 09, IDIAP, 2000.
- [14] M. KANEVSKI AND N. GILARDI, *Numerical experiments with support vector machines*, IDIAP-RR 15, IDIAP, 1999.
- [15] M. KANEVSKI, N. GILARDI, E. MAYORAZ, AND M. MAIGNAN, *Environmental spatial data classification with support vector machines*, IDIAP-RR 7, IDIAP, 1999.
- [16] K. KELLER, S. BEN-YACOUB, AND C. MOKBEL, *Combining wavelet-domain hidden Markov trees with hidden Markov models*, IDIAP-RR 14, IDIAP, 1999.
- [17] C. KERMORVANT, *A comparison of noise reduction techniques for robust speech recognition*, IDIAP-RR 10, IDIAP, 1999.
- [18] M. KURIMO, *Thematic indexing of spoken documents by using self-organizing maps*, IDIAP-RR 05, IDIAP, 2000.
- [19] J. LUETTIN, *Speaker verification experiments on the XM2VTS database*, IDIAP-RR 2, IDIAP, 1999.
- [20] S. MARCEL, *Approches génératives pour le traitement de séquences d'images: application à la reconnaissance dynamique des gestes de la main*, IDIAP-RR 45, IDIAP, 2000.
- [21] J. MARIÉTHOZ, D. GENOUD, F. BIMBOT, AND C. MOKBEL, *Client / world model synchronous alignment for speaker verification*, IDIAP-RR 23, IDIAP, 1999.
- [22] J. MARIÉTHOZ AND C. MOKBEL, *Synchronous alignment*, IDIAP-RR 06, IDIAP, 1999.
- [23] S. C. MIKHAIL KANEVSKI, PATRICK WONG, *Environmental data mapping with support vector regression and geostatistics*, IDIAP-RR 10, IDIAP, 2000.
- [24] P. MOERLAND, *Mixtures of latent variable models for density estimation and classification*, IDIAP-RR 25, IDIAP, 2000.

- [25] P. MOERLAND AND E. MAYORAZ, *Dynaboost: Combining boosted hypotheses in a dynamic way*, IDIAP-RR 9, IDIAP, 1999.
- [26] M. MOREIRA, *The use of boolean concepts in general classification contexts*, IDIAP-RR 46, IDIAP, Martigny, Switzerland, December 2000.
- [27] B. NEDIC AND H. BOURLARD, *Recent developments in speaker verification at IDIAP*, IDIAP-RR 26, IDIAP, 2000.
- [28] K. SHEARER, S. VENKATESH, AND H. BUNKE, *Video sequence matching via decision tree path following*, IDIAP-RR 12, IDIAP, 2000.
- [29] M.-C. SILAGHI AND H. BOURLARD, *Iterative posterior-based keyword spotting without filler models: Iterative viterbi decoding and one-pass approach*, IDIAP-RR 27, IDIAP, 1999.
- [30] T. A. STEPHENSON, *An introduction to Bayesian network theory and usage*, IDIAP-RR 03, IDIAP, 2000.
- [31] T. A. STEPHENSON, M. MAGIMAI DOSS, AND H. BOURLARD, *Automatic speech recognition using pitch information in dynamic Bayesian networks*, IDIAP-RR 41, IDIAP, 2000.
- [32] G. THIMM, *Segmentation of X-ray image sequences showing the vocal tract*, IDIAP-RR 1, IDIAP, January 1999.
- [33] G. THIMM, *Segmentation of X-ray image sequences showing the vocal tract (with tool documentation)*, IDIAP-RR 1, IDIAP, January 1999.
- [34] A. VINCIARELLI, *A survey on off-line cursive script recognition*, IDIAP-RR 43, IDIAP, 2000.
- [35] A. VINCIARELLI AND J. LUETTIN, *Off-line cursive script recognition based on continuous density HMM*, IDIAP-RR 25, IDIAP, 1999.
- [36] A. VINCIARELLI AND J. LUETTIN, *A new normalization technique for cursive handwritten words*, IDIAP-RR 32, IDIAP, 2000.
- [37] K. WEBER, S. BENGIO, AND H. BOURLARD, *HMM2- extraction of formant features and their use for robust ASR*, IDIAP-RR 42, IDIAP, Martigny, Switzerland, 2000.

## 8.5 IDIAP Communications

- [1] R. COLLOBERT, *Support vector machines, théorie et application*, IDIAP-Com 03, IDIAP, 2000.
- [2] H. GLOTIN, *Various adaptive weighting schemes for large vocabulary robust audio-visual ASR, with particular reference to the cocktail party effect*, IDIAP-COM 4, IDIAP, 2000.
- [3] IDIAP, *Activity report 1999*, IDIAP-COM 01, IDIAP, 2000.
- [4] A. MORRIS, *Latent variable decomposition for posteriors or likelihood based subband ASR*, IDIAP-COM 04, IDIAP, 1999.
- [5] V. SIIVOLA, *Language modeling based on neural clustering of words*, IDIAP-COM 02, IDIAP, 2000.

## 8.6 Other Documents

- [1] D. GENOUD, *Reconnaissance et Transformation de Locuteurs*, PhD thesis, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, January 1999.
- [2] P. MOERLAND, *Mixture Models for Unsupervised and Supervised Learning*, PhD thesis, École Polytechnique Fédérale de Lausanne, Computer Science Department, Lausanne, Switzerland, June 2000.
- [3] M. MOREIRA, *The use of Boolean concepts in general classification contexts*, PhD thesis, Ecole Polytechnique Federale de Lausanne, Switzerland, January 2001.