Activity Report 2002

IDIAP–Com 2003-01

February 2003

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

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

Est. 1991

INTERACTIVE
MULTIMODAL
INFORMATION
MANAGEMENT

Est. 2001

Contents

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 Confederation (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 2002, 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 **15%** of the annual budget) and the Federal Office for Education and Science (OFES) for European projects (representing about **15%** of the budget). For the first time in 2002 the National Centre of Competence in Research (IM)2, funded by the SNSF and lead by IDIAP, is another important source, about **30%** of the budget. The rest of the funding (about **10%**) comes from collaboration with industry, and one CTI (Commission for Technology and Innovation) project.

In 2002, IDIAP grew from an average of 40-45 to 60 collaborators, including permanent scientific staff, postdoctoral fellows, PhD students (around 25), system and development engineers, and short-term to medium-term visitors.

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

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

In 2002, IDIAP's activities have continued to flourish, with a reasonable growth of the number of collaborative projects and publications, together with a constant increase in the quality of the research, now recognized at the international level. For example, the number of **national and international projects** has significantly increased, and many new projects were granted or started in 2002. As of this writing, the 12 individual SNSF projects have been integrated into one global project, the (IM)2 NCCR is fully active, and 7 European (EC/OFES) projects are active at IDIAP, not counting the 6 which ended in 2002. VoiceInPack, a national project from CTI (Commission for Technology and Innovation), done in collaboration with ETHZ, and Komodo Entertainment Software, 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, 16 journal papers (compared to 14 in the previous Activity Report), 59 international conference papers (compared to 51 in the previous Activity Report), and 54 unpublished (or not yet published) internal research reports (compared to 46 in the previous Activity Report).

Thanks to the continued support of our authorities, and to our most competent personnel, motivated to the highest level, IDIAP is thus recognized as a highly sought partner in the areas they 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 2002, ces financements représentaient 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 **15%** du budget annuel) pour des projets de recherche fondamentale (étudiants doctorants) ainsi que de l'OFES pour les projets européens (représentant environ **15%** du budget). Pour la première fois en 2002, le Pôle de Recherche National (IM)2, financé par le Fonds National Suisse de la Recherche Scientifique et dirigé par l'IDIAP, représente une source supplémentaire à hauteur d'environ **30%** du budget. Le reste du financement de l'IDIAP (environ **10%**) provient de collaborations avec l'industrie et d'un projet CTI (Commission pour la Technologie et l'Innovation).

En 2002, l'IDIAP est passé d'environ 40-45 à 60 collaborateurs, composés essentiellement de chercheurs permanents, de chercheurs post-doctoraux, d'ingénieurs doctorants (environ 25), d'ingénieurs systèmes et de développement, et de visiteurs à court ou moyen terme.

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

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

En 2002, 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 2002. A ce jour, les 12 projets individuels du Fonds National Suisse de la Recherche Scientifique ont été intégrés dans un projet global, le PRN (IM)2 a atteint son rythme de croisière, et 7 projets européens (EC/OFES) sont actifs à l'IDIAP, sans compter les 6 projets terminés en 2002. VoiceInPack, un projet national de la CTI (Commission pour la Technologie et l'Innovation), en partenariat avec l'EPFZ et la socité Komodo Entertainment Software, 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, 16 articles dans des revues internationales (comparé à 14 pour le rapport d'activité précédent), 59 articles dans des conférences internationales (comparé à 51 pour le rapport d'activité précédent), et 54 rapports scientifiques internes non publiés ou en cours de publication (comparé à 46 pour le rapport d'activité précédent),.

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

General contact information:

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

Phone: +41 - 27 - 721 77 11

Fax: +41 - 27 - 721 77 12

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

In 2002, 21 new researchers joined IDIAP while only 4 left. 10 students were at IDIAP for an internship. System management and administrative staff were also strengthened.

3.1 Scientific Staff

Mr	Jitendra AJMERA Jitendra.Ajmera@idiap.ch	Research assistant +41 27 721 77 48	
Mr	Silèye BA Sileye.Ba@idiap.ch	Research assistant +41 27 721 77 61	01.10.02 →
Mr	Marc BARNARD Mark.Barnard@idiap.ch	Research assistant +41 27 721 77 29	
Dr	Samy BENGIO Samy.Bengio@idiap.ch	Machine Learning Group Leader +41 27 721 77 39	
Mr	Mohamed F. BENZEGHIBA Mohamed.Benzeghiba@idiap.ch	Research assistant +41 27 721 77 35	
Prof.	Hervé BOURLARD Herve.Bourlard.@idiap.ch	Director +41 27 721 77 20	
Mr	Fabien CARDINAUX Fabien.Cardinaux@idiap.ch	Research assistant +41 27 721 77 55	
Mr	Datong CHEN Datong.Chen@idiap.ch	Research assistant +41 27 721 77 56	
Ms	Silvia CHIAPPA Silvia.Chiappa@idiap.ch	Research assistant +41 27 721 77 30	
Mr	Ronan COLLOBERT Ronan.Collobert@idiap.ch	Research assistant +41 27 721 77 31	01.08.02 →
Mr	Christos DIMITRAKAKIS Christos.Dimitrakakis@idiap.ch	Research assistant +41 27 721 77 40	
Mr	Beat FASEL Beat.Fasel@idiap.ch	Research assistant +41 27 721 77 23	

Dr	Daniel GATICA-PEREZ Daniel.Gatica-Perez@idiap.ch	Research scientist +41 27 721 77 33	01.01.02 →
Dr	Nicolas GILARDI Nicolas.Gilardi@idiap.ch	Research assistant	→ 30.11.02
Mr	Maël GUILLEMOT Mael.Guillemot@idiap.ch	Development engineer +41 27 721 77 61	21.10.02 →
Mr	Shajith IKBAL Shajith.Ikbal@idiap.ch	Research assistant +41 27 721 77 46	
Ms	Agnès JUST Agnes.Just@idiap.ch	Research assistant +41 27 721 77 68	01.10.02 →
Prof.	Michael KANEVSKI Michael.Kanevski@idiap.ch	Research scientist	→ 31.03.02
Ms	Mikaela KELLER Mikaela.Keller@idiap.ch	Research assistant +41 27 721 77 75	01.12.02 →
Mr	Itshak LAPIDOT Itsak.Lapidot@idiap.ch	Research scientist +41 27 721 77 60	01.03.02 →
Mr	Guillaume LATHOUD Guillaume.Lathoud@idiap.ch	Research assistant +41 27 721 77 63	01.03.02 →
Mr	Quan LE Quan.Le@idiap.ch	Research assistant +41 27 721 77 36	
Dr	Vincent LEMAIRE Vincent.Lemaire@idiap.ch	Research scientist	01.02.02 → 31.07.02
Mr	Mathew MAGIMAI DOSS Mathew@idiap.ch	Research assistant +41 27 721 77 51	
Dr	Sebastien MARCEL Sebastien.Marcel@idiap.ch	Research scientist +41 27 721 77 27	
Mrs	Christine MARCEL Christine.Marcel@idiap.ch	Development engineer +41 27 721 77 50	
Mr	Johnny MARIÉTHOZ Johnny.Mariethoz@idiap.ch	Development engineer +41 27 721 77 44	
Mr	Olivier MASSON Olivier.Masson@idiap.ch	Development engineer +41 27 721 77 66	01.07.02 →
Dr	Iain MCCOWAN Iain.Mccowan@idiap.ch	Research scientist +41 27 721 77 32	
Mr	Michael MCGREEVY Michael.McGreevy@idiap.ch	Research assistant +41 27 721 77 35	15.01.03 →
Dr	José MILLAN José.Millan@idiap.ch	Research scientist +41 27 721 77 70	01.09.02 →

Mr	Hemant MISRA Hemant.Misra@idiap.ch	Research assistant +41 27 721 77 57	
Mr	Florent MONAY Florent.Monay@idiap.ch	Research assistant +41 27 721 77 69	01.08.02 →
Mr	Darren MOORE Darren.Moore@idiap.ch	Development engineer +41 27 721 77 34	07.01.02 →
Dr	Andrew MORRIS Andrew.Morris@idiap.ch	Research scientist	→ 31.12.02
Dr	Jean-Marc ODOBEZ Jean-Marc.Odobez@idiap.ch	Research scientist +41 27 721 77 26	
Mr	Norman POH HOON THIAN Norman.Poh@idiap.ch	Research assistant +41 27 721 77 53	01.09.02 →
Mr	Alexei POZDNOUKHOV Alexei.Pozdnoukhov@idiap.ch	Research assistant +41 27 721 77 65	01.01.03 →
Mr	Pedro QUELHAS Pedro.Quelhas@idiap.ch	Research assistant +41 27 721 77 74	01.11.02 →
Mr	Yann RODRIGUEZ Yann.Rodriguez@idiap.ch	Research assistant +41 27 721 77 72	01.09.02 →
Dr	Conrad SANDERSON Conrad.Sanderson@idiap.ch	Research scientist +41 27 721 77 43	01.08.02 →
Mr	Kevin SMITH Kevin.Smith@idiap.ch	Research assistant +41 27 721 77 67	21.11.02 →
Mr	Todd STEPHENSON Todd.Stephenson@idiap.ch	Research assistant +41 27 721 77 52	
Mr	Alex TRUTNEV Alex.Trutnev@idiap.ch	Research assistant +41 27 721 77 38	
Mr	Vivek TYAGY Vivek.Tyagi@idiap.ch	Research assistant +41 27 721 77 58	
Mr	Alessandro VINCIARELLI Alessandro.Vinciarelli@idiap.ch	Research assistant +41 27 721 77 24	
Mrs	Katrin WEBER Katrin.Weber@idiap.ch	Research assistant +41 27 721 77 37	
Dr	Pierre WELLNER Pierre.Wellner@idiap.ch	Research scientist +41 27 721 77 62	01.08.02 →

3.2 Students

Mr	Aïssa AIT-HASSOU Aissa.Ait-Hassou@idiap.ch	01.07.02 → 15.08.02
Mr	Jaume ESCOFET Jaume.Escofet@idiap.ch	30.01.02 → 31.07.02
Mr	François GARDIEN Francois.Gardien@idiap.ch	01.03.03 → 30.06.02
Mr	Maël GUILLEMOT Mael.Guillemot@idiap.ch	01.05.02 → 301.08.02
Mr	Guillaume HEUSCH Guillaume.Heusch@idiap.ch	05.08.02 → 11.10.02
Mr	Cédric NORMAND Cedric.Normand@idiap.ch	05.08.02 → 11.10.02
Mr	Simon PAYNE Simon.Payne@idiap.ch	01.02.02 → 31.05.02
Mr	Norman POH HOON THIAN Norman.Poh@idiap.ch	01.02.02 → 30.06.02
Mr	Alexei POZDNOUKHOV Alexei.Pozdnoukhov@idiap.ch	01.07.02 → 31.12.02
Mr	Himanshu VARSHNEY Himanshu.Varshney@idiap.ch	06.05.02 → 18.07.02

3.3 System and Development Staff

Mr	Tristan CARRON Tristan.Carron@idiap.ch	System engineer +41 27 721 77 77	01.01.03 →
Mr	Thierry COLLADO Thierry.Collado@idiap.ch	Webmaster & system engineer +41 27 721 77 42	
Mr	Norbert CRETTOL Norbert.Crettol@idiap.ch	System engineer +41 27 721 77 25	01.03.02 →
Mr	Laurent DEFAGO Laurent.Defago@idiap.ch	System engineer +41 27 721 77 25	→ 28.02.02
Mr	Frank FORMAZ Frank.Formaz@idiap.ch	System Management Group Leader +41 27 721 77 28	
Mrs	Haiyan WANG Haiyan.Wang@idiap.ch	Development engineer +41 27 721 77 54	

3.4 Administrative Staff

Ms	Rosanna BARBUTO Rosanna.Barbuto@idiap.ch	Marketing & Communications +41 27 721 77 73	01.09.02 →
Mr	Pierre DAL PONT Pierre.DalPont@idiap.ch	Financial Manager +41 27 721 77 45	
Dr	Jean-Albert FERREZ Jean-Albert.Ferrez@idiap.ch	Program Manager +41 27 721 77 19	
Mrs	Sylvie MILLIUS Sylvie.Millius@idiap.ch	Secretary +41 27 721 77 21	
Mrs	Nadine ROUSSEAU Nadine.Rousseau@idiap.ch	Secretary +41 27 721 77 22	
Mr	Michel SALAMIN	French teacher	01.09.02 →
Mrs	Joanne SCHULZ (MOORE) Joanne.Schulz@idiap.ch	HR assistant +41 27 721 77 49	01.05.02 →

4 Major events in 2002

4.1 Grand opening of the Pavillon Dalle Molle

Figure 1: The Villa Tissières

Figure 2: The Pavillon Dalle Molle

The major event of the year 2002 took place on June, 24, when the Pavillon Dalle Molle was ready to host about 40 researchers. Aside from 21 offices, the Pavillon also offers a fully equipped conference room for 60 to 70 people, with audioconferencing, and the Smart Meeting Room (see below). After a long preparation, the Pavillon was finally built in a record 16 weeks and fullfilled an urgent need for more space caused by the growth of the institute.

4.2 The first year of the (IM)2 NCCR

The (IM)2 NCCR officially started on January 1st, 2002. While the first months were still dedicated to bootstraping the various projects, hiring researchers and students, and setting up structures, the NCCR has now reached its cruising pace. The impact of the NCCR on IDIAP is obvious through the new "Pavillon Dalle Molle" building, the hiring of more than 20 new full time staff members, and the acquisition and setting up of the Smart Meeting Room. As a brief reminder, we can list some of the (IM)2-related events of 2002:

4.2.1 New or improved partnerships

The Network brought new partnerships to IDIAP or strenghtened existing ones: EPFL, ETHZ, the Universities of Geneva, Fribourg and Bern, ICSI (a formal visitor exchange agreement has been signed, two students including IDIAP's Jitendra Ajmera are now at ICSI), Eurecom, HEVs, CIMTEC, ...

4.2.2 The (IM)2 Innauguration

IDIAP, in collaboration with the City of Martigny, organised the official Inauguration of the NCCR on May, 4, 2002. The ceremony featured talks by M. Pierre Crittin, President of Martigny and of IDIAP Foundation Council, M. Jean-René Fournier, State Councilor of Valais, Prof Hervé Bourlard, Director of IDIAP and IM2, Prof Martin Hasler, Member of SNF Research Committee, Prof Stefan Catsicas, EPFL Vice-President for Research, and M. Thierry Gattlen, Director of SportAccess Kudelski SA.

4.2.3 The SNSF Review Panel Site Visit

In addition to the annual progress report, the NCCR was assessed during a two day visit of the SNSF-appointed Review Panel. The panel members are: Dr Phil Janson (SNSF, Chairman), Prof. Marco Baggiolini (SNSF), Dr Giordano Bruno Beretta (Hewlett Packard Laboratories, Palo Alto, USA), Prof. Shih-Fu Chang (Columbia University, New York, USA), Prof. Beat Hirsbrunner (SNSF), Prof. Mari Ostendorf (University of Washington, Seattle, USA), Prof. Steve Renals (University of Sheffield, UK), Prof. Gerhard Rigoll (Technische Universität München, D), and Dr Andrew William Senior (IBM T.J. Watson Research Center, Hawthorne, USA). The panel reports to SNSF, however the feedback IDIAP received was highly positive.

4.2.4 The (IM)2 Summer Institute

On October 3 and 4, the "Centre du Parc" in Martigny hosted the first internal workshop that brought together about 80 (IM)2 scientists. A similar event will be organised again in 2003.

4.2.5 The (IM)2 web site, http://www.im2.ch/

IDIAP hosts a common web site that acts as an entry point for all activities related to the NCCR. On this web site, one can also find copies of the monthly IM2 Newsletter, featuring activities of the NCCR and related news. A hard copy of this Newsletter can also be recieved by regular mail upon request to the IDIAP secretaries.

4.3 The Smart Meeting Room and the Rhonedata Multimodal Media File Server

In order to record meetings in the framework of M4 and (IM)2, IDIAP has equipped one of the rooms of the Pavillon Dalle Molle with state-of-the-art audio and video recording equipment. Currently, this allows recording of 24 audio and 3 video tracks, all fully synchronised, which are then stored and distributed through the RhoneData Multimodal Media File Server. Full details about the room and the server can be found at: http://www.idiap.ch/˜mccowan/meeting/ and http://mmm.idiap.ch/.

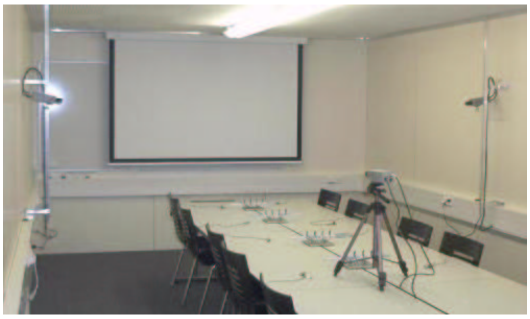


Figure 3: The IDIAP Smart Meeting Room

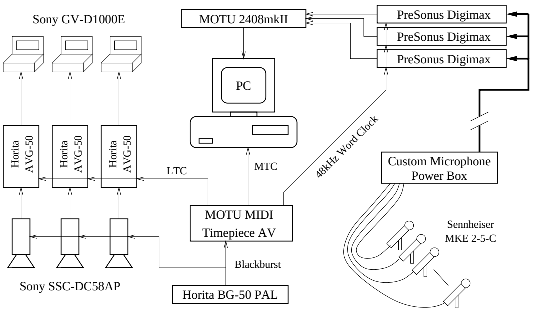


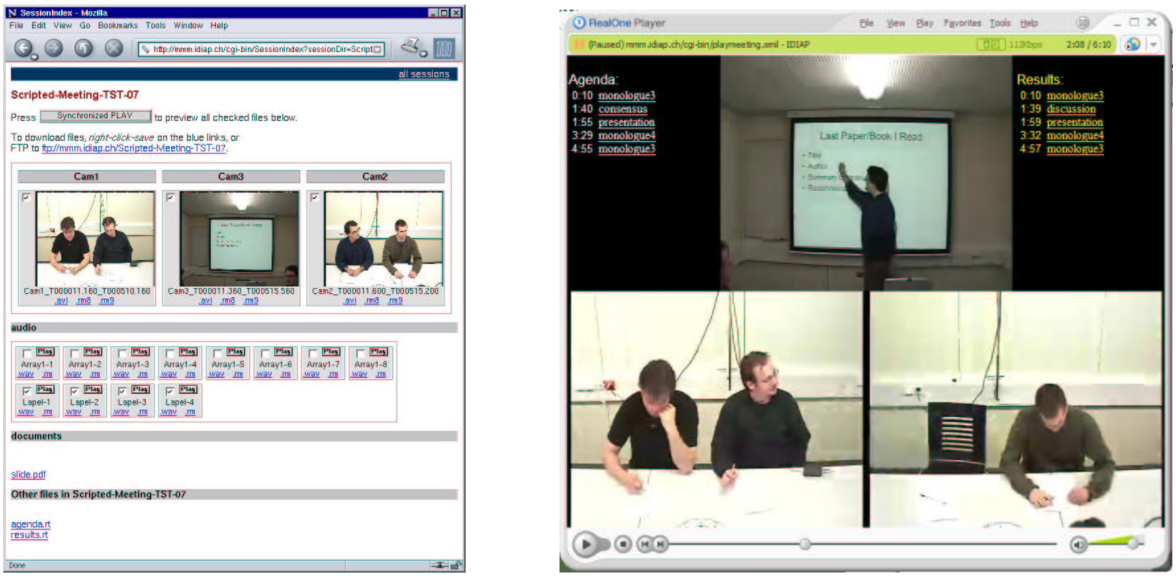
Figure 4: Hardware equipment of the IDIAP Smart Meeting Room

4.4 New regime for SNSF projects

In 2001, IDIAP was invited by SNSF to replace its 10-15 recurring small (1-2 PhD student) SNSF projects by a single global project, covering/extending all the active projects. This global project, referred to as MULTI, was submitted and accepted by the SNSF scientific committee. After synchronization of all active projects, the change to the global project took place on October 1st, 2002. While simplifying the management and reporting of research activities funded by SNSF which are not part of the NCCR, it is also expected that this opportunity will further increase the quality and continuity of the projects funded by SNSF.

4.5 Partnership with CIMTEC

In the framework of (IM)2, but also for its own needs, in June 2002 IDIAP started a very active partnership with CIMTEC (www.cimtec.ch), a leading office for business innovation and technology transfer. This collaboration has already identified approximately 10 potentially exploitable technologies, and will most likely result in 2003 in the creation of two to three new IDIAP spin-offs. The existing IDIAP spin-off, VoxAccess, moved into our offices and benefited from closer contacts with IDIAP staff.


SessionIndex - Mozilla
File Edit View Go Bookmarks Tools Window Help
http://mmm.idiap.ch/cgi-bin/SessionIndex?sessionDir=Script
all sessions

Scripted-Meeting-TST-07

Press Synchronized PLAY to preview all checked files below.

To download files, *right-click-save* on the blue links, or FTP to ftp://mmm.idiap.ch/Scripted-Meeting-TST-07.

Cam1	Cam3	Cam2
☑ Cam1_T000011.160_T000510.160 .avi .rm8 .rm9	☑ Cam3_T000011.360_T000515.560 .avi .rm8 .rm9	☑ Cam2_T000011.600_T000515.200 .avi .rm8 .rm9

audio

□ Play Array1-1 .wav .rm	□ Play Array1-2 .wav .rm	□ Play Array1-3 .wav .rm	□ Play Array1-4 .wav .rm	□ Play Array1-5 .wav .rm	□ Play Array1-6 .wav .rm	□ Play Array1-7 .wav .rm	□ Play Array1-8 .wav .rm
☑ Play Lapel-1 .wav .rm	☑ Play Lapel-2 .wav .rm	☑ Play Lapel-3 .wav .rm	☑ Play Lapel-4 .wav .rm				

documents

slide.pdf

Other files in Scripted-Meeting-TST-07

agenda.rt
results.rt

Done
RealOne Player
File View Play Favorites Tools Help
(Paused) mmm.idiap.ch/cgi-bin/playmeeting.smil - IDIAP
112Kbps
2:08 / 6:10
Last Paper/Book I Read
Title
Author
Summary
Recommend

Agenda:
0:10 monologue3
1:40 consensus
1:55 presentation
3:29 monologue4
4:55 monologue3

Results:
0:10 monologue3
1:39 discussion
1:59 presentation
3:32 monologue4
4:57 monologue3

Figure 5: (*Left*) Interface of the media file server currently resulting of Im2 and the EC-IST M4 project, see also http://mmm.idiap.ch. All audio and video files can be accessed/downloaded separately, or can be played back synchronously (*Right*). This server will be used as an initial version of the media file server to distribute and annotate the multimodal data. Interfaced with an SQL database server, it should also allow for retrieving and presenting information.

4.6 New web site

The IDIAP web site, `http://www.idiap.ch/`, which is one of the most important tools for communication, has more than 500 visits every day. The site was redeveloped in September 2002, adopting a more fashionable graphic line.

4.7 The analysis of the Ben Laden tapes

In late November, IDIAP was asked by French TV channel France 2 to analyse the latest recording of Ben Laden. Following the announcement that the recorded voice may not be that of Ben Laden, IDIAP was featured on many TV and radio stations, newpapers, online magazines, etc. The press release is available at `http://www.idiap.ch/pages/press/bin-laden-eval.pdf`.

4.8 IDIAP Research Committee

IDIAP has set up an internal Research Committee, responsible for evaluating internal reseach proposals in the framework of (IM)2 and the SNSF MULTI project. This Committee is also responsible for evaluating potential candidates for a scientific position at IDIAP.

4.9 Best PhD Student Award

In 2002, IDIAP insitituted the annual PhD award, based on a selective process taking into account the quality of the scientific research, inter-project collaboration, and the candidate's social qualities. This award will be given every year to the student who has the best scientific and social activity, focusing especially on the collaboration with other IDIAP students and seniors. The selection for the prize is made through nomination and selections by all seniors and postdocs. Alessandro Vinciarelli was the first winner of the annual IDIAP Best PhD Student award.

4.10 IEEE Neural Networks for Signal Processing, NNSP'02

In September 2002, IDIAP organized in Martigny the 2002 IEEE International Workshop on Neural Networks for Signal Processing (NNSP02) (details at http://eivind.imm.dtu.dk/nnsp2002/). Some 150 researchers from all around the world came to Martigny for this important event.

4.11 Eurospeech'03

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

4.12 European project meetings

Several of the European projects currently active at IDIAP have held technical meetings in Martigny: M4 on 29-30.08.2002, FgNet on 12-13.09.2002, LAVA on 20-21.3.2003.

Furthermore, on June 5, IDIAP (together with HEVs, ECAL and IUKB) took part in the European-wide celebration of the 10th anniversary of the ERASMUS student exchange program.

4.13 TAM: Tuesday Afternoon Meetings

With more than 50 researchers, almost half of them at IDIAP for less than a year, there was a growing need to strengthen mutual awareness. Thanks to the new conference room, all staff are now able to meet on Tuesday afternoon for a presentation (followed by a short discussion) by one of IDIAP's scientists. and one of the IDIAP scientist presents his or her work, followed by a short discussion. This gives everyone, and in particular students, the opportunity to present and discuss their activity approximately once a year. The TAMs also provide an opportunity for general announcements regarding life at IDIAP.

4.14 Lectures on Statistical Machine Learning

In October 2002, Dr Samy Bengio started to give a series of weekly lectures on statistical machine learning.

4.15 Medienseminar BTF in Bern

On April, 18th, IDIAP and (IM)2 representatives were in Bern for the mid-term presentation of the 2000-2003 message from the Federal Council on the encouragement of teaching, research and technology. There were speeches from Federal Councilors R. Dreyfuss and P. Couchepin, and various other representatives. Actual examples of the current Swiss strategy were presented. SNF chose (IM)2 to demonstrate the NCCR concept.

4.16 Important visits at IDIAP

Several important bodies have visited IDIAP in 2002:

- The Control Committee (*Commission de Gestion - Geschäftsprüfungskommission*) of the Council of States (*Conseil des Etats - Ständerat*).
- A delegation of the Swiss Science and Technology Council
- A delegation of the *Conseil des Ecoles Polytechniques Fédérales*, (CEPF, ETHRat)
- The Director of Economic Affairs of the Canton du Valais, Léonard Favre.

5 Research Activities

The focus of our activities is the development of advanced (multimodal) natural input and output computer interfaces through speech and vision, and of 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 follows:

- **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 a computer interface (on-line recognition). **Research issues** include: analysis of documents with a complex layout, recognition of degraded printed text, recognition of running handwriting.

- **Visual input**: Shape tracking (including lip tracking, face tracking); gesture recognition; facial expressions; images (e.g., sketches, signatures, photos) used as input. **Research issues** include: robustness of the algorithms; combination of colour, motion, texture, and shape in the analysis; more accurate model-based analysis; computational complexity.

- **Input (spoken, written, visual) analysis and understanding**, involving parsing and syntactic and semantic analysis and modeling. **Research issues** include: specification and formalism of unimodal and multimodal syntactical and semantic constraints, using these constraints into unimodal and multimodal input signal processing, merging modalities through multimodal "grammars".

- **Protecting information access**, involving: speaker verification, signature recognition, face recognition; bio-metric (multimodal) user authentication. **Research issues** include: increasing robustness of user authentication techniques, multimodal user authentication (mixture of experts, confidence-based weighting of the different media, etc).

- **Modality integration**, involving, e.g.: Speech and gestures, facial movement and speech recognition, facial movement and speech synthesis, and interface agents. **Research issues** include: merging of different (media) data streams, possibly non-synchronous and with different data rate, fusion of the different modalities (e.g., based on signal-to-noise ratio or confidence level estimation).

- **Mathematical methods**, including: Statistical modeling and statistical pattern classification, signal processing techniques, connectionist techniques, expert fusion, support vector machines.

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

5.1 Speech Processing Group

Figure 6: Some members of the IDIAP 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.

The IDIAP Speech Processing group is currently involved in numerous European, Swiss National Science Foundation, and DARPA projects, for example:

- MultiModal Meeting Manager (M4), from the EC/IST 5th Framework Program, http://www.dcs.shef.ac.uk/spandh/projects/m4/index.html. This project is concerned with the construction of a demonstrable system to enable structuring, browsing and querying of an archive of automatically analysed meetings. The archive will have been created in the Smart Meeting Room installed at IDIAP, equipped with multimodal sensors. See the annual report at http://www.dcs.shef.ac.uk/spandh/projects/m4/M4-AnnualReport2002/.
- Hearing, Organization and Recognition of Speech in Europe (HOARSE, http://www.hoarsenet.org)
- DARPA EARS (Effective Affordable Reusable Speech-to-text), see http://www.darpa.mil/iao/EARS.htm.
- Speaker source localization, microphone arrays and beamforming.

The IDIAP Speech Group is also significantly contributing to the Swiss National Center of Competence in Research (NCCR) IM2 (see http://www.im2.ch) on Interactive Multimodal Information Management, through “Individual Projects” IM2.SP (www.im2.ch/SP.php) and IM2.ACP (www.im2.ch/ACP.php).

5.1.1 Research Themes

The research areas of the Speech Processing group currently focus on:

- Automatic recognition of (isolated, continuous, or natural) speech based on phonetic (sub-word) modeling, using spectral-temporal profiles of speech, as well as articulatory.
- Development and improvement of state-of-the-art speech recognition systems based on hidden Markov models (HMM).

- Speaker verification: development and improvement of state-of-the-art text-dependent, text-independent, and user-customized speaker verification systems.
- Using discriminant artificial neural networks (ANN) to estimate a posteriori probabilities. In this regard, IDIAP (in collaboration with ICSI, Berkeley, http://www.icsi.berkeley.edu) is recognized as a leader in the use of hybrid HMM/ANN systems, exhibiting several advantages compared to standard HMM approaches.
- 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 detecting out-of-vocabulary words is also investigated.
- 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 the 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.
- Multi-stream and multi-channel 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.
- 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.
- 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).
- Statistical language modelling: Extending current language models to better cope with natural speech, out-of-vocabulary word, and word classes.
- Speaker adaptation: Improving recognition accuracy by automatically adapting (a subset of) the parameters of the recognition system.
- Development and adaptation of efficient software for large vocabulary continuous speech recognition, on different computer platforms (mainly UNIX and Windows NT), all compatible with the TORCH (http://www.torch.ch) libraries developed at IDIAP.
- Speaker source localization, microphone arrays and beamforming, as illustrated in Figure 7.
- Development and testing of applications prototypes.

5.1.2 Application Examples

1. Command and control systems, possibly used in noisy environments, e.g., to operate a speech enabled cellular phone in cars. See, e.g., the RESPITE and SPHEAR projects.
2. Speech enabled information systems: Building speech-enabled kiosks, desk tablets, and personal data assistants to enable users to find and display current information.
3. Information retrieval for audio documents: Using transcriptions automatically generated by a large-vocabulary speech recogniser to build indexes that can be queried by information retrieval engines for searchable audio archives. See, e.g., the ASSAVID (see Figure 8) and CIMWOS projects, allowing for:

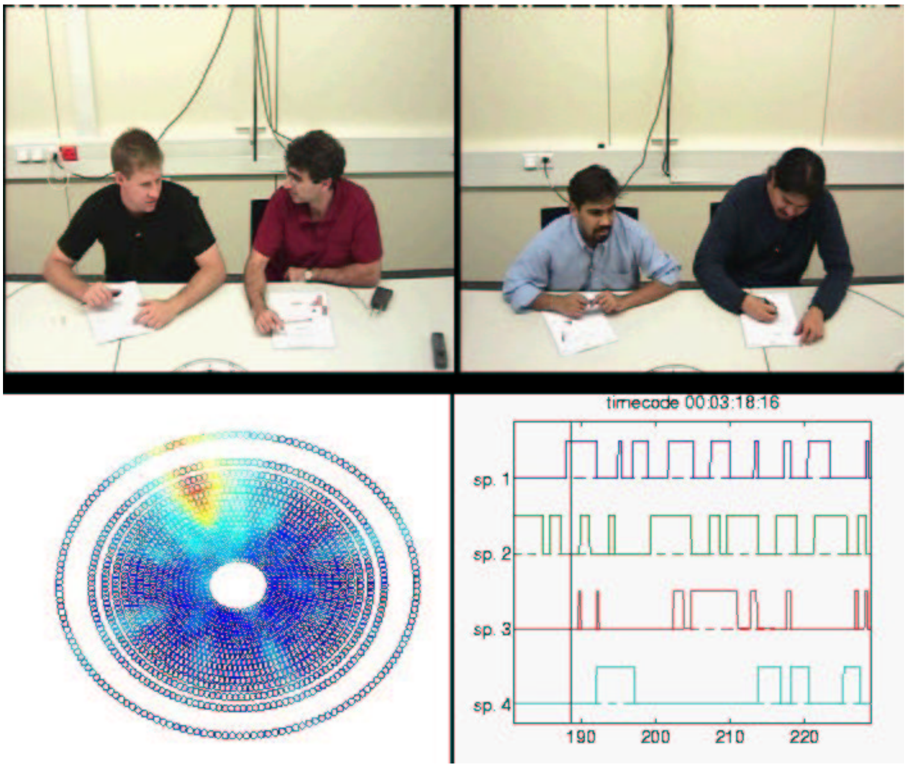


Figure 7: Illustration of the audio tracking and speaker beamforming achieved by using the microphone arrays installed in the IDIAP smart meeting room.

  - Automatic transcription of broadcast speech by an automatic speech recognition system
  - Automatic indexing of the generated audio archives
  - Content-based retrieval from typed or spoken input queries.

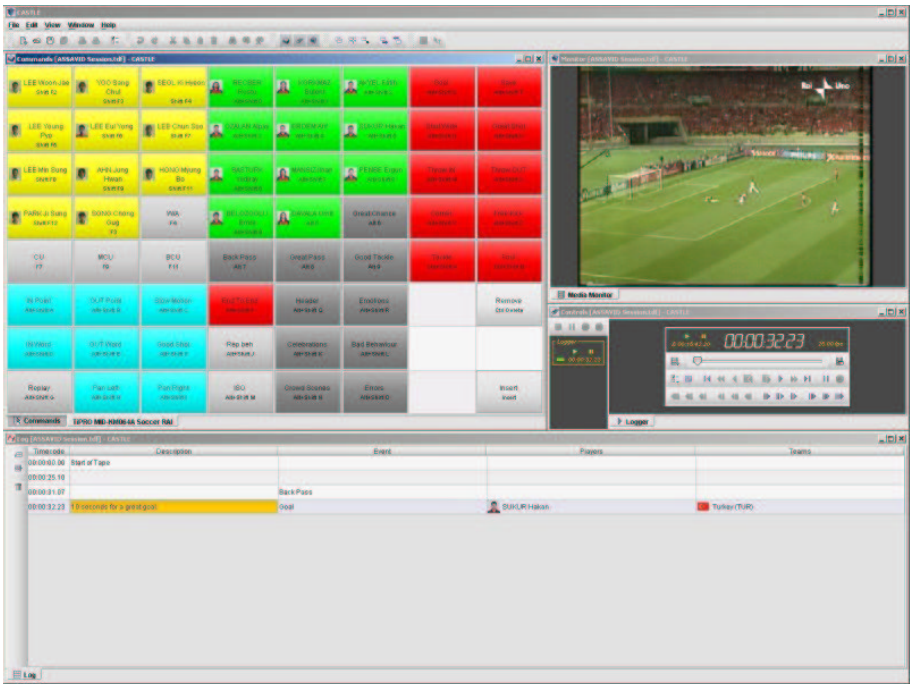
4. Automatic meeting manager: processing of multiple audio (and video) streams for structuring, browsing and querying of an archive of automatically analysed meetings. See, e.g., the M4 project or a typical meeting browser.

5.2 Computer Vision Group

The computer vision group at IDIAP investigates and develops principled methods and algorithms for analysis of visual and multimedia data, and addresses a number of specific problems including people detection and tracking, gesture and facial expression recognition, handwriting recognition, and multimedia content analysis. Their work frequently involves collaboration with the two other groups at IDIAP, speech processing and machine learning, as complementary expertise is needed for many of the research problems. The group is active in all of their areas of expertise under a number of collaborative European and Swiss national projects.

5.2.1 Research Themes

1. Handwriting recognition: Offline handwriting recognition is the automatic transcription of handwritten data when only its image is available. Members of the group have worked on methods to improve modeling of handwritten data, and have developed a recognizer based on continuous density HMMs which can deal with single words as well as handwritten texts (with the help of Statistical Language Models).



Timecode	Description	Event	Players	Teams
00:00:00.00	Start of Tape			
00:00:25.10				
00:00:31.07		Back Pass		
00:00:32.23	10 seconds for a great goal	Goal	SUKUR Hakan	Turkey (TUR)

Figure 8: Interface of the audio-video indexing and retrieval system developed in the framework of ASSAVID.

Figure 9: Some members of the IDIAP Computer Vision Group

2. Face Algorithms: Face algorithms can be divided into four different areas.

   - Face detection: The goal of face detection is to identify and locate human faces in images at different positions, scales, orientations and lighting conditions.
   - Face localization: Face localization is a simplified face detection problem with the assumption that the image contains only one face.
   - Face verification: Face verification is concerned with validating a claimed identity based on the image of its face, and either accepting or rejecting the identity claim.
   - Face recognition: The goal of face recognition is to identify a person based on the image of its face. This face image has to be compared with all registered persons. Therefore, face recognition is computationally expensive with respect to the number of registered persons.

The group is mainly interested in face detection and verification using neural networks, SVM based methods and boosted weak classifiers (see Figure 10).

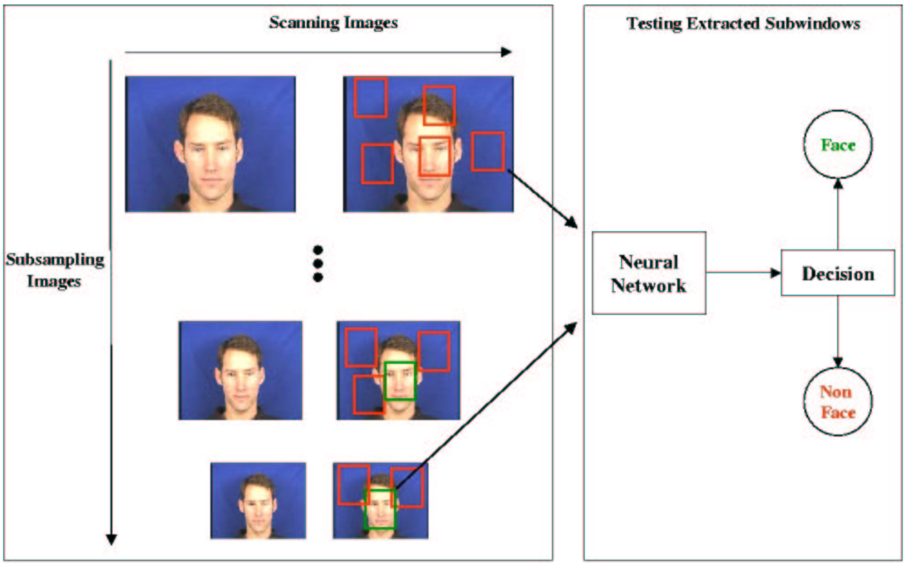


Figure 10: Face detection in an image using several subsampling stages.

3. Gesture Recognition: Gestural interaction based on the image is the most natural method for the construction of advanced man-machine interfaces. Thus, machines would be easier to use by associating the gestural command with the vocal command. This includes recognition of gestures such as facial expressions, hand postures, hand gestures and body postures. Current work on facial expression recognition is based on convolutional neural networks. Statistical approaches (skin color blobs) for object segmentation (faces and hands) in color images are investigated. The vision group is also interested in gesture recognition using hybrid models (Hidden Markov Models and Neural Networks) such as Input/Output Hidden Markov Models.

4. Tracking and activity recognition: Object tracking represents an essential component of gesture recognition, human behavior monitoring, and video indexing. IDIAP is investigating the design of stable trackers that are robust against ambiguities, image measurements, changes in the acquisition setting, and object intra-class variability. The group focuses on two areas: (1) the development of sequential Monte Carlo (SMC) techniques, and the combination of SMC and finite state motion models based on HMMs for joint tracking and recognition of people activity, and (2) the fusion of multiple visual and multimodal (audio-visual) features, for example for speaker tracking.

5. Multimedia content analysis: The vision group is developing statistical models, algorithms and tools to automatically extract relevant information from audio-visual streams, which can be used for structuring, annotating, indexing and retrieving multimedia databases. Some of the current research directions include:

   - Media structuring: The structure of videos is needed both at the individual and at the database levels. On one hand, finding structure in individual videos (shots, scenes) is useful to generate video summaries for browsing and retrieval, and usually constitutes the starting point to extract higher-level information. On the other hand, structuring a whole video database is useful for access and filtering (locating video replicas, organizing by “visual topic”, etc.).
   - Event classification: The group is developing audio-visual feature extraction and data fusion algorithms for event classification in sports video and meeting databases. Current efforts have been

directed to define semantically meaningful events, and to learn their statistical models for classification.

- 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 exploit the temporal redundancy to fuse recognition results of the same text obtained at different times.
- Modeling of textual and visual features: Members of the group investigate joint statistical models of words and visual features in multimedia databases, to relate low-level visual information with semantics. Such models would allow for important information retrieval functionalities, like clustering (grouping images that refer to the same text topics), annotation (attaching words to visual content), and illustration (attaching images to words).

5.2.2 Application Examples

Three typical applications of the methods developed at IDIAP are the following:

- Hand drawn character and cursive writing recognition is useful for such tasks as automated address reading for postal services, and interfaces for 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.

- Identity verification is a general task for security applications like access control, transaction authentication (in telephone banking or remote credit card purchases), voice mail, or secure teleworking. Face detection is a fundamental step before the verification procedure. Its reliability and time-response have a major influence on the performance and usability of the whole face verification system.

- 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 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.

5.3 Machine Learning Group

Figure 11: Some members of the IDIAP Machine Learning Group

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

5.3.1 Research Themes

1. Large scale data analysis: most actual powerful machine learning algorithms have been used for medium scale datasets: less than one hundred features describing one example and less than ten thousand example 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 algorithms 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 algorithms 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. Brain-Computer Interfaces: Using specially design helmets, EEG signals of a patient can be recorded and analyzed by advanced machine learning techniques, in order to extract corresponding commands uttered mentally by the patient (such as "left", "right", etc). Both high-level and low-level processing of these very noisy sequences are taken into account, from simple FFTs to Hidden Markov Models.

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

5.3.2 Application Examples

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

- Data Mining: how to extract interesting information from huge database wharehouses (for instance, churn detection, client modeling and prediction).
- Finance and Economy: financial portfolio management, asset prediction, portfolio selection, auction analysis.
- Pattern Recognition: handwritten character recognition, speech recognition, face detection.
- Biological Sequence Analysis: classification of DNA or RNA sequences.

6 Current Projects

ASSAVID – Automatic Segmentation and Semantic Annotation of Sports Videos

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

Duration: February 2000 – July 2002

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

Contact persons: Sebastien Marcel, Iain McCowan, Jean-Marc Odobez, Mark Barnard, Jitendra Ajmera, Datong Chen

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

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

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

AudioSkim – Automatic Segmentation of Large Audio and Multimedia Documents

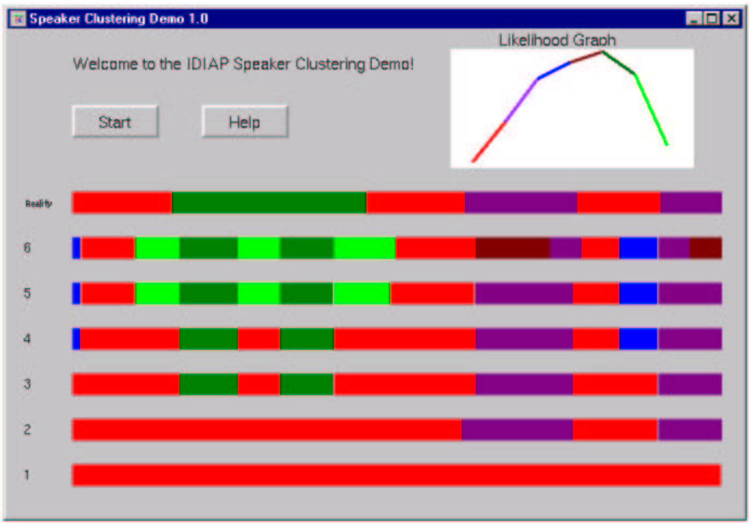
Funding: Swiss National Science Foundation

Duration: March 2001 - September 2002

Contact persons: Jitendra Ajmera, Iain McCowan, Hervé Bourlard

Description: The problem of distinguishing speech signals from other audio signals (e.g., music) has become increasingly important as automatic speech recognition (ASR) systems are applied to more and more real-world multimedia domains. Furthermore, audio and speech segmentation will always be needed to break and structure the continuous audio stream into manageable chunks applicable to the configuration of the ASR system.

This project thus aims at developing and testing on large audio databases (possibly as part of multimedia databases) such as broadcast news and sport videos, different approaches towards automatic segmentation of (multimedia) sound tracks, including, among others, changes in acoustic environments, speaker



Speaker Clustering Demo 1.0

Welcome to the IDIAP Speaker Clustering Demo!

Figure 12: Interface and illustration of the IDIAP unsupervised speaker segmentation algorithm. The first row gives the true speaker segmentation (one color per speaker). The second row gives the optimal segmentation for a given, over-estimated, number of clusters, while the upper right plot represents the likelihood resulting of the segmentation. At each iteration, the number of clusters and their parameters are re-estimated, until the global likelhood reaches an optimum value.

change detection, speaker identification and tracking, and speech/music discrimination. During its first year, this project resulted in an automatic system allowing for online segmentation of an audio signal into speech/non-speech segments and which is apparently outperforming other state-of-the-art approaches (see previous activity report). In 2002, as illustrated by Figure 12, the main emphasis of AudioSkim has been put on the automatic (unsupervised) speaker clustering and speaker turn detection. In this framework, a threshold free (BIC like) algorithm was tested and evaluated on big database (Hub-4 97 evaluation set). The results obtained are comparable to the best results when BIC is used with optimal (manually optimized, database dependent) threshold/penalty term. This will be a key development in many applications, including: meeting data segmentation and indexing, multimedia database segmentation and indexing, etc. This has led to several publication; see, e.g., IDIAP Research Reports 02-39 and 02-23.

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

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

Duration: February 2000 – May 2003

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

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

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

Funding: Swiss National Science Foundation

Duration: March 1999 - September 2002

Contact persons: Todd Stephenson, Andrew Morris, Hervé Bourlard

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

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

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

CARTANN – Cartography by Artificial Neural Networks

Funding: Swiss National Science Foundation

Duration: January 1999 - January 2003

Partners: Lausanne University (prof. Michel Maignan)

Contact persons: Nicolas Gilardi, Mikhael Kanevski, Samy Bengio

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

- highly non stationary spatial processes,
- cartography of distribution functions, as opposed to cartography of the mean value,
- user and data-driven parameterization for the discrimination between a stochastic trend and auto-correlated residuals,
- cartography of stochastic deviations related to advection-diffusion models.

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

CIMWOS – Combined IMages and WOrd Spotting

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

Duration: April 2001 - October 2003

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

Contact persons: Iain McCowan, Jean-Marc Odobez, Jitendra Ajmera, Hervé Bourlard

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

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

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

Duration: June 2001 - May 2005

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

Contact persons: Samy Bengio, Hervé Bourlard

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

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

The secondary objectives are as follows:

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

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

COST 278 – Spoken Language Interaction in Telecommunication

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

Duration: June 2001 - May 2005

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

Contact persons: Hervé Bourlard, Sébastien Marcel

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

1. To improve the knowledge of the issues and problems involved in general in spoken language interaction in telecommunication.
2. To achieve knowledge of issues related to robustness and multi-linguality within spoken language processing.
3. To achieve knowledge of spoken language interaction in the context of multi-modal communication.
4. To achieve knowledge of human-computer dialogue theories, models and systems and associated tools for the establishment of such systems.
5. To achieve knowledge of and evaluate telecommunication applications that apply spoken language as one out of more input or output modalities.

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

Divide and Learn – Improved Learning for Large Classification Problems

Funding: Swiss National Science Foundation

Duration: October 2000 - September 2002

Partners: Swiss Federal Institute of Technology (EPFL)

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

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

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

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

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

Funding: European project, INTAS foundation

Duration: June 2000 – June 2002

Contact persons: Samy Bengio, Mikhail Kanevski

Description: To support the ongoing effort to develop indicators for environmentally sustainable development, there is a real need for research to enhance the development of technologies which contribute to the maintenance of environmental quality (water, air, soil). The first step of such a research program consists of 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.

EARS - Effective Affordable Reusable Speech-to-text

Funding: DARPA - US

Duration: July 2002 - June 2007

Contact persons: Hervé Bourlard

Description: As part of the DARPA EARS (Effective Affordable Reusable Speech-to-text) program, and in collaboration with ICSI/Berkeley, SRI, the University of Washington, and Columbia University, we are working towards significanlty improving speech recognition in a project refered to as "Pushing the Envelope - Aside" where we are studying both replacements of the standard spectral envelope as the speech representation of choice (typically with cepstral transformation). This includes work on the acoustic "front end", but also includes research on statistical modeling for the new features that are being generated.

FaceX – Facial Expression Recognition through Temporal and Appearance Based Models

Funding: Swiss National Science Foundation

Duration: October 1998 – September 2002

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

Contact person: Beat Fasel

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

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

FGnet – Face and Gesture Recognition Working Group

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

Duration: 36 months, September 2001-August 2004

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

Contact person: Sebastien Marcel

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

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

GHOST - Gesturing Hand recOgition baSed on user Tracking

Funding: France Telecom R&D

Duration: 24 months, February 2002-February 2004

Partners: Telecommunication and Neural Techniques group of France Telecom R&D DTL/TIC

Contact person: Sébastien Marcel

Description: The aim of the project is thus to recognize up to 15-20 hand gestures. It is necessary to distinguish two aspects of hand gestures :

- the static aspect is, for instance, characterized by a posture of the hand in an image,
- the dynamic aspect is defined either by the trajectory of the hand, or by the sequence of hand postures in a sequence of images.

In this project hand gestures are represented by trajectories of the hand in 3D. The hand gesture database is provided by France Telecom R&D and is acquired using a stereo camera framework. Our work has as an ambition to develop hybrid techniques of statistical training (Hidden Markov Models and Neural Networks) for the recognition of hand gestures (pointing gestures or drawing gestures).

HOARSE – Hearing Organisation and Recognition of Speech in Europe

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

Duration: 48 months, September 2002-August 2006

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

Contact person: Hervé Bourlard

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

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

HMM2 – A New Framework for Robust and Adaptive Speech Recognition

Funding: Swiss National Science Foundation

Duration: October 2000 - September 2002

Contact persons: Ikbal Shajith, Hervé Bourlard

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

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

HPVWI - HP Visual Web Initiative

Funding: HP

Duration: January 2002 - December 2002

Contact persons: Hervé Bourlard, Samy Bengio

Description: This project is part of the HP philanthropy programme, and is concerned with user authentication based on biometric modalities. Amongst various biometric technologies, the ones based on facial features (face verification) and voiceprint (speaker verification) are considered here. Software dealing with these modalities has already been developed based on a number of distinct approaches in the speech and vision groups at IDIAP. The present project mainly aims at integrating, demonstrating and testing effective methods for the joint use of these modalities to facilitate, protect and promote various services requiring secured transactions.

The present project is developing an integrated demonstration system, illustrated with figures and teaching material, of a multimodal user authentication system. This system will be developed on the basis of the IDIAP TORCH software platform recently developed at IDIAP.

IM2.ACP - Access and Content Protection

Funding: SNSF, through the (IM)2 NCCR

Duration: January 2002 – December 2003

Contact persons: Norman Poh, Conrad Sanderson, Samy Bengio, Hervé Bourlard

Description: In the framework of IM2.ACP, IDIAP is mainly developing new text-dependent (user-customized) and text-independent speaker verification systems. IDIAP is also investigating advanced multimodal biometric identification/verification systems, trypically based on voice and face verification, and involving different fusion algorithms.

Experiments on scalability of speaker verification and fusion algorithms have been performed in order to verify how the system degrades when the size of the model needs to be small. The results are very interesting:

- the performance of the speaker verification system alone degrades slowly with respect to the reduction of the number of parameters (logarithmically at the beginning, hence we can remove a lot of parameters without a big performance loss)
- the performance of the fusion system degrades even more slowly than the speaker verification system, since the degradation of unimodality systems (speaker and face systems) are independent from each other.

These experiments were carried out on the English BANCA database and its associated protocol. Further experiments on variations on the creation and use of World Models for text independent speaker verification have shown promising results with respect to cross-gender attacks. These experiments were carried out on the NIST and PolyVar databases and novel protocols were designed to verify specific attacks.

IM2.LH - Leading House projects

Funding: SNSF, through the (IM)2 NCCR

Duration: January 2002 – December 2005

Contact persons: Pierre Wellner, Hervé Bourlard

Description: The general goal of the Leading House project is to leverage on the research and development efforts available at IDIAP, and to make sure that this will directly benefit all the IM2 partners. IM2.LH is also responsible for consolidating as much as possible all of the IM2 outcomes in common projects and applications. IM2.LH is also responsible for initiating new research areas or for covering researc and developments required by IM2 (but currently missing).

During the first year of IM2, the majority of time spent on the Leading House project was spent on:

1. Setting up a fully operational Smart Meeting Room, and integrating and developing the necessary software.

2. Development of audio software related to microphone arrays. In the framework of IM2, a new research direction at IDIAP over the past year has been the use of microphone arrays. Microphone arrays use directional discrimination to permit distant, hands-free signal acquisition. In this way, the microphone arrays can provide both speech enhancement (potentially for subsequent recognition), and also estimation of speaker location.
3. Building some initial demonstration systems.
4. Integrating some initial forms of the IM2 technologies.

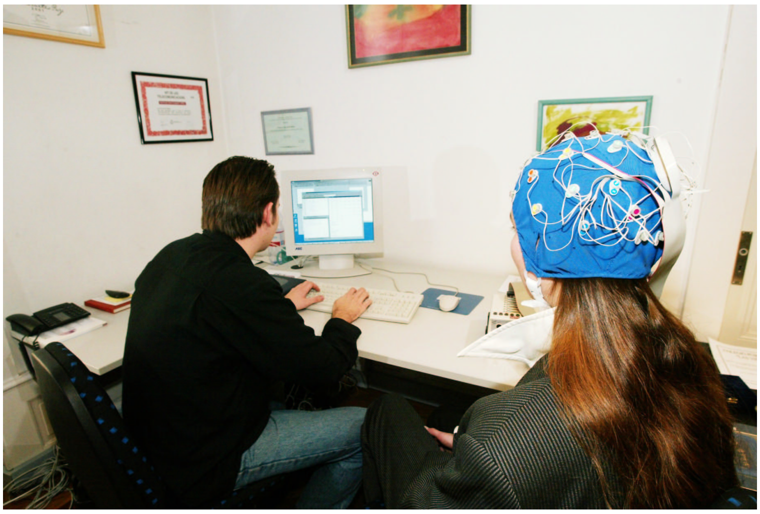


Figure 13: IDIAP brain-machine interface.

IM2.BMI - Brain Machine Interfaces

Funding: SNSF, through the (IM)2 NCCR

Duration: July 2002 – June 2004

Contact persons: Silvia Chiappa, Josè del R. Millàn, Hervé Bourlard

Description: In the framework of IM2, a project on Brain Machine Interface has also been defined. Addressing a particular kind of (advanced) man-machine interface, this project aims to investigate the possibility of classifying spontaneous brain activity based on either reconstructed brain activity maps, or directly from EEG recordings (thus, a particular kind of multi-channel processing).

During the first few months of this project, the following has been achieved:

- A communication infrastructure was created in order to facilitate the relations between the partners of the consortium. An email list as well as a protected website were created and are now maintained at IDIAP.
- The experimental system has been set up (see Figure 13), and an initial set of data has been recorded and is currently being analyzed by the partners.
- State-of-the-art survey: At least 20 identified research groups are working on Brain- Computer/Machine Interfaces (BCI or BMI). Currently, it is possible to classify from 2 to 10 different mental states, within a precision ranging from 80% to 100%. This means a bit rate of 1 to 100 bits/min, which is 10 to 100 times lower than keyboard bit rate.

IM2.MI - Multimodal Integration

Funding: SNSF, through the (IM)2 NCCR

Duration: January 2002 – December 2003

Contact persons: Mikaela Keller, Samy Bengio, Hervé Bourlard

Description: The goals of IM2.MI are the research and development of principled methods for the fusion and efficient decoding of different input modalities (multi-channel processing). The objectives can thus be decomposed as follows:

- Development of new multi-channel, multi-rate, signal processing techniques (including EEG processing)
- Development of new data fusion algorithms, as well as new decision strategies
- Development of a multichannel statistical model for the combination of asynchronous input streams
- Development of an efficient multimodal decoder
- Implementation of all these algorithms into a common software platform.

During the first year of this project, IDIAP started working on a new HMM model to handle asynchronous streams of data. This model is mainly inspired by two other models, namely:

- Asynchronous Input/Output Hidden Markov Models (AIOHMM), which enables modelling of an output sequence conditioned on an input sequence asynchronously, and
- Multi-stream HMMs.

In a recent paper (part of a special ICASSP session; see also IDIAP Research Report 02-59), the above approach has been used to analyse multimodal meeting behaviors. In this case, multiple streams containing audio and video information related to all participants of a meeting were merged in order to decode the general behavior of the meeting. The possible behaviors follow a language model with such events as monologues, discussions, presentations, consensus, disagreements, etc. The overall objective of the project is to summarize a new meeting according to this language. Several small meetings were recorded in order to train the model. Preliminary experiments show that such a multimodal decoder can obtain up to 80% event recognition on new meetings involving persons that were not present in any of the training meetings. Further experiments involving asynchronous HMMs will be performed soon.

IM2.MUCATAR - Multiple Camera Tracking and Activity Recognition

Funding: SNSF, through the (IM)2 NCCR

Duration: July 2002 – June 2004

Contact persons: Sileye Ba, Kevin Smith, Jean-Marc Odobez, Daniel Gatica-Perez

Description: In the context of IM2, human tracking and event/activity recognition represent essential components towards multimodal interaction and analysis of multimedia databases. The understanding of human activities in indoor environments is important, both as a component of a multimodal interface, and to extract semantic clues from videos for indexing and retrieval. The main research problems that will be addressed in the project are :

- the development of exemplar-based models of typical people activity that can be constructed directly from training sets in a probabilistic setting;

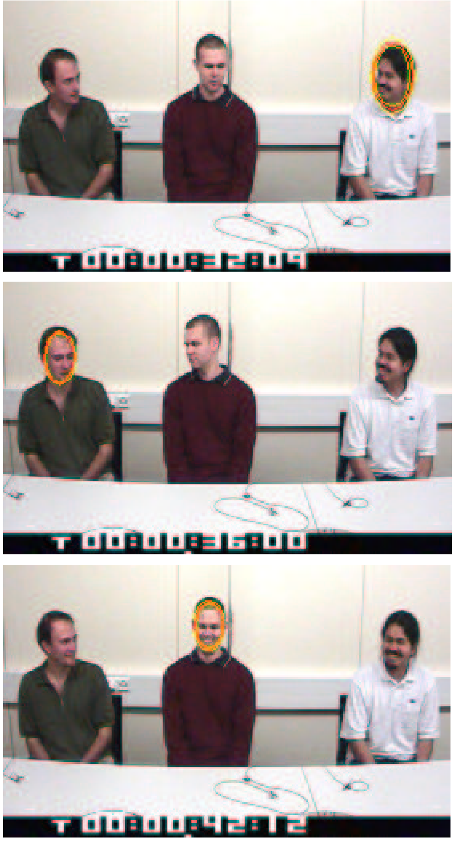


Figure 14: People tracking in the IDIAP smart meeting room.

- the combination of multiple visual features for tracking;
- the study of the combination of Sequential Monte Carlo and Hidden Markov Models for performing jointly the tracking and recognition tasks;
- the extension of such formulations to a multiple camera scenario.

The project started with the study and the implementation of standard particle filters (PF), with shape and color object models. After, we explored the use of Importance PF to perform the assymmetrical integration of audio-visual information in a way that efficiently exploits the complementary features of each modality. This research has been applied to the audio-visual tracking of multiple speakers in meeting rooms, as illustrated in Figure 14. At the same time, a new probabilistic model for visual tracking has been proposed. This model allows for an implicitly modeling of motion, and early results show that this model leads to more stable and discriminative trackers than those using generic object modeling alone (e.g. with shape and color).

IM2.RTMAP - Reat Time Microphone Array Processing

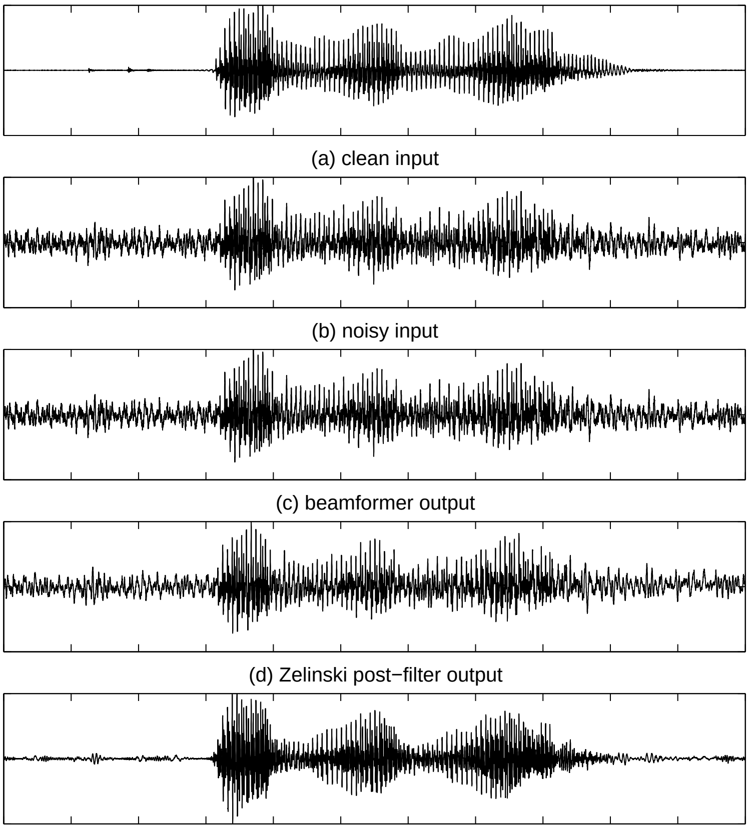
Funding: SNSF, through the (IM)2 NCCR

Duration: July 2002 – June 2004

Contact persons: Olivier Masson, Darren Moore, Iain McCowan, Hervé Bourlard

Description: An IM2 white paper project, entitled “Real-time Microphone Array Processing for Meeting Room” has commenced within the framework of IM2.SP. The project aims to use microphone arrays to address the problems of acquiring clean speech, detecting periods of voice activity and dynamically determining the location for each meeting participant. The major objectives are to research novel array processing techniques aimed at making arrays more viable in terms of cost, processing and space requirements, and also to produce a stand-alone system that will facilitate further research in IM2. In this framework, and as illustrated in Figure 15, new microphone array algorithms have been developed using new post-filter formulated for diffuse noise field (see IDIAP RR-39, to be published in IEEE Trans on Signal Processing).

Within the scope of the white paper, a sub-project & contract to HEV-Sion has been defined, entitled “Small Microphone Array Low Level Development Project”. In this sub-project, IDIAP has subcontracted to HEV Sion the task of implementing the hardware and low-level software required for the proposed modular array architecture. The sub-project commenced in August under the direction of Dr. Joseph Moerschell and has an expected duration of 6 months. The major deliverable for the sub-project is the demonstration of a stand-alone 8 channel microphone array with basic functionality.



(a) clean input

(b) noisy input

(c) beamformer output

(d) Zelinski post−filter output

(e) proposed post−filter output

Figure 15: Plots demonstrating speech enhancement using a 5 element microphone array. (a) original clean input signal; (b) noise corrupted input signal; (c) output of standard superdirective beamformer; (d) output of standard array post-filter; (e) output of proposed IDIAP post-filter optimised for diffuse noise field.

IM2.SP - Speech Processing

Funding: SNSF, through the (IM)2 NCCR

Duration: January 2002 – December 2003

Contact persons: Hemant Misra, Hervé Bourlard

Description: The goal of IM2.SP is to provide the IM2 NCCR with advanced and flexible speech processing modules which can be used as an input mode (voice input), as an audio indexing tool (requiring large vocabulary, continuous speech recognition systems) turning audio files into text, and as an output mode (mainly based on text-to-speech systems). The research goals of this IP are thus to improve state-of-the-art speech recognition algorithms (with respect to performance and robustness to noise and speech style), as well as (to a lesser extent) text-to-speech technologies.

During the first 9 months of IM2, and following the planned activites briefly presented above, the following has been achieved:

- Development of the TODE Decoder:

  The TODE recogniser developed at IDIAP is designed to meet the speech decoding needs of researchers at IDIAP and in the wider speech research community, and to insure easy adaptation and porting to different tasks and environments. In its current implementation, the main functionalities and features of TODE are:

  – Based on a time synchronous Beam Search decoding algorithm
  – Integrated with the TORCH machine learning toolkit
  – Accepts feature vectors or acoustic probability vectors as input
  – Supports both GMM and ANN-based acoustic modelling
  – Arbitrary N-gram language modelling
  – Compatible with many popular file formats
  – Linear lexicon
  – Integrated Word Error Rate (WER) calculation
  – Supported - development ongoing

  Manual of the speech decoder can now be downloaded from http://www.torch.ch/documentation.php.

- New algorithms for automatic audio segmentation have been developed and tested, including:

  – A particularly performant system for speech/non-speech detection has been developed and tested on several international databases.
  – A new approach towards automatic speaker clustering and speaker turn detection has been developed and tested. This system is based on a new information theory based clustering (generalizing the BIC criterion originally proposed by IBM) determining the optimal number of clusters without the need of any penalty term (as opposed to BIC).

- Work on microphone array (IDIAP) recording and sound source localization has just started. In this framework, one new PhD student (Guillame Lathoud) has been hired to work on this. See Section 2.2.1 for further detail.

KERNEL – Kernel Methods for Sequence Processing

Funding: Swiss National Science Foundation

Duration: February 2001 - February 2003

Contact persons: Quan Le, Samy Bengio

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

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

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

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

LAVA – Learning for Adaptable Visual Assistants

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

Duration: 36 months, March 2002-February 2005

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

Contact person: Samy Bengio, Jean-Marc Odobez, Mark Barnard, Pedro Quelhas, Alexei Pozdnoukhov

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

- Robust and efficient categorisation and interpretation of large numbers of objects, scenes and events, in real settings
- Automatic acquisition of knowledge of categories, for convenient construction or extension of applications.

M4 – MultiModal Meeting Manager

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

Duration: 36 months, March 2002-February 2005

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

Contact person: Hervé Bourlard, Daniel Gatica-Perez

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

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

FNSNF MULTI - Multimodal Interaction and Multimedia Data Mining

Funding: Swiss National Science Foundation

Duration: October 2002 - September 2004

Contact persons: Hervé Bourlard

Description: Since October 1, 2002, this new NSF project is integrating (and extending) all the SNSF research activities (apart from the IM2-NCCR activities) reported in the present document. As a unified research theme, the goal of the present Swiss National Science Foundation (SNSF) project will be to carry out fundamental research in the field of multimodal interaction, which covers a wide range of critical activities and applications, including recognition and interpretation of spoken, written and gestural language. Other key subthemes of this project will include the control of information access, typically through biometric user authentication techniques (including speaker and face verification). Building upon the same technologies, the present project will also investigate advanced approaches towards the structuring, retrieval and presentation of multimedia information, also referred to as "multimedia data mining". This is indeed a wide-ranging and important research area that includes not only the multimodal interaction described above, but also multimedia document analysis, indexing, and information retrieval, thus involving complex computer vision and data fusion algorithms.

FNSNF PROMO – PROnunciation MOdelling in Automatic Speech Recognition Systems

Funding: Swiss National Science Foundation

Duration: August 2000 - July 2002

Contact persons: Mathew Magimai Doss, Hervé Bourlard

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

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

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

RESPITE – REcognition of Speech by Partial Information TEchniques

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

Duration: January 1999 - September 2002

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

Contact persons: Hervé Bourlard, Andrew Morris

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

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

SCRIPT – Cursive Handwriting Recognition

Funding: Swiss National Science Foundation

Duration: October 1999 - September 2002

Contact person: Alessandro Vinciarelli

Description: The recognition of cursive handwritten words when only the image of the data is available is called Off-Line Cursive Script Recognition (CSR). The great variability of handwriting styles and the fact that the letters are connected are the major difficulties of the problem.
A system for single word recognition was developed. It presents an original normalisation method (based

on statistics) that improved significantly the performance with respect to traditional normalisation methods.
We are extending now the recognition problem to the automatic reading of sentences.

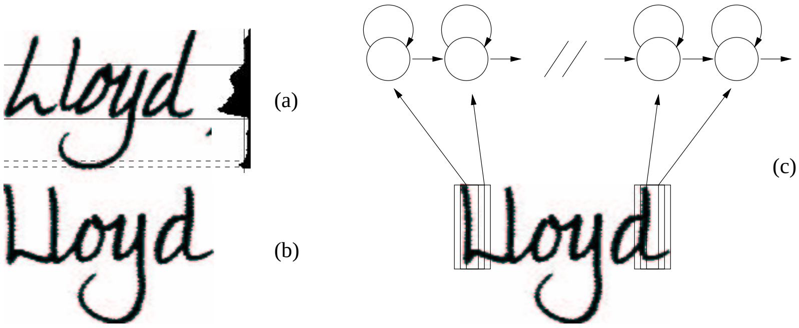


Figure 16: Single word recognition. The original image (a) is normalized (b) and modeled with HMMs (c). A HMM is created for every word in a list of possible interpretations of the data. The most likely model is assumed as transcription of the data.

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.

SPHEAR – SPeech, HEAring and Recognition

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

Duration: March 1998 - September 2002

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

Contact persons: Astrid Hagen, Hervé Bourlard, Andrew Morris

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

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

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

Funding: Swiss National Science Foundation

Duration: January 1999 - September 2002

Contact persons: Mohamed Benzeghiba, Hervé Bourlard

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

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

VOCR - Text Recognition for Video Retrieval

Funding: Swiss National Science Foundation

Duration: December 1999 - September 2002

Contact persons: Datong Chen, Jean-Marc Odobez

Description: The objective of this project is the investigation and development of algorithms for the detection, segmentation, and recognition of text in images and videos to be used for indexing and retrieval.

This year, the research has focused on three main issues :

- the design of a fast text localization focusing step, which enables text size normalization. It relies on a machine learning text verification step applied on background independent features.
- the improvement of text recognition results. It is addressed by a text segmentation step followed by an traditional OCR algorithm within a multi-hypotheses framework relying on multiple segments, language modeling and OCR statistics.
- the exploitation of temporal information to reach a better decision, correct errors. A selection mechanism of the best solution over time has been designed. Cuurrently, we are studying the possibility of merging solutions to produce a new recognition string.

All these techniques have been implemented and incorporated into software to be used in the european projects ASSAVID and CIMWOS. Experiments conducted on large databases of real broadcast documents provided by the partners (BBC, Canal+) have proven the robustness of our approach.

VoiceInPack – Low bit-rate speech transmission based on speech recognition and speech synthesis for online multiplayer games

Funding: Swiss Commission for Technology and Innovation (CTI)

Duration: November 2002 - April 2004

Partners: ETHZ and Komodo Entertainment Software SàRL

Contact persons: Hervé Bourlard

Description: VoiceInPack will contribute to the development of a Massively Online Game created by Komodo. VoiceInPack will develop a technology that allows, with a high level of compression, to fullfill Komodo's game voice requirements. VoiceInPack aims at developing very low bit rate speech transmission (over the internet, for game applications) by using the front-end of a speech recognition system to turn the voice signal into a sequence a phone (developed by IDIAP), complemented by additional prosodic parameters (such as pitch and duration). These phoneme sequences will then be sent over the communication channel, and later used as the input of the back-end of a text-to-speech system (developed by ETHZ). The output signal will also be modulated accordingly to the player's avatar. While interesting from an application point of view, this project will also allow further research and development activities in speech recognition (improvement of phonetic recognition) and speech synthesis (using prosodic features to improve naturalness).

7 Educational Activities

7.1 Current PhD Theses

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

7.2 PhD Defenses

- **Ph.D. candidate:** Nicolas Gilardi

  Supervisor: Prof. M. Maignan

  Examiners: Dr. S. Bengio (IDIAP), Prof. M. Maignan, Prof. M. Kanevski, Prof. S. Canu

  University: Lausanne

  Title: Machine Learning for Spacial Data Analysis

7.3 Participation in PhD Thesis Committees

- **Ph.D. candidate:** P.E. Sottas

  Committee member: Samy Bengio

  University: EPFL, Lausanne

  Date: 29.08.2002

  Title: Temporal Sequence Learning with Non-Equilibrium Recurrent Neural Networks

- **Ph.D. candidate:** Nicolas Gilardi

  Committee member: Samy Bengio

  University: Lausanne

  Date: 25.11.2002

  Title: Machine Learning for Spacial Data Analysis

- **Ph.D. candidate:** Antoine Rozenknop

  Committee member: Hervé Bourlard

  University: EPFL, Lausanne

  Date: 9.12.2002

  Title: Modèles syntaxiques probabilistes non-génératifs

- **Ph.D. candidate:** Patrick Nguyen

  Committee member: Hervé Bourlard

  University: EPFL, Lausanne

  Date: 25.09.2002

  Title: Speaker Adaptation: Modeling Variabilities

PhD Students	Project	Expected PhD	At IDIAP since	PhD Status	Thesis Supervisor	Thesis Director
SNSF PROJECTS						
AJMERA Jitendra	AudioSkim *	2004	01.01.01	3rd year, Speech Processing	H. Bourlard	Prof. H. Bourlard, EPFL
BENZEGHIBA Mohamed	SV-UCP *	2004	01.08.00	3rd year, Speech Processing	H. Bourlard	Prof. H. Bourlard, EPFL
CHEN Datong	VOCR *	2003	01.11.99	4th year, Computer Vision	J.M.Odobez	Dr. J.-P. Thiran, EPFL
COLLOBERT Ronan	Divide&Learn *	2004	01.08.02	2nd year, Machine Learning	S. Bengio	Not decided yet
DIMITRIKAKIS Christos	Divide&Learn *	2005	01.10.01	1st year, Machine Learning	S. Bengio	Not decided yet
FASEL Beat	FaceX *	2002	01.10.98	4th year, Computer Vision	D. Gatica-Perez	Prof. Van Goal, ETHZ
GILARDI Nicolas	CARTANN	2002	01.01.99	**Accepted**, Machine Learning	S. Bengio	Prof. Maignan, UNIL
IKBAL Shajith	HMM2 *	2004	01.05.00	3rd year, Speech Processing	H. Bourlard	Prof. H. Bourlard, EPFL
LE Quan	KERNEL *	2004	01.02.01	2nd year, Machine Learning	S. Bengio	Not decided yet
MAGIMAI DOSS Mathew	PROMO *	2004	25.10.99	3rd year, Speech Processing	H. Bourlard	Prof. H. Bourlard, EPFL
STEPHENSON Todd	BNS 4 ASR *	2003	01.03.99	4th year, Speech Processing	H. Bourlard	Prof. H. Bourlard, EPFL
TRUTNEV Alex	INSPECT	2004	01.08.00	3rd year, Speech Processing	M. Rajman	Prof. H. Bourlard, EPFL
VINCIARELLI Alessandro	SCRIPT *	2003	01.10.99	4th year, Computer Vision	S. Bengio	Prof. H. Bunke, Univ. Berne
WEBER Katrin	CORREL	2002	01.01.98	4th year, Speech Processing	H. Bourlard	Prof. H. Bourlard, EPFL
** as of 1.10.2002, all these individual SNSF projects have been merged into the MULTI project.*						
PROJECTS WITHIN THE (IM)2 NCCR						
BA Silèye	MUCATAR	2006	01.10.02	1st year, Computer Vision	J.M. Odobez	Not decided yet
CHIAPPA Silvia	BMI	2005	01.11.01	1st year, Machine Learning	S. Bengio	Not decided yet
KELLER Mikaela	MI	2006	01.12.02	1st year, Machine Learning	S. Bengio	Not decided yet
MISRA Hemant	SP	2005	24.07.01	2nd year, Speech Processing	H. Bourlard	Prof. H. Bourlard, EPFL
MONAY Florent	LH	2006	01.08.02	1st year, Computer Vision	D. Gatica-Perez	Not decided yet
POH HOON THIAN Norman	ACP	2006	01.09.02	1st year, Machine Learning	S. Bengio	Not decided yet
SMITH Kevin	MUCATAR	2006	21.11.02	1st year, Computer Vision	D. Gatica-Perez	Not decided yet
EUROPEAN PROJECTS, FUNDED BY OFES						
BARNARD Mark	ASSAVID / LAVA	2005	15.03.01	3rd year, Computer Vision	J.M. Odobez+S. Bengio	Prof. H. Bourlard, EPFL
CARDINAUX Fabien	CIMWOS	2005	01.10.01	2nd year, Computer Vision	S. Marcel	Not decided yet
LATHOUD Guillaume	SPHEAR / M4	2006	01.03.02	1st year, Speech Processing	H. Bourlard	Not decided yet
QUELHAS Pedro	LAVA	2006	01.11.02	1st year, Computer Vision	J.M. Odobez	Not decided yet
RODRIGUEZ Yann	COST 275	2006	01.09.02	1st year, Machine Learning	S. Marcel+S. Bengio	Not decided yet
POZDNOUKHOV Alexei	LAVA	2006	01.07.02	1st year, Machine Learning	S. Bengio	Not decided yet
OTHER PROJECTS						
JUST Agnès	GHOST (France Télécom R&D)	2006	01.10.02	1st Year, Computer Vision	S. Marcel	Not decided yet
McGREEVY Michael	EARS (DARPA)	2007	15.01.03	1st year, Speech Processing	H. Bourlard	Prof. Sridharan, QUT
TYAGI Vivek	EARS (DARPA)	2005	01.06.01	1st year, Speech Processing	H. Bourlard	Prof. H. Bourlard, EPFL

7.4 Courses

- **Title:** Speech and Language Engineering
  Lecturer and Director of the course: Prof. H Bourlard
  School: EPFL, Postgraduate
- **Title:** Decision, estimation and statistical pattern recognition: Application to speech recognition
  Lecturer: Prof. H Bourlard
  School: EPFL, DI/DSC Predoctoral School
- **Title:** Speech Processing
  Lecturer: Prof. H Bourlard
  School: EPFL, Undergraduate (2nd cycle)
- **Title:** Statistical Pattern Recognition
  Lecturer: Prof. H Bourlard
  School: European School of Medical Physics, Archamps (F), Nov. 16, 2002.
- **Title:** Statistical Machine Learning
  Lecturer: Dr Samy Bengio
  School: IDIAP

7.5 Other student projects

- **Trainee:** Alain Anthamatten
  Committee member: Pierre Dal Pont, Jean-Albert Ferrez
  University/School: HEVs
  Date: September 2002 - January 2003
  Title: Etude de l'impact économique de l'IDIAP et d'IM2

8 Scientific Activities

8.1 Editorship

Prof. Hervé Bourlard is

- Editor-in-Chief of Speech Communication
- Action Editor of Neural Network
- Associate Editor of Intl. Journal of Pattern Recognition and Artificial Intelligence
- Member of the Editorial Board of Futur(e)

Dr Samy Bengio is

- Associate Editor for the Journal of Computational Statistics

8.2 Scientific and Technical Committees

Prof. Hervé Bourlard is/was:

- Member of the Board of Trustees, Intl. Computer Science Institute, Berkeley, CA, USA.
- Member of the European Information Society Technology Advisory Group (ISTAG)
- Member of the Board of Trustees of the Swiss Network for Innovation
- Member of the Advisory Council of ISCA (International Speech Communication Association)
- Member of the IEEE Technical Committee on Neural Network Signal Processing
- Member of the Advisory Board of the European Speech Technology Network
- General Chairman, IEEE Neural Network for Signal Processing workshop, Martigny, 2002
- General Chairman, Eurospeech 2003, Geneva
- Co-Technical Chairman, IEEE Intl. Conference of Acoustics, Speech, and Signal Processing (ICASSP), Orlando, May 2002
- Member of the Scientific Committee, European Symposium of Artificial Neural Networks (ESANN), 2002
- Expert for several European projects

Dr Samy Bengio is

- Program Chair of the 2002 IEEE Workshop on Neural Networks for Signal Processing (NNSP'02)
- Program Committee of the 4th international conference on audio and video based biometric person authentication (AVBPA'2003)

8.3 Short Term Visits

- **Location:** University of Montreal, Yoshua Bengio's LISA laboratory
  Visitor: Samy Bengio
  Date: from 23 nov 2002 to 30 nov 2002
- **Location:** IBM Watson Research Center, Yorktown Heights (NY), USA.
  Visitor: Alessandro Vinciarelli
  Date: from May 13 to August 2

8.4 Scientific Presentations (other than conferences)

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

- **Event:** Workshop on HP Philantropic Program, 27 Sept. 2002

  Speaker: Samy Bengio

  Title: Presentation of the project MUST

- **Event:** Seminar given in the MANTRA group at EPFL, 8 Feb. 2002

  Speaker: Samy Bengio

  Title: A parallel Mixture of SVMs

- **Event:** Cognitive Vision Workshop, Zürich, 19-20 Sept. 2002

  Speaker: Jean-Marc Odobez

  Title: LAVA: Research directions and current results

- **Even:** Seminar given at ETHZ-Zurich, 11 Dec. 2002

  Visitor: Mohamed Faouzi BenZeghiba and Hervé Bourlard

  Title: Speaker verification based on user-customized password

- **Location:** IBM Watson Research Center, Yorktown Heights (NY), USA, August 16, 2002.

  Visitor: Hervé Bourlard

  Title: IDIAP Activities in Interactive Multimodal Information Management

- **Location:** CAST-EPFL, August 27, 2002

  Visitor: Hervé Bourlard

  Title: Interactions multimodales

- **Event:** 5th Catalan Conf. on Artificial Intelligence, Castellón, Spain. Keynote Talk

  Speaker: José del R. Millán

  Title: Adaptive Brain Interfaces for Communication and Control

  Date: from 24/10/02 to 25/10/02

- **Event:** Santa Lucia Institute (Hospital for neuromotor rehabilitation and home of the new European Brain Research Institute), Rome, Italy. Invited seminar

  Speaker: José del R. Millán

  Title: Non-Invasive Brain-Actuated Control of a Mobile Robot

  Date: 7 Nov. 2002

9 Publications (2001 and 2002)

9.1 Books and Book Chapters

[1] H. BOURLARD, T. ADALI, S. BENGIO, J. LARSEN, AND S. DOUGLAS, eds., *Proceedings of the Twelfth IEEE Workshop on Neural Networks for Signal Processing (NNSP)*, IEEE Press, 2002.

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

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

[4] J. MILLÁN, *Brain-computer interfaces*, in The Handbook of Brain Theory and Neural Networks: The Second Edition, M. A. Arbib, ed., The MIT Press, 2002.

[5] J. MILLÁN, *Robot navigation*, in The Handbook of Brain Theory and Neural Networks: The Second Edition, M. A. Arbib, ed., The MIT Press, 2002.

9.2 Articles in International Journals

[1] J. AJMERA, I. MCCOWAN, AND H. BOURLARD, *Speech/music discrimination using entropy and dynamism features in a hmm classification framework*, to be published in Speech Communication, (2002).

[2] F. BEAUFAYS, H. BOURLARD, H. FRANCO AND N. MORGAN, *Neural networks in automatic speech recognition*, (2001).

[3] S. BENGIO, C. MARCEL, S. MARCEL, AND J. MARIÉTHOZ, *Confidence measures for multimodal identity verification*, Information Fusion, 3 (2002), pp. 267–276.

[4] F. CAMASTRA AND A. VINCIARELLI, *Cursive character recognition by Learning Vector Quantization*, Pattern Recognition Letters, 22 (2001), pp. 625–629.

[5] F. CAMASTRA AND A. VINCIARELLI, *Intrinsic dimension estimation of data: an approach based on Grassberger-Procaccia's algorithm*, Neural Processing Letters, 14 (2001), pp. 27–34.

[6] F. CAMASTRA AND A. VINCIARELLI, *Estimating the intrinsic dimension of data with a fractal-based method*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 24 (2002), pp. 1404–1407.

[7] F. CAMASTRA AND A. VINCIARELLI, *Combining Neural Gas and Learning Vector Quantization for cursive character recognition*, to appear on Neurocomputing, (2003).

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

[9] R. COLLOBERT, S. BENGIO, AND Y. BENGIO, *A parallel mixture of SVMs for very large scale problems*, Neural Computation, 14 (2002), pp. 1105–1114.

[10] B. FASEL AND J. LUETTIN, *Automatic Facial Expression Analysis: A Survey*, Pattern Recognition, 36 (2003), pp. 259–275.

[11] I. LAPIDOT AND H. GUTERMAN, *Dichotomy between clustering performance and minimum distortion in piecewise-dependent-data (PDD) clustering*, to be published in IEEE Signal Processing Letters, (2003).

[12] S. MOELLER AND H. BOURLARD, *Analytic assessment of telephone transmission impact on asr performance using a simulation model*, Speech Communication, (2002).

[13] A. MORRIS, A. HAGEN, H. GLOTIN, AND H. BOURLARD, *Multi-stream adaptive evidence combination for noise robust asr*, Speech Communication, (2001).

[14] A. VINCIARELLI, *A survey on off-line cursive word recognition*, Pattern Recognition, 35 (2002), pp. 1433–1446.

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

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

9.3 Articles in Conference Proceedings

[1] J. AJMERA, H. BOURLARD, I. LAPIDOT, AND I. MCCOWAN, *Unknown-multiple speaker clustering using hmm*, in ICSLP, 2002.

[2] J. AJMERA, I. MCCOWAN, AND H. BOURLARD, *Robust hmm-based speech/music segmentation*, in ICASSP, 2002.

[3] S. BENGIO, *An asynchronous hidden markov model for audio-visual speech recognition*, in Advances in Neural Information Processing Systems, NIPS 15, S. Becker, S. Thrun, and K. Obermayer, eds., Vancouver, Canada, 2003, MIT Press.

[4] S. BENGIO AND J. MARIÉTHOZ, *Learning the decision function for speaker verification*, in IEEE International Conference on Acoustic, Speech, and Signal Processing, ICASSP, Salt Lake, City, USA, 2001.

[5] M. F. BENZEGHIBA AND H. BOURLARD, *Hybrid HMM/ANN and GMM Combination for User-Customized Password Speaker Verification*, IDIAP-RR 45, IDIAP, 2002.

[6] M. F. BENZEGHIBA AND H. BOURLARD, *User-Customized Password HMM Based Speaker Verification*, IDIAP-RR 35, IDIAP, 2002.

[7] M. F. BENZEGHIBA AND H. BOURLARD, *User-Customized Password Speaker Verification based on HMM/ANN and GMM models*, IDIAP-RR 10, IDIAP, 2002.

[8] 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.

[9] F. CARDINAUX AND S. MARCEL, *Face verification using MLP and SVM*, in XI Journees NeuroSciences et Sciences pour l'Ingenieur (NSI 2002), no. 21, La Londe Les Maures, France, 15-19 September 2002.

[10] D. CHEN, H. BOURLARD, AND J.-P. THRIAN, *Text Identification in Complex Background using SVM*, in Proceedings of the Int. Conf. on computer vision and pattern recognition, Hawaii, USA, Dec 2001.

[11] D. CHEN, J.-M. ODOBEZ, AND H. BOURLARD, *Text Segmentation and Recognition in Complex Background Based on Markov Random Field*, in Int. Conf. Pattern Recognition 2002, Quebec city, Canada, Oct 2002.

[12] D. CHEN, K. SHEARER, AND H. BOURLARD, *Text Enhancement with Asymmetric Filter for Video OCR*, in Proceedings of the 11th International Conference on Image Analysis and Processing, Palermo, Italy, Sep 2001, pp. 192–198.

[13] D. CHEN, K. SHEARER, AND H. BOURLARD, *Video OCR for Sport Video Annotation and Retrieval*, in Proceedings of the 8th IEEE International Conference on Mechatronics and Machine Vision in Practice, Hong Kong SAR, China, Aug 2001, pp. 57–62.

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

[15] R. COLLOBERT, Y. BENGIO, AND S. BENGIO, *Scaling large learning problems with hard parallel mixtures*, in International Workshop on Pattern Recognition with Support Vector Machines, SVM'2002, 2002.

[16] C. SANDERSON AND K. K. PALIWAL, *Noise Resistant Audio-Visual Verification via Structural Constraints*.

[17] B. FASEL, *Facial Expression Analysis using Shape and Motion Information Extracted by Convolutional Neural Networks*, in International IEEE Workshop on Neural Networks for Signal Processing (NNSP 02), Martigny, Switzerland, sep 2002, pp. 607–616.

[18] B. FASEL, *Head-Pose Invariant Facial Expression Recognition using Convolutional Neural Networks*, in International IEEE Conference on Multimodal Interfaces (ICMI 02), Pittsburgh, USA, oct 2002, pp. 529–534.

[19] B. FASEL, *Mutliscale Facial Expression Recognition using Convolutional Neural Networks*, in Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP 02), Ahmedabad, India, dec 2002.

[20] B. FASEL, *Robust Face Analysis using Convolutional Neural Networks*, in Proceedings of the International Conference on Pattern Recognition (ICPR 02), vol. 2, Quebec, Canada, aug 2002, pp. 40–43.

[21] D. GATICA-PEREZ AND M.-T. SUN, *Linking objects in videos by importance sampling*, in IEEE International Conference on Multimedia and Expo, 2002.

[22] D. GATICA-PEREZ AND M.-T. SUN, *Object localization in metric spaces for video linking*, in IEEE Workshop on Motion and Video Computing, 2002.

[23] D. GATICA-PEREZ, M.-T. SUN, AND A. LOUI, *Probabilistic home video structuring: Feature selection and performance evaluation*, in IEEE International Conference on Image Processing, 2002.

[24] N. GILARDI, S. BENGIO, AND M. KANEVSKI, *Conditional gaussian mixture models for environmental risk mapping*, in IEEE International Workshop on Neural Networks for Signal Processing (NNSP), 2002.

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

[26] A. HAGEN AND H. BOURLARD, *Error correcting posterior combination for robust multi-band speech recognition*, IDIAP-RR 10, IDIAP, Martigny, Switzerland, March 2001.

[27] A. HAGEN, H. BOURLARD, AND A. MORRIS, *Adaptive ml-weighting in multi-band recombination of gaussian mixture asr*, in ICASSP, vol. 1, 2001.

[28] S. IKBAL, H. MISRA, AND H. BOURLARD, *Phase AutoCorrelation (PAC) derived Robust Speech Features*, in Proceedings of the 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP-03), Hong Kong, April 2003.

[29] S. IKBAL, K. WEBER, AND H. BOURLARD, *Speaker Normalization using HMM2*, in Proceedings of the 2002 IEEE International Workshop on Neural Networks for Signal Processing (NNSP-02), Martigny, Switzerland, September 2002, pp. 647–656.

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

[31] G. LATHOUD AND I. MCCOWAN, *Location Based Speaker Segmentation*, in Proceedings of the 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP-03), Hong Kong, April 2003.

[32] S. MARCEL AND S. BENGIO, *Improving face verification using skin color information*, in Proceedings of the 16th International Conference on Pattern Recognition, IEEE Computer Society Press, 2002.

[33] S. MARCEL, C. MARCEL, AND S. BENGIO, *A state-of-the-art Neural Network for robust face verification*, in Proceedings of the COST275 Workshop on The Advent of Biometrics on the Internet, Rome, Italy, 2002.

[34] MARIÉTHOZ, J. AND BENGIO, S., *A comparative study of adaptation methods for speaker verification*, in International Conference on Spoken Language Processing ICSLP, Denver, CO, USA, September 2002, pp. 581–584.

[35] I. MCCOWAN, S. BENGIO, D. GATICA-PEREZ, G. LATHOUD, F. MONAY, D. MOORE, P. WELLNER, AND H. BOURLARD, *Modeling human interaction in meetings*, in Proceedings of International Conference on Acoustics, Speech and Signal Processing, Hong Kong, April 2003.

[36] I. MCCOWAN AND H. BOURLARD, *Microphone array post-filter for diffuse noise field*, IDIAP-RR 39, IDIAP, Martigny, Switzerland, 2001.

[37] I. McCowan, A. Morris, and H. Bourlard, *Robust speech recognition with small microphone arrays using the missing data approach*, IDIAP-RR 09, IDIAP, Martigny, Switzerland, 2002.

[38] H. Misra, H. Bourlard, and V. Tyagi, *New entropy based combination rules in HMM/ANN multi-stream ASR*, in Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Hong Kong, April 2003.

[39] D. Moore and I. McCowan, *Microphone array speech recognition : Experiments on overlapping speech in meetings*, in Proceedings of the 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP-03), Hong Kong, April 2003.

[40] A. Morris, *Data utility modelling for mismatch reduction*, in Proc. CRAC (workshop on Consistent & Reliable Acoustic Cues for sound analysis), Aalborg, Denmark, September 2 2001.

[41] A. Morris, J. Barker, and H. Bourlard, *From missing data to maybe useful data: soft data modelling for noise robust asr*, in Proc. WISP, no. 06, Stratford-upon-Avon, England, April 2-3 2001.

[42] A. Morris, A. Hagen, and H. Bourlard, *Map combination of multi-stream hmm or hmm/ann experts*, in Proc. Eurospeech, Aalborg, Denmark, September 3-7 2001.

[43] A. Morris, B. Obermaier, and G. Pfurtscheller, *Eeg pattern recognition through multi-stream evidence combination*, in Proc. World Congress on Neuroinformatics, Vienna University of Technology, Austria, September 24-29 2001.

[44] A. Morris, S. Payne, and H. Bourlard, *Low cost duration modelling for noise robust speech recognition*, in Proc. ICSLP, Denver, Colorado, USA, September 16-20 2002.

[45] J. Mouriño, S. Chiappa, R. Jané, and J. Millán, *Evolution of the mental states operating a brain-computer interface*, in Proceedings of the International Federation for Medical and Biological Engineering, Vienna, Austria, December 2002.

[46] J.-M. Odobez and D. Chen, *Video Text Recognition based on Markov Random Field and Grayscale Consistency Constraint*, in Int. Conf. Image Processing 2002, Rochester, NY, USA, Sept 2002.

[47] N. Poh, S. Bengio, and J. Korczak, *A multi-sample multi-source model for biometric authentication*, in IEEE International Workshop on Neural Networks for Signal Processing (NNSP), 2002.

[48] N. Poh, S. Marcel, and S. Bengio, *Improving face authetication using virtual samples*, in IEEE International Conference on Acoustics, Speech, and Signal Processing, no. 40, 2003.

[49] T. A. Stephenson, J. Escofet, M. Magimai-Doss, and H. Bourlard, *Dynamic Bayesian network based speech recognition with pitch and energy as auxiliary variables*, in 2002 IEEE International Workshop on Neural Networks for for Signal Processing (NNSP 2002), Martigny, Switzerland, September 2002, pp. 637–646.

[50] T. A. Stephenson, M. Magimai-Doss, and H. Bourlard, *Auxiliary variables in conditional Gaussian mixtures for automatic speech recognition*, in Seventh International Conference on Spoken Language Processing (ICSLP 2002), vol. 4, Denver, CO, USA, September 2002, pp. 2665–2668.

[51] T. A. Stephenson, M. Magimai-Doss, and H. Bourlard, *Mixed Bayesian networks with auxiliary variables for automatic speech recognition*, in International Conference on Pattern Recognition (ICPR 2002), vol. 4, Quebec City, PQ, Canada, August 2002, pp. 293–296.

[52] T. A. Stephenson, M. Magimai-Doss, and H. Bourlard, *Speech recognition of spontaneous, noisy speech using auxiliary information in Bayesian networks*, in Proceedings of the 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP-03), Hong Kong, April 2003.

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

[54] A. Vinciarelli and S. Bengio, *Offline cursive word recognition using continuous density Hidden Markov Models trained with PCA or ICA features*, in Proceedings of International Conference on Pattern Recognition, vol. III, Quebec City (Canada), 2002, pp. 81–84.

[55] A. Vinciarelli and S. Bengio, *Writer adaptation techniques in HMM based off-line cursive script recognition*, in Proceedings of 8^{th} International Conference on Frontiers on Handwriting Recognition, Niagara on the Lake (Canada), 2002, pp. 287–291.

[56] K. WEBER, S. BENGIO, AND H. BOURLARD, *HMM2- Extraction of Formant Features and their Use for Robust ASR*, in European Conference on Speech Communication and Technology (Eurospeech 2001), Aalborg, Denmark, September 2001.

[57] K. WEBER, S. BENGIO, AND H. BOURLARD, *Speech Recognition Using Advanced HMM2 Features*, in Automatic Speech Recognition and Understanding Workshop, Madonna di Campiglio, Italy, December 2001.

[58] K. WEBER, S. BENGIO, AND H. BOURLARD, *Increasing Speech Recognition Noise Robustness with HMM2*, in IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 02), Orlando, Florida, USA, May 2002.

[59] K. WEBER, F. DE WET, B. CRANEN, L. BOVES, S. BENGIO, AND H. BOURLARD, *Evaluation of Formant-Like Features for ASR*, in International Conference on Spoken Language Processing (ICSLP 2002), Denver, CO, USA, September 2002.

9.4 IDIAP Research Reports

[1] J. AJMERA, H. BOURLARD, AND I. LAPIDOT, *Improved unknown-multiple speaker clustering using hmm*, IDIAP-RR 23, IDIAP, Martigny, Switzerland, 2002.

[2] J. AJMERA, I. MCCOWAN, AND H. BOURLARD, *Bic revisited for speaker change detection*, IDIAP-RR 39, IDIAP, Martigny, Switzerland, 2002.

[3] A.POZDNOUKHOV, *The analysis of kernel ridge regression learning algorithm.*, IDIAP-RR 54, IDIAP, Martigny, Switzerland, 2002.

[4] M. BARNARD, J.-M. ODOBEZ, AND S. BENGIO, *Multi-modal audio-visual event recognition for football analysis*, IDIAP-RR 12, IDIAP, 2003.

[5] B. H. F. H. BEAUFAYS, F AND N. MORGAN, *Neural networks in automatic speech recognition*, IDIAP-RR 09, IDIAP, 2001.

[6] S. BENGIO, *Multimodal authentication using asynchronous HMMs*, IDIAP-RR 02, IDIAP, 2003.

[7] S. BENGIO, F. BIMBOT, J. MARIÉTHOZ, V. POPOVICI, F. PORÉE, E. BAILLY-BAILLIÈRE, G. MATAS, AND B. RUIZ, *Experimental protocol on the BANCA database*, IDIAP-RR 05, IDIAP, 2002.

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

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

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

[11] M. F. BENZEGHIBA, H. BOURLARD, AND J. MARIÉTHOZ, *Speaker Verification Based On User-Customized Password*, IDIAP-RR 13, IDIAP, 2001.

[12] F. CAMASTRA AND A. VINCIARELLI, *Estimating the intrinsic dimension of data with a fractal-based method*, IDIAP-RR 2, IDIAP, 2002.

[13] F. CARDINAUX AND S. MARCEL, *Face verification using MLP and SVM*, IDIAP-RR 21, IDIAP, 2002.

[14] F. CARDINAUX, C. SANDERSON, AND S. MARCEL, *Comparison of MLP and GMM classifiers for face verification on XM2VTS*, IDIAP-RR 10, IDIAP, 2003.

[15] D. CHEN AND J.-M. ODOBEZ, *A New Method of Contrast Normalization for Verification of Extracted Video Text having Complex Backgrounds*, IDIAP-RR-02 16, IDIAP, Apr 2002.

[16] D. CHEN AND J.-M. ODOBEZ, *Comparison of Support Vector Machine and Neural Network for Text Texture Verification*, IDIAP-RR-02 19, IDIAP, Martigny, Apr 2002.

[17] D. CHEN AND J.-M. ODOBEZ, *Monte Carlo Video Text Segmentation*, IDIAP-RR-03 07, IDIAP, Jan 2003.

[18] D. CHEN, J.-M. ODOBEZ, AND H. BOURLARD, *Text Detection and Recognition in Images and Videos*, IDIAP-RR-02 61, IDIAP, Dec 2002.

[19] R. COLLOBERT, S. BENGIO, AND J. MARIÉTHOZ, *Torch: a modular machine learning software library*, IDIAP-RR 46, IDIAP, 2002.

[20] C. SANDERSON AND S. BENGIO, *Robust Features for Frontal Face Authentication in Difficult Image Conditions*, IDIAP-RR 05, IDIAP, January 2003.

[21] J. CZYZ, S. BENGIO, C. MARCEL, AND L. VANDENDORPE, *Scalability analysis of audio-visual person identity verification*, IDIAP-RR 04, IDIAP, 2003.

[22] F. DE WET, K. WEBER, L. BOVES, B. CRANEN, S. BENGIO, AND H. BOURLARD, *Evaluation of formant-like features for automatic speech recognition*, IDIAP-RR 08, IDIAP, 2003.

[23] C. DIMITRAKAKIS AND S. BENGIO, *Online policy adaptation for ensemble algorithms*, IDIAP-RR 28, IDIAP, 2002.

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

[25] D. GATICA-PEREZ, G. LATHOUD, I. MCCOWAN, J.-M. ODOBEZ, AND D. MOORE, *Audio-visual speaker tracking with importance particle filters*, IDIAP-RR 37, IDIAP, 2002.

[26] N. GILARDI, S. BENGIO, AND M. KANEVSKI, *Estimation of conditional distributions using gaussian mixture models*, IDIAP-RR 03, IDIAP, 2002.

[27] A. HAGEN, *Robust speech recognition based on multi-stream processing*, IDIAP-RR 41, Lausanne, Switzerland, December 2001.

[28] A. HAGEN AND A. C. MORRIS, *Recent advances in the multi-stream hmm/ann hybrid approach to noise robust asr*, IDIAP-RR 57, IDIAP, 2002.

[29] S. IKBAL, H. BOURLARD, S.BENGIO, AND K. WEBER, *IDIAP HMM/HMM2 System: Theoretical Basis and Software Specifications*, IDIAP-RR 27, IDIAP, Martigny, Switzerland, 2001.

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

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

[32] I. LAPIDOT, *Self-organizing-maps with BIC for speaker clustering*, IDIAP-RR 60, IDIAP, Martigny, Switzerland, 2002.

[33] I. LAPIDOT, *What is better: GMM of two gaussians or two clusters with one gaussian?*, IDIAP-RR 56, IDIAP, Martigny, Switzerland, 2002.

[34] I. LAPIDOT AND A. MORRIS, *Extended BIC criterion for model selection*, IDIAP-RR 42, IDIAP, Martigny, Switzerland, 2002.

[35] Q. LE AND S. BENGIO, *Hybrid generative-discriminative models for speech and speaker recognition*, IDIAP-RR 06, IDIAP, 2002.

[36] Q. LE AND S. BENGIO, *Client dependent gmm-svm models for speaker verification*, IDIAP-RR 03, IDIAP, 2003.

[37] V. LEMAIRE, *Bagging using the vmse cost function*, IDIAP-RR 27, France Telecom Research and Development, 2002.

[38] V. LEMAIRE AND F. CLÉROT, *Som-based clustering for on-line fraud behavior classification: a case study*, IDIAP-RR 30, France Telecom Research and Development, 2002.

[39] M. MAGIMAI.-DOSS, T. A. STEPHENSON, AND H. BOURLARD, *Modelling auxiliary information (pitch frequency) in hybrid HMM/ANN based ASR systems*, IDIAP-RR 62, IDIAP, 2002.

[40] S. MARCEL, *Evaluation protocols and comparative results for the Triesch hand posture database*, IDIAP-RR 50, IDIAP, 2002.

[41] S. MARCEL, *Gestures for multi-modal interfaces: A review*, IDIAP-RR 34, IDIAP, 2002.

[42] S. MARCEL, *Robust face verification using skin color and Neural Networks*, IDIAP-RR 49, IDIAP, 2002.

[43] A. C. MORRIS, *Noise pdf transformation in secondary feature processing*, IDIAP-RR 29, IDIAP, 2002.

[44] A. C. MORRIS AND H. MISRA, *Confusion matrix based posterior probabilities correction*, IDIAP-RR 53, IDIAP, 2002.

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

[46] F. PORÉE, J. MARIÉTHOZ, S. BENGIO, AND F. BIMBOT, *The BANCA database and experimental protocol for speaker verification*, IDIAP-RR 13, IDIAP, 2002.

[47] C. SANDERSON, S. BENGIO, H. BOURLARD, J. MARIETHOZ, R. COLLOBERT, M. F. BENZEGHIBA, F. CARDINAUX, AND S. MARCEL, *Speech & Face Based Biometric Authentication at IDIAP*, IDIAP-RR 13, IDIAP, February 2003.

[48] K. SHEARER, C. DORAI, AND S. VENKATESH, *Detection of narrative structure for annotation of news broadcasts*, IDIAP-RR 3, IDIAP, 2001.

[49] K. SHEARER AND S. VENKATESH, *Artifacts of the colour coherence vector and an alternative similarity measure*, IDIAP-RR 2, IDIAP, 2001.

[50] T. A. STEPHENSON, *Conditional Gaussian mixtures*, IDIAP-RR 11, IDIAP, 2003.

[51] T. A. STEPHENSON, M. MAGIMAI-DOSS, AND H. BOURLARD, *Speech recognition with auxiliary information*, IDIAP-RR 58, IDIAP, 2002.

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

[53] A. VINCIARELLI AND S. BENGIO, *Transforming the feature vectors to improve HMM based cursive word recognition systems*, IDIAP-RR 32, IDIAP, 2002.

[54] A. VINCIARELLI, S. BENGIO, AND H. BUNKE, *Offline recognition of large vocabulary cursive handwritten text*, IDIAP-RR 01, IDIAP, 2003.

[55] K. WEBER, S. BENGIO, AND H. BOURLARD, *A Pragmatic View of the Application of HMM2 for ASR*, IDIAP-RR 23, IDIAP, Martigny, Switzerland, 2001.

9.5 IDIAP Communications

[1] T. COLLADO, *Developement d'un systeme de demande interactif via le telephone (infovox)*, IDIAP-Com 08, IDIAP, 2001.

[2] C. SANDERSON, *Speech Processing & Text-Independent Automatic Person Verification*, IDIAP-Com 08, IDIAP, 2002.

[3] C. SANDERSON, *The VidTIMIT Database*, IDIAP-Com 06, IDIAP, 2002.

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

[5] IDIAP, *Activity report 2001*, IDIAP-COM 01, IDIAP, 2002.

[6] IDIAP, *Activity report 2002*, IDIAP-COM 01, IDIAP, 2003.

[7] S. KRSTULOVIĆ, *Epfl lab session 1/2: Introduction to Gaussian statistics and pattern recognition*, IDIAP-Com 6, IDIAP, 2001.

[8] S. KRSTULOVIĆ, *Epfl lab session 2/2: Introduction to hidden markov models*, IDIAP-Com 7, IDIAP, 2001.

[9] D. MOORE, *The idiap smart meeting room*, IDIAP-COM 07, IDIAP, 2002.

[10] D. MOORE, *Tode: A decoder for continuous speech recognition*, IDIAP-COM 09, IDIAP, 2002.

[11] A. C. MORRIS, *An information theoretic measure of sequence recognition performance*, IDIAP-COM 03, IDIAP, 2002.

[12] H. WANG, *Rebuilding speech recognition on windows*, IDIAP-Com 09, IDIAP, 2001.

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

[14] H. WANG AND S. BENGIO, *The mnist database of handwritten upper-case letters*, IDIAP-Com 04, IDIAP, 2002.

[15] H. WANG, A. VINCIARELLI, AND F. FORMAZ, *Handwriting recognition demo*, IDIAP-Com 02, IDIAP, 2002.

9.6 Other Documents

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

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