IDIAP COMMUNICATION

IDIAP

Martigny - Valais - Suisse

Activity Report 1997

IDIAP–Com 98-01

February 98

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

IDIAP

Martigny - Valais - Suisse

Institut Dalle Molle d'Intelligence Artificielle Perceptive

Contents

1 General Presentation of IDIAP

1.1 Status and Structure

The Dalle Molle Institute for Perceptual Artificial Intelligence (IDIAP, “Institut Dalle Molle d'Intelligence Artificielle Perceptive”) is a semi-private non-profit research institute founded in 1991 to celebrate the 20th anniversary of the Dalle Molle Foundation. It is the third research center initiated by the Dalle Molle Foundation, after ISSCO in Geneva and IDSIA in Lugano.

In November 1996, and as initially planned at the establishment of the institute, IDIAP acquired the status of Research Foundation (IDIAP Foundation), now independent of the Dalle Molle Foundation, and is officially affiliated with the Swiss Federal Institute of Technology in Lausanne (EPFL) and the University of Geneva.

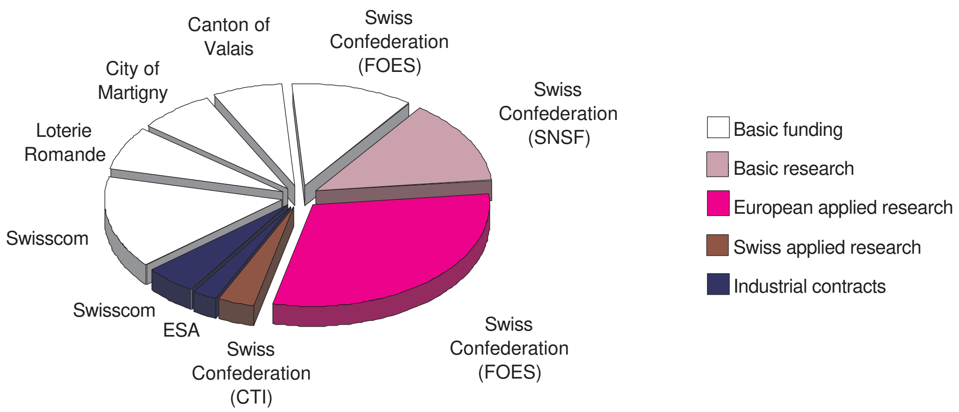

Figure 1: *Relative distribution of IDIAP funding in 1996.*

Today, IDIAP is primarily funded by long-term support from the the Swiss Confederation (FOES), the Canton of Valais, the City of Martigny, and Swisscom. The “Loterie Romande” is also supporting our research efforts with annual grants. In addition, IDIAP receives substantial research grants from the Swiss National Science Foundation (SNSF) for basic research projects and the Federal Office for Education and Science (FOES) for European projects. The relative distribution of IDIAP's funding in 1997 is illustrated in Figure 1.

For the last few years, IDIAP has counted an average of about 25-30 scientists in residence at IDIAP including permanent staff, postdoctoral fellows, PhD students, and short-medium term visitors.

The general management structure of IDIAP is illustrated in Figure 2 and is composed of a Foundation Council, a Board of Directors, and a Scientific Committee (advising the Board of Directors). It is also intended to set up an Economic Relation Committee, which will be responsible for publicizing IDIAP's research results across the industrial world, as well as providing IDIAP with new research opportunities of particular interest to industry.

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. Further information about IDIAP can be found on our web server `http://www.idiap.ch`.

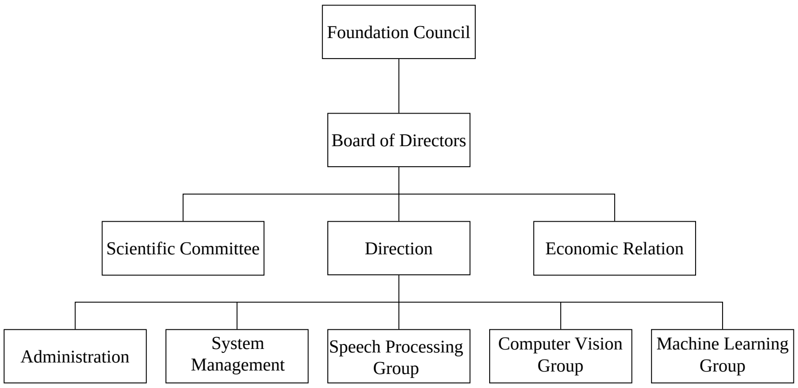

Figure 2: *IDIAP Structure.*

1.2 Research and Development Activities

The main 1997 research activities of IDIAP, which focus on medium-long term objectives, are described in detail in the present Activity Report. Focusing on a few well defined research axes, IDIAP carries out fundamental research and develops prototype systems (to validate its research results) along three complementary research directions:

- **Speech Processing, including all aspects of automatic speech recognition and speaker verification**.

  This involves the development and testing of state-of-the-art, as well as advanced, speech recognition systems (ranging from small to large vocabularies, from speaker dependent to speaker independent, from isolated words to continuous speech and keyword/keyphrase spotting). While mainly focusing on telephone speech (in the framework of our contract with Swisscom), this work is also applied to microphone input. Current research activities mainly focus on improving speech unit models towards better robustness to noise and speaking styles. Amongst other activities, this involves further developments of hybrid systems using hidden Markov models and artificial neural networks, as well as advanced research in sub-band and multi-stream processing (as pioneered by IDIAP, together with the Faculté Polytechnique de Mons in Belgium, the International Computer Science Institute in Berkeley, USA, and the Oregon Graduate Institute in Portland, USA). Large vocabulary speech recognition systems, involving complex pronunciation dictionaries and grammars, are also developed and tested. As described below, the IDIAP Speech Processing group is involved in numerous national and international (ESPRIT, ACTS, COST, TMR) projects.

  In speaker verification, most of the research activities so far have focused on the improvement of current state-of-the-art algorithms, and on the development of innovative solutions combining concurrent and/or complementary strategies. Recently, IDIAP participated in the international NIST (National Institute of Standards and Technology, USA) evaluation and showed that their technology was at the leading edge in that field.

  The main applications and prototype systems which have been developed and tested so far were oriented towards: advanced voice servers (e.g., for accessing remote databases), personal call assistants, calling card applications (involving voice dialing and speaker verification), automatic audio indexing and retrieval, and multimodal user verification systems. Finally, to facilitate research, as well as multi-lingual system development, IDIAP is also actively involved in speech database collection and management activities, which have mainly taken place in the framework of our contract with Swisscom (Polyphone and GSM data), as well as in an EC project.

- **Computer Vision, including object recognition, motion analysis, sensor fusion, and document recognition**.

  The work focuses on specific research topics which address problems in three target applications: multimodal interfaces, access security, and information management and retrieval.

  Research in object recognition has addressed the problem of fast object detection (e.g. face detection) in grey-level images, paying particular attention to the appearance variability of objects. This has lead to the development of a new object detection method based on the combination of an FFT with a MLP, which is of the order of a magnitude faster than state-of-the-art methods while obtaining identical performance levels. In document analysis, our activities have focused on the basic problems of feature representation and similarity measures. We have investigated methods based on the combination of rotated histograms with an MLP which have lead to state-of-the-art performance.

  Research in motion analysis has been concerned with the modeling and recognition of deformable objects and their tracking in image sequences. The work has addressed the independent modeling of shape and intensity using appearance based modeling techniques which were applied to the tracking of lips and the tracking of articulators in X-ray images. Based on the lip tracker, algorithms have been developed for the modeling and recognition of visual speech (lip-reading) and for the verification of a person's identity.

  Work in sensor fusion has addressed specific problems in the fusion of classifiers for person authentication and in the combination and modeling of audio with visual information in speech recognition. Much of this work has largely benefited from collaboration with the speech and machine learning groups and has lead to substantial advances in research.

- **Machine Learning, including pattern classification, data analysis and knowledge extraction**.

  This involves a high level of skill in all state-of-the-art methodologies for automated learning in order to apply them to the concrete problems of classification, pattern recognition, and expert fusion arising in perceptual artificial intelligence.

  A strong expertise is maintained in techniques of nature as diverse as artificial neural networks (connectionism), Bayesian networks (statistical learning), decision trees (symbolic learning), support vector machines (optimization) and logical analysis of data (Boolean functions theory).

  The fruitful cross fertilization between this wide base of learning techniques and specific applications in speech processing and pattern recognition has already lead to some original approaches and unexpected results. However, it requires an important research effort to adapt general methods to problems with characteristics such as large and noisy databases (speech databases). In particular, significant research effort is applied to the decomposition of large scale problems into sets of simpler subproblems.

  Other prospective activities are also carried out in which our know-how in automated learning is applied to time series prediction as well as to the design of assisted diagnosis systems.

In 1997, these research activities produced more than 60 research papers, of which over 45 were published in authoritative books and journals or presented at international conferences (see Section 7 for more detail).

As briefly described above, and as illustrated by Figure 3, the activities in the three research groups have been defined to be as complementary as possible, while fostering active collaboration across the different research themes. While speech processing and computer vision are often complementary in (multimodal) applications, they are also often based on common theories and mathematical tools, and can benefit strongly from interaction. Some recent developments in handwriting recognition, for instance, have been using hidden Markov models, initially developed in speech recognition. Similarly, some recent advances in multi-stream processing will be exploited in both groups and will also directly

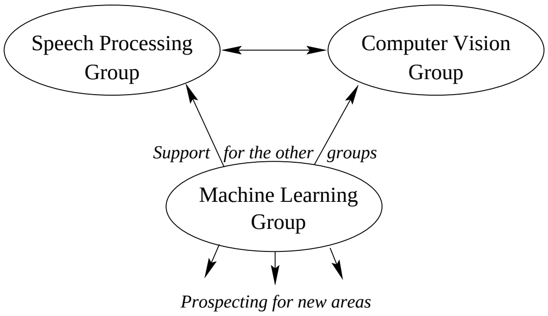

Figure 3: *Inter-dependencies between the three research groups.*

benefit the development of multimodal systems (as recently shown at IDIAP by some preliminary work on audio-visual speech recognition). The Machine Learning group investigates and develops new technologies that are common and useful to speech processing, computer vision, and multimodal processing. It thereby plays the role of technical support to the other, more application oriented, groups. For example, the new methods proposed for decomposing a learning problem into subproblems have been successfully applied to speaker verification, and research on the ways of recombining the submodules of such decomposition contributes directly to the problem of the fusion of multimodal experts. As a secondary goal, the Machine Learning group is also responsible for the identification of new application areas which could directly benefit from the available technology (primarily developed in the framework of speech and vision) and which could become important to future IDIAP activities (e.g., times series prediction). In this latter case, research will often be done in collaboration with other institutes with more expertise in the identified area.

1.3 Participation in National and European Community Research Projects

The activity of IDIAP within the framework of national and European Community projects has been particularly intense for the last few years, and IDIAP has played a leading role in the conception and coordination of many projects.

In 1997, with regards to projects funded by the European Community in the framework of the fourth Program for Research and Technology, IDIAP was actively involved in five EC projects:

- 2 Telematics projects: CAVE, to develop speaker verification systems in banking and telecommunication, and SpeechDat, to produce, standardize, and evaluate large multi-lingual speech corpora.
- 2 COST projects: COST249 on automatic speech recognition over the telephone, and COST250 on automatic speaker verification over the telephone. In comparison to non-Swiss partners, it is important to note here that IDIAP has received substantial funding from FOES in the framework of these two COST projects, which allowed us to significantly boost our expertise in speech and speaker recognition, and to initiate many new industrial and EC projects.
- 1 ACTS (Advanced Communications Technologies and Services) project on multimodal verification for teleservices and security applications (M2VTS).

In 1997, four additional EC projects (often initiated by IDIAP) were also granted to IDIAP:

- 1 ESPRIT Long Term Research project: THematic Indexing of Spoken Language (THISL) on automatic indexing and vocal access of recorded broadcast news (from BBC).
- In the TMR (Training and Mobility of Researchers) program that aims to promote international mobility of researchers, IDIAP played a key role in the conception of the SPHEAR (SPeech and

HEARing) project which has the objective of a better understanding of auditory processing, and to deploy this understanding in automatic speech recognition in adverse conditions.

- 1 in Telematics: PIoneering Caller Authentication for Secure Service Operation (PICASSO) on the use of automatic speaker verification systems in calling card and personal communication assistant applications.
- 1 within the SOCRATES/ERASMUS program, in which IDIAP is the only Swiss representative. The goal of this project is to define and initiate a European Master program in language and speech. Common core courses will be given in all the represented European countries, followed by specialized courses and projects which will be given in specific institutions.

In the framework of national projects, IDIAP was mainly involved in several Swiss National Science Foundation projects (mainly for the education and training of PhD students), such as:

- GLAD: Generalization of "Logical Analysis of Data" techniques
- Compact hardware-friendly neural networks (now mainly focusing on mixture of experts)
- ETCv: A system to help labeling of speech corpora (finished in Oct.97)
- ZEPHYR: Time series prediction with hybrid Markov models (granted in 1997)
- MULTICHAN: Non-stationary multichannel signal processing (granted in 1997)
- ARTIST: Articulatory representation towards improved speech technology (granted in 1997).
- AV-COM: Audio-visual combination (granted in 1997, for equipment).

IDIAP was also involved in one CTI (Commission for Technology and Innovation) project with Sun Microsystems, aComm, Swisscom and EPFL to develop workstation-based speech and speaker recognition systems for SwissNet.

More information about these projects is given in Section 3.

1.4 Collaboration with other Organizations and Companies

Throughout the last few years, IDIAP has maintained close contacts with research organizations, universities, and industries working in the same research and developments areas. Those contacts typically originate from the follow-up of successful projects, or are based on personal long-term relationships and regular exchanges with some particular institutions. Just as a few examples, we can mention here:

- As a follow-up of contacts generated by a CTI project, IDIAP keeps collaborating with Swisscom and Sun Microsystems. A follow-up project is currently in discussion, as well as the possibility of a start-up company exploiting recent results obtained in the framework of voice servers.
- Initiated by European projects, IDIAP now has very good contacts with several companies, including Cerberus (CH), BBC (UK), Daimler-Benz (D), Thomson (F), and several universities including Cambridge University (UK), Sheffield University (UK), Faculté Polytechnique de Mons (BE), and IMT (Neuchatel).
- Based on personal contacts and regular information exchange, we can mention here Rutgers University (RUTCOR) and the International Computer Science Institute (Berkeley, CA) with which further collaborator, including student exchange, is currently planned.
- More recently, in the framework of a project funded by the Catalyst Foundation (USA), we started a 4 year collaboration with Johns Hopkins University (Baltimore, USA) and the Indian Institute of Technology (Delhi) on micropower analog VLSI implementation of continuous speech recognition systems.

1.5 Training Activities and Regional Development

On top of high quality research and development, we consider that two other major functions of IDIAP are:

Training and supervision of PhD students (most of the time affiliated with EPFL or University of Geneva) and postdoctoral fellows, as well as short-term or medium-term visitors from academia (including ETS) and industry. As an example of this specific concern, IDIAP is currently working (as the only Swiss representative) with other European partners (in the framework of a European Socrates/Erasmus project) on defining the content of a European Masters in Language Technology (in which core courses would be given in all countries, followed by specialized courses and projects that would be taken in pre-defined countries). The first test of this Master program should start in 1999.

IDIAP is thus very active in the training of researchers and engineers, as well as in the training of highly qualified personnel in the scientific and technical fields. As of this writing, IDIAP is host to 12 PhD students, as well as several graduating students preparing their final thesis and coming from EPFL, Eurecom (F), ENST (F) and the ETS (Superior Technical School) of Sion. Every year, IDIAP also has a budget for 36 months (12 months for each group) of short-term visits for external fellows or students.

Technology transfer and industrial support, with two motivations: (1) keeping up-to-date with the technology (since it has often become too expensive for even the largest companies to maintain in-house competence in all important areas), and (2) developing and testing prototypes oriented towards applications of special interest to some of our sponsors (e.g., Swisscom). In this framework, IDIAP should thus be able to perform qualitative and quantitative analyses, enhance the applicability of current base technologies, integrate technologies into pilot systems, and engage in active technology transfer. Through open and intensive industrial collaboration, IDIAP aims to play an important role in promoting the economic development of the Canton of Valais.

Finally, IDIAP is regularly involved in the organization of scientific events such as:

- The first international conference on "Audio-and Video-based Biometric Person Authentication" (AVBPA), held in Crans-Montana on March 12-14, 1997.
- "Journées d'Etude sur la Parole" (JEP'98), which will be held in Martigny on June 15-19, 1998.

2 Staff

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

Phone: +41 – 27 – 721 77 11

Fax: +41 – 27 – 721 77 12

Internet: http://www.idiap.ch

2.1 Scientific Staff

Persons at IDIAP in 1997 or as of this writing:

Mr.	Johan Myhre Andersen Johan.Myhre.Andersen@idiap.ch	research assistant +41 – 27 – 721 77 24
Dr.	Souheil Ben Yacoub Souheil.Ben-Yacoub@idiap.ch	research scientist +41 – 27 – 721 77 38
Ms.	Giulia Bernardis Giulia.Bernardis@idiap.ch	research assistant +41 – 27 – 721 77 36
Mr.	Olivier Bornet Olivier.Bornet@idiap.ch	research engineer, System Management group leader +41 – 27 – 721 77 40
Prof.	Hervé Bourlard Herve.Bourlard@idiap.ch	director, Speech Processing group leader +41 – 27 – 721 77 20
Mr.	Gilles Caloz Gilles.Caloz@idiap.ch	research assistant +41 – 27 – 721 77 23
Dr.	Gérard Chollet Gerard.Chollet@idiap.ch	consultant until April 97
Dr.	Jean-Luc Cochard Jean-Luc.Cochard@idiap.ch	research director until October 97
Mr.	Andrei Constantinescu Andrei.Constantinescu@idiap.ch	research assistant until September 97
Dr.	Emile Fiesler Emile.Fiesler@idiap.ch	research director until March 97
Mr.	Frank Formaz Frank.Formaz@idiap.ch	research engineer +41 – 27 – 721 77 28
Mr.	Philippe Fu Philippe.Fu@idiap.ch	research assistant until March 97
Mr.	Arnaud Gaudinat Arnaud.Gaudinat@idiap.ch	research assistant until September 97
Mr.	Dominique Genoud Dominique.Genoud@idiap.ch	research assistant +41 – 27 – 721 77 26
Mr.	Hervé Glotin Herve.Glotin@idiap.ch	research assistant +41 – 27 – 721 77 33
Mr.	Frédéric Gobry Frederic.Gobry@idiap.ch	research assistant +41 – 27 – 721 77 31

Ms.	Astrid Hagen Astrid.Hagen@idiap.ch	research assistant +41 – 27 – 721 77 34
Mr.	Cédric Jaboulet Cedric.Jaboulet@idiap.ch	research engineer until November 97
Ms.	Katrin Keller Katrin.Keller@idiap.ch	research assistant +41 – 27 – 721 77 37
Mr.	Christopher Kermorvant Christopher.Kermorvant@idiap.ch	research assistant +41 – 27 – 721 77 46
Mr.	Sacha Krstulović Sacha.Krstulovic@idiap.ch	research assistant +41 – 27 – 721 77 43
Dr.	Jürgen Lüttin Juergen.Luettin@idiap.ch	research director, Computer Vision group leader +41 – 27 – 721 77 27
Dr.	Gilbert Maître Gilbert.Maitre@idiap.ch	research scientist +41 – 27 – 721 77 30
Mr.	Johnny Mariéthoz Johnny.Mariethoz@idiap.ch	research engineer +41 – 27 – 721 77 44
Dr.	Eddy Mayoraz Eddy.Mayoraz@idiap.ch	research director, Machine Learning group leader +41 – 27 – 721 77 29
Mr.	Guillaume Melin Guillaume.Melin@idiap.ch	research engineer until May 97
Mr.	Perry Moerland Perry.Moerland@idiap.ch	research assistant +41 – 27 – 721 77 32
Mr.	Miguel Moreira Miguel.Moreira@idiap.ch	research assistant +41 – 27 – 721 77 45
Dr.	Andrew Morris Andrew.Morris@idiap.ch	research scientist +41 – 27 – 721 77 35
Dr.	Indu Saxena Indu.Saxena@idiap.ch	research scientist until March 97
Dr.	Georg Thimm Georg.Thimm@idiap.ch	research scientist +41 – 27 – 721 77 39

2.2 Visitors

Prof.	Ethem Alpaydin Ethem.Alpaydin@idiap.ch	Bogaziçi University, Istanbul, Turkey from February to June 1998
Dr.	Sergio Curinga Sergio.Curinga@idiap.ch	University of Genova, Italy from September to December 97
Mr.	Stéphane Dupont Stephane.Dupont@idiap.ch	École Polytechnique de Mons, Belgium from April to June 97
Mr.	Christophe Ris Christophe.Ris@idiap.ch	École Polytechnique de Mons, Belgium from June to July 97
Mr.	Patrick Verlinde Patrick.Verlinde@idiap.ch	École Royale Militaire, Brussels, Belgium from July to September 97

2.3 Students

Mr.	Jean-Luc Beuchat	Jean-Luc.Beuchat@idiap.ch	until February 97
Mr.	Beat Fasel	Beat.Fasel@idiap.ch	+41 – 27 – 721 77 41
Ms.	Rachel Fournier	Rachel.Fournier@idiap.ch	until June 97
Ms.	Ana Merchan	Ana.Merchan@idiap.ch	+41 – 27 – 721 77 25
Mr.	Tomas Lundin		July 96 to February 97
Mr.	Cédric Roserens	Cedric.Roserens@idiap.ch	July to August 97
Mr.	Michael Schmal	Michael.Schmal@idiap.ch	July to September 97
Mr.	Samuel Vannay	Samuel.Vannay@idiap.ch	July 97 to January 98
Mr.	Robbert Visscher		May 96 to February 97

2.4 Administrative Staff

Ms.	Sandrine Dély-Schindler	annotator
	Sandrine.Dely@idiap.ch	+41 – 27 – 721 77 37
Ms.	Sophie Loye	interviewer
		from October to November 97
Mr.	Pierre Friche	librarian
	Friche97@uni2a.unige.ch	+41 – 27 – 721 77 21
Ms.	Sylvie Millius	secretary
	Sylvie.Millius@idiap.ch	+41 – 27 – 721 77 21
Ms.	Nadine Rousseau	secretary
	Nadine.Rousseau@idiap.ch	+41 – 27 – 721 77 21

3 Research Activities

3.1 Speech Processing Group

3.1.1 Research Focus

The Speech Processing group is working on all aspects of automatic speech recognition and speaker verification, including fundamental research and development of prototype systems. Its research activities can be described along the following main axes:

- Automatic speech recognition (ASR) for isolated words and continuous speech, and keyword spotting techniques.
- Speaker recognition/verification over the telephone.
- Development and management of large databases of speech samples required to test our research results and to develop application prototypes.
- Development and testing of prototype systems.

Some of our most recent research and development activities related to these themes are briefly discussed below.

Automatic Speech Recognition
In 1997, the work in automatic speech recognition mainly focused on (1) improving speech unit models, (2) research, development and testing of large vocabulary speech recognition systems (involving pronunciation dictionaries and grammars), (3) prototype development and technology transfer. More specifically:

1. Speech unit models — Development of better phonetic models based on hidden Markov models (HMM) and/or artificial neural networks (ANN) and which are more robust to speech and channel characteristics. Ideally, these models should also be more independent of the training databases, allowing for more flexible application developments (i.e., easy adaptation of the ASR system to a new lexicon and a new grammar). This research is performed on the basis of internal state-of-the-art software as well as external software (HTK) as a reference point. Among other research areas, we can mention here: further development of hybrid HMM/ANN systems, multiband speech recognition, multiple time scale processing, confidence level, and music/speech detection.

2. Development and testing of large vocabulary speech recognition systems, involving: software development (in collaboration with our partners from different EC projects), software exchange and adaptation to local databases, lexica and learning of lexica, and training and testing (stochastic) of language models (as well as language models adaptation).

3. Prototype development — In collaboration with Swisscom, IDIAP defined and developed application prototypes involving speech recognition (and speaker verification) technology. For example, interactive voice server prototypes have been developed to assess the technology as well as the possible applications.

Most of these developments have been carried out in the framework of formal international (EC) and national (SNSF and CTI) projects. The group also benefited from extensive (formal and informal) collaboration (and sometime complementary expertise) with other research labs like the International Computer Science Institute (ICSI, Berkeley, partly funded by the SNSF), Faculté Polytechnique de Mons (FPMs, Mons, Belgium), and EPFL.

Speaker Recognition and Speaker Verification

Speaker verification at IDIAP is mainly concerned with the improvement of the current state-of-the-art algorithms and the development of innovative solutions combining concurrent and/or complementary strategies. This work is mainly carried on in the framework of one national and two European projects (CAVE and COST 250) described below. Most of the research is performed in the context of 3 existing databases of speech samples designed especially to address the problem of inter- and intra-speaker variability (Polyvar, Polycode, Polycost). In 1997, IDIAP also participated to the international NIST (National Institute of Standards and Technology, USA) evaluation and showed that their system was at the leading edge of the technology in text independent speaker verification. IDIAP is also member of an European consortium (ELISA) created to exchange the know-how in text independent speaker verification between academic partners. IDIAP will lead, for the ELISA consortium, the fusion tests of the 1998 NIST evaluation. In 1997 also, technology transfer in the text dependent speaker verification domain, was conducted for external companies (ASCOM-TECH).

Speech Databases

The importance of publicly accessible speech databases for research and development is now widely recognized. Many speech databases are now publicly available from LDC (Linguistic Data Consortium, USA) or ELRA (European Language Resource Association, EC). It is however often crucial to extend these databases to cover various purposes including the development of robust techniques and establishment of common methods of assessing basic algorithms and system performance. Consequently, since 1992, IDIAP has been recording and managing large speech databases over telephone lines. To manage and use these large databases, a set of software packages has been developed. On top of completing the collection and annotation of some standard telephone databases, a GSM database has also been collected to initiate further research on speech recognition in non-ideal environments. However, we have to keep in mind that it will never be possible to collect enough data to cover all the acoustic and task variabilities. Therefore, task-specific database collection is not the ideal way to deal with the variability problem. In the future, speech and text (as well as video) databases should be collected with the purpose of learning about the sources of variabilities so that algorithms can be designed to identify and properly handle such variabilities.

3.1.2 Research Grants

◇ CAVE – Speaker Verification in Banking and Telecommunication

Funding: European project LE 1930, Telematics Program, supported by OFES

Duration: December 95 – November 97

Partners: Dutch PTT Telecom (NL), Ubilab (CH), Vocalis (GB), IDIAP (CH), ENST (FR), KTH (SE), KUN (NL), Swisscom (CH)

Contact: Prof. Hervé Bourlard

Staff: Gilles Caloz, Dominique Genoud, Cédric Jaboulet, Guillaume Melin.

Description: CAVE addresses one of the key issues in telematics transaction services, namely speaker verification to provide secure transactions. Its goal is twofold: (1) improving current state-of-the-art technology and (2) assessing this technology in the framework of real application as well as its acceptability by the users.

In this project, IDIAP is primarily a technology provider and its involvement is mainly research oriented. IDIAP is thus mainly working on improving the performance of the speaker verification module, as well as on providing procedures and databases to allow objective tests of speaker verification systems.

◇ SpeechDat II – Spoken Language Resources Dissemination

Funding: European project LE2-4001, Telematics Program, supported by OFES
Duration: March 96 – February 98
Partners: Aalborg University (DK), British Telecom (UK), European Commission (L), CSELT (I), Tampere Univ. of Technology (FIN), ELRA (F), GEC-Marconi Ltd (UK), GPT Ltd (UK), IDIAP, INESC (P), Knowledge S.A. (GR), Kungl Tekniska Hogskolan (S), Lernout & Hauspie Speech Products (B), Matra Communication (F), Philips (NL), Philips (D), Portugal Telecom (P), Siemens AG (D), Speech Processing Expertise Centre (NL), Swisscom (CH), Telenor R&D (N), Univ. of Maribor (SL), Univ. München (D), Univ. of Patras (GR), Univ. Politecnica de Catalunya (E), Vocalis Ltd (UK)
Contact: Prof. Hervé Bourlard
Staff: Gilles Caloz, Andrei Constantinescu
Description: Due to the progress reached in speech processing technology, more and more powerful voice driven teleservices can be implemented which allow easy access to information services (e.g. train table information), transaction services (e.g. home shopping), and call processing services (e.g. voice mail handling) via the tele-network. Many European companies are active in the field of creating such services and delivering the required speech technology. However, for research purposes, as well as the implementation of the speech processing technology (speech recognition and speaker verification), spoken language resources (including speech databases, lexica, and related tools) are necessary.

The current project aims at producing, standardizing, evaluating and disseminating large speech databases realizing a large coverage of languages (most European languages) and applications. In the framework of this project, IDIAP participates to the specification of these databases and (together with ETH) is responsible for collecting and labeling the data in Swiss French and Swiss German.

◇ COST 249 – Automatic Speech Recognition over the Telephone

Funding: European project, COST action, supported by OFES
Duration: October 95 – September 98
Partners: Belgium, Switzerland, Czech Republic, Germany, Denmark, Spain, France, Greece, Hungary, Italy, Lithuania, Norway, The Netherlands, Portugal, Poland, Sweden, Slovakia, Slovania, United Kingdom
Contact: Prof. Hervé Bourlard
Staff: Giulia Bernardis, Hervé Glotin, Johan Andersen
Description: This collaborative COST action aims at improving state-of-the-art speech recognition systems over telephone lines. This is a very broad project, addressing all important aspects of continuous speech recognition systems, including:

1. Concept establishment: overall system configuration, task complexity and dialog modeling.
2. Linguistic processing: lexical knowledge, parsing strategies, higher order constraints, language models, and speaker adaptation.
3. Phonetic decoding: neural networks and HMMs, task and language independence, and recognitions units.
4. Acoustic signal processing: feature extraction, noise suppression, and speech corpora.

In the framework of this COST action, IDIAP is more particularly involved in the development and improvement of acoustic decoding algorithms for continuous speech recognition over the telephone, and their integration with higher level knowledge such as phonological and syntactical constraints. At the national level, this work is mainly carried out in collaboration with ETH also involved in the current project.

◇ COST 250 – Automatic Speaker Recognition over the Telephone Network

Funding: European project, COST action, supported by OFES
Duration: October 95 – September 98
Partners: France, Italy, United Kingdom, Sweden, The Netherlands, Spain, Portugal, Ireland, Denmark, Greece, Slovenia, Switzerland, Turkey
Contact: Dominique Genoud
Staff: Dominique Genoud, Gilles Caloz
Description: This collaborative COST action aims at: (1) studying the technology, the economical and social feasibility of the use of Automatic Speaker Recognition /Verification technologies from speech utterances, (2) analyzing in detail the applications in telecommunications, (3) obtaining the needed databases to finalize and evaluate the automatic speaker recognition, (4) completing and transmitting between European laboratories the know-how in this field, and (5) elaborating demonstration prototypes of these technologies.

In 1996, a speaker verification database was created for the COST 250 project. About 130 people calling from 13 different countries were recorded. The recording, processing, annotation, and distribution of the Database has been performed by IDIAP, EPFL, and KTH. This database was distributed for free to each partner of the project.

In 1997, in the framework of this COST project, IDIAP participated in the last international NIST (National Institute of Standards and Technology, USA) evaluation and showed that their technology was at the leading edge in that field.

◇ Likelihood ratio adjustment for the compensation of model mismatch in speaker verification

Funding: Swisscom
Duration: July 96 – March 97
Partners: Swisscom, ENST (F)
Contact: Dominique Genoud
Staff: Dominique Genoud
Description: The problem of threshold setting in speaker verification is crucial for deployed systems. In this project we show that, under relatively weak hypotheses, the log Likelihood Ratio follows a Gaussian distribution, the moments of which can be estimated from the frame-by-frame likelihood. A consequence of this property is the possibility to design a simple procedure for adjustment of the decision threshold depending on the speaker, on the risk conditions, and on the length of the test utterance. We show experimentally the benefit that can be gained from such a procedure.

◇ Automatic French speech recognition on Unix workstations with SwissNet Connection

Funding: CTI (formerly CERS)
Duration: September 95 – August 97
Partners: Sun Microsystems (CH), aComm (CH), Swisscom, EPFL, IDIAP
Contact: Olivier Bornet
Staff: Olivier Bornet
Description: This project considers applications of automatic speech recognition and speaker verification techniques in developing efficient Interactive Voice Servers (IVS) for Integrated Services Digital Network (ISDN) based communication systems. The goal of the project is to make available basic technologies for automatic speech recognition and speaker verification on multi-processor SunSPARC workstations and Swissnet (ISDN) platforms to industrial partners and particularly to Swiss industry for Swiss French.

The developed algorithms provide the necessary tools to design and implement workstation oriented voice messaging and voice response demonstrators for telephone quality Swiss French. The speech recognition algorithms are based on a speaker independent flexible vocabulary technology and speaker verification is performed by a number of techniques executed in parallel, and combined for optimal decision.

The recognition results validate the flexible vocabulary approach which offers the potential to build word models for any application vocabulary from a single set of phonetic sub-word units trained on the Swiss French Polyphone database (see the SpeechDat project). This project resulted in many encouraging real-time demonstration systems. Consequently, the possibility of a follow-up project, as well as the start-up of a small company exploiting the results of this project, are currently being investigated.

◇ $\mathsf{ETC}_{\mathsf{v\acute{e}rif}}$: a system to help labeling of speech corpora

Funding: Swiss National Science Foundation, FN 20-43 494.95

Duration: November 95 – October 97

Contact: Dr. Jean-Luc Cochard

Staff: Philippe Fu, Arnaud Gaudinat, Hervé Glotin

Description: $\mathsf{ETC}_{\mathsf{v\acute{e}rif}}$ is addressing the problem of verification of speech utterances described in terms of the acoustic signal and the (supposedly) associated text. This is an important problem, generally related to the training of state-of-the-art systems and for which the appropriate sub-units of representation (e.g., phones) have to be defined.

The $\mathsf{ETC}_{\mathsf{v\acute{e}rif}}$ system that has been tested at IDIAP is implemented as a multi-agent system, and is based on a general purpose platform called ETC, for cooperative treatment environment ("Environnement de Traitement Coopératif").

◇ Enhanced automatic speaker recognition in telephony

Funding: Swiss National Science Foundation, FN 21-45 624.95

Duration: April 96 – December 98

Contact: Dominique Genoud

Staff: Dominique Genoud

Description: This research project aims at performing more fundamental research in speaker recognition and speaker verification, including:

- Analysis of intra- and inter-speaker variability, and selection of better parameters for speaker characterization.
- Development of most suitable algorithms specific to speaker verification tasks.
- Development of adaptive environment techniques (noise, transmission channel, ...), and a decision taken from complementary or cooperative tokens.
- Evaluation of speaker recognition technology with regards to or in synergy with other biometrical technologies.

◇ ESA-SPACT

Funding: European Space Agency (ESA), Noordwijk (NL)
Duration: January 96 – August 97.
Partners: ESTEC (NL), NLR (NL), ORIGIN/BSO (NL), TNO-TM (NL), TCD (IR), IDIAP (CH)
Contact: Luis Miguel Moreira
Staff: Luis Miguel Moreira, Gilles Caloz
Description: A recent development for the task support of crew in space missions is the Advanced Crew Terminal (ACT), a laptop computer offering crew support applications. This project aims at extending the ACT and its applications, with a speech I/O interface. Such interface will allow the crew: to command the ACT applications by voice, thus allowing him/her to obtain the required information in a hands-busy situation, to command the ACT to read out text or other information by synthesized voice, thus accessing information in an eyes-busy situation. The main objectives of this project are:

- To build/implement a flight Speech I/O equipped Advanced Crew Terminal which supports the Smart Gas Sensor experiment, and
- to evaluate the performance of the implemented system during the MIR mission in August 1997.

◇ Polyphone-CHall – Recording and annotation of a German speech corpus

Funding: Swisscom
Duration: 15 months, December 95 – April 97
Partner: TIK/ETH Zürich
Contact: Dr. Jean-Luc Cochard
Staff: Gilles Caloz, Andrei Constantinescu
Description: Based on the experience gained by IDIAP with the Swiss-French Polyphone project, IDIAP has been helping TIK/ETH for the Swiss German speech corpus. This includes: content specification, sheet preparation, speaker recruiting, speech recording. More specifically, the tasks dedicated to IDIAP concern: (1) the preparation of the prompting sheets (to guarantee maximum statistically balanced speech material) and (2) adaptation of the IDIAP software tools used for text preparation and for speech annotation to the new requirements.

◇ SWISSCOM: Advanced Vocal Interfaces Services

Funding: Swisscom
Duration: 1 year, January – December 97
Partners: Swisscom
Contact: Prof. Hervé Bourlard
Staff: Olivier Bornet, Gilles Caloz, Johan Andersen
Description: The main goal of this project was to provide Swisscom with the necessary baseline technology to develop advanced vocal intercase services. During its first year, this project involved several research and development directions, including:

- Research on linguistic modeling of spontaneous requests, involving collaboration with other Swiss institutions like EPFL (Lausanne) and ISSCO (Geneva).
- Setting up a first reference system for Swiss-French continuous speech recognition, with initial tests on the (Swiss-French) Polyphone database (containing prompted sentences and simulated spontaneous 111 requests).
- Management and distribution of the Swiss-Polyphone database, with extension to and preliminary testing on GSM data.

3.1.3 Projects granted in 1997

◇ SOCRATES – European Masters in Language and Speech

Funding: European Project, DG XXII

Duration: September 97 – September 2000

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

Contact: Prof. Hervé Bourlard

Description: The purpose of this project is to organize an advanced course (recognized as European Masters) allowing students to qualify for multidisciplinary team-working in the language industries. Besides in depth knowledge of Speech Science, Natural Language Processing or Computer Science, that has been provided by undergraduate studies, the student will obtain through this Masters the contextual knowledge from the fields that were not part of his/her specialization.

As the only Swiss representative, IDIAP has the objective to create a center of excellence in the domain of Speech Processing for graduated students, which would also be part of a large European teaching network. At the European level, this would cover well defined common courses (taught in every participating countries), as well as specialized courses and research projects (in countries where special expertise has been identified).

◇ THISL – THematic Indexing of Spoken Language

Funding: European project, ESPRIT Program, Long Term Research supported by OFES

Duration: 3 years, February 97 – January 2000

Partners: IDIAP, Sheffield University (UK), Cambridge University (UK), Thomsom (FR), BBC (UK), ICSI (Berkeley, CA).

Contact: Prof. Hervé Bourlard

Staff: Johan Myhre Andersen, Dr. Jürgen Lüttin, Giulia Bernardis

Description: The objective of THISL is to show the feasibility of integrating state of the art *Natural Language Processing* (NLP) and *Large Vocabulary Continuous Speech Recognition* (LVCSR) technologies, towards advanced multimedia applications. In this framework, the present proposal will focus on R&D aimed at retrieving multimedia information (written or spoken text) using a spoken language interface. Most of the tests will be performed on recordings of BBC broadcast news.

The expected result of the project is a real-time prototype system for navigating in the sound-track of a TV news broadcast. Significant intermediate results will include transcription of broadcast speech, development of audio editing tools, content-based retrieval from audio/video archives, and a robust spoken language interface for search and retrieval of multimedia data.

◇ SPHEAR – SPeech and HEARing

Funding: European project, European DGXII TMR Research Network, supported by OFES
Duration: March 1998 - February 2002
Partners: Ruhr-Universität Bochum (Germany), Ericsson Eurolab (Germany), Institut National Polytechnique de Grenoble (F), University of Keele (UK), University of Patras (GR).
Contact: Prof. Hervé Bourlard
Staff: Astrid Hagen, Christopher Kermorvant
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 (recently introduced by Hervé Bourlard) based on multi-band and multi-stream processing.

◇ PICASSO – PIoneering Caller Authentication for Secure Service Operation

Funding: European project, Telematics project from DGIII, supported by OFES
Duration: March 1998 – February 2001
Partners: IDIAP, Ubilab (CH), Swisscom (CH), ENST (F), IRISA (F), PTT-Telecom (NL), KPNResearch (NL), KUN (NL), Fortis (NL), KTH (SE), Telia (SE), and Vocalis (UK).
Contact: Prof. Hervé Bourlard
Staff: Giles Caloz, Johnny Mariethoz
Description: PICASSO builds upon the work done in the CAVE project which has advanced speaker verification technology and conducted security experiments with a range of prototype implementations. Work within CAVE found that there is often a trade-off between the level of security provided, which deters fraud, and the usability of the system. Voice based verification poses little hardware device problem for users, as all that is required is a standard telephone. The number of telephones available is also being swelled by enormous growth in the mobile and GSM markets.

PICASSO will integrate verification with Automatic Speech Recognition (ASR) to develop a new generation of telephone enquiry systems which combine high-accuracy customer verification with easy-to-use interfaces which understand speech. The first applications of the project results will be telephone calling cards/accounts, a messaging service ('voice mail') and retail banking.

◇ MULTICHAN – Non-stationary multichannel signal processing

Funding: Swiss National Science Foundation
Duration: January 98 – December 99
Contact: Prof. Hervé Bourlard
Staff: Dr. Andrew Morris, Katrin Keller
Description: The purpose of the present project is to investigate a new multichannel signal processing technique which has recently shown much promise in the framework of multiband speech processing. In multiband speech recognition, the frequency range is split into several bands, and information in the bands is used for phonetic probability estimation by independent modules. These probabilities are then combined for recognition later in the process at some segmental level. This multiband paradigm is motivated by psycho-acoustic studies, and by its potential robustness to noise. However, research in this potentially important new approach is still preliminary. The current project will thus start by investigating further important issues related to this approach, including trade-offs between segment choices, features, and recombination approaches. Furthermore, it seems that the same multichannel paradigm could also be used to address the problem of multiple time scale analysis (e.g., towards incorporating multiple time scale information) in current ASR system.

The multichannel approach considered here is a pretty new research area. It is however already attracting a lot of interest and could have an important impact not only on speech recognition research but also on many problems dealing with non-stationary temporal signals.

◇ ARTIST – Articulatory Representations To Improve Speech Technologies

See Section 3.2.3.

3.2 Computer Vision Group

3.2.1 Research Focus

Computer vision deals with the automatic analysis and interpretation of visual scenes. Research areas investigated within the group address general computer vision topics and application themes defined by IDIAP. Research topics include:

- Object recognition
- Motion analysis and recognition
- Sensor fusion
- Document analysis and recognition

These topics were investigated to address both fundamental research issues and application driven problems in the areas of human-computer interaction (visual speech recognition, hand-writing recognition), access security (multimodal person authentication), and X-ray image sequence analysis.

Object recognition
Object recognition represents one of the fundamental problems in computer vision and deals with the recognition and classification of visual objects. A sub-problem is concerned with the detection of objects of known classes. We have investigated and developed an algorithm for the fast detection of objects which is based on an MLP and which exploits the computational efficiency of an FFT. In comparison to conventional approaches, the method leeds to a speed up factor of about 8 - 16, depending on image and object size, while retaining identical performance levels.

We have applied the object detection algorithm to the problem of face detection which is a pertinent yet difficult task in several vision applications, e.g. face recognition, visual speech recognition, image and video indexing, video conferencing, and video coding. Face detection experiments that were performed on difficult test images of natural scenes, containing several subjects, have shown good performance of the system. The algorithm is currently being implemented for a real-time face detection system.

Motion analysis and recognition
Our work in motion analysis has been concerned with the problem of visual contour tracking and we have developed a technique which enables the tracking of deformable objects. The work has addressed the problem of lip-tracking which constitutes a difficult task due to the large appearance variability.

We have also applied the tracker to the analysis of X-ray videos which are difficult to process due to different kind of noise sources and occlusions. The objective of this research is to provide detailed articulatory data that will lead to a better understanding and insight into human speech production which is believed to lead to significant improvements in current speech recognition and verification technology.

Research in motion recognition has concentrated on visual speech recognition (lip-reading) and on motion-based person authentication. We have developed one of the first speaker independent continuous lip-reading systems which achieved performance levels similar to the ones of human lip-readers with no lip-reading knowledge. We have also introduced a novel technique for person authentication, based on visual spatio-temporal models of talking persons. Both methods were extensively evaluated on large audio-visual databases.

Sensor fusion
Sensor fusion is a powerful solution to pattern recognition problems involving large sets of classes and noisy input since it allows the simultaneous use of different information sources. Typical problems which arise in our research topics are the combination of different information sources (acoustic, visual, articulatory) in speech recognition and the combination of different modalities (face, profile, speech, motion) in person authentication.

The group has investigated and developed several novel methods of sensor fusion. Several of them address the fusion of multiple classifiers in the case of person verification. We have developed a technique for audio-visual person verification which combines acoustic information with temporal visual information from the talking face. In collaboration with the Machine Learning Group we have investigated the use of binary classifiers to combine identity information from different classifiers. We have also implemented a PC-prototype version of an acoustic-only person authentication system which has been tested locally within the institute.

Another sensor fusion method was concerned with the fusion of asynchronous, possibly noisy, data for the purpose of audio-visual speech recognition. In this case we developed a method based on the Multi-Stream approach (see Speech Group) which enables the synchronous decoding of audio-visual speech but which can also account for asynchrony between the two modalities. The system has been tested for audio-visual speech recognition tasks and has shown to considerably improve the performance of acoustic-only systems when background noise is present.

Document analysis and recognition

This subject is concerned with the recognition of machine- printed, hand-printed, and hand-written documents. Most of our work has been concentrated on hand-printed character recognition, namely visual feature representations and similarity measures. We have developed a new feature representation method based on rotated histograms which was combined with an MLP based recognition system. This work was performed as part of a diploma thesis and has been granted the "NCR Award".

One of the fundamental problems in handwriting recognition is the segmentation of characters. This problem is similar to the segmentation of phones in speech recognition and can be addressed by joint segmentation and recognition using stochastic methods. This strategy also allows the incorporation of multiple knowledge sources such as lexica, language models, and the topic of the document. We have started to apply these techniques to the recognition of ancient script documents.

3.2.2 Research Grants

◇ M2VTS – Multimodal Verification for Teleservices and Security Applications

Funding: European project AC 102, ACTS Program, supported by OFES

Duration: October 95 – September 98

Partners: Matra Communication (F), Cerberus AG (CH), Ibermática S.A. (E), Ecole Polytechnique Fédérale de Lausanne (CH), Université de Neuchâtel (CH), Université Catholique de Louvain (B), University of Surrey (GB), Renaissance (B), Aristotle University of Thessaloniki (GR), Compagnie Européenne de Télésécurité (F), Universidad Carlos III (E), Banco Bilbao Vizcaya (E), Unidad Tecnica Auxiliar de la Policia (E)

Contact: Dr. Jürgen Lüttin

Staff: Dr. Souheil Ben-Yacoub, Dr. Gilbert Maître

Companion projects: CAVE, COST 250 (see Section 3.1.2)

Description: The primary goal of the M2VTS project is to address the issue of secured access to local and centralized services in a multi-media environment. The main objective is to extend the scope of application of network-based services by adding novel and intelligent functionalities, enabled by automatic verification systems combining multimodal strategies (secured access based on speech, image and other information). The objectives are also to show that limitations of individual technologies (speech recognition, speaker verification...) can be overcome by relying on multi-modal decisions (combination or fusion of these technologies) and can find practical and important applications in the new emerging fields of advanced interfaces for tele-services.

3.2.3 Projects granted in 1997

◇ ARTIST – Articulatory Representations To Improve Speech Technologies

Funding: Swiss National Science Foundation
Duration: April 97 – March 99
Contact: Dr. Jürgen Lüttin
Staff: Sacha Krstulović, Dr. Georg Thimm
Description: This research project aims at using articulatory features in speech recognition (SR) and speaker verification (SV) applications. Such features are believed to lead to significant improvements of SV/SR systems, in accordance with the statements of Liberman's "Motor Theory of Speech Perception", with European ACCOR project's results, and with several other studies.

Subtasks involved in this research include :

- Automatic segmentation of an X-ray video database displaying the vocal tract by means of computer vision techniques. This will provide a set of matched acoustic/articulatory data suitable for the training or validation of acoustic-to-articulatory conversion schemes.
- Implementation of robust acoustic-to-articulatory conversion methods. This will enable the extraction of articulatory features from a sound input, thus making the use of articulatory features compatible with existing SV/SR systems.
- Use of extracted articulatory feature in SR/SV systems. This will validate the original concept of "Motor Speech Perception" and improve the existing SV/SR applications' performances.

◇ AV-COM – Audio-Visual Combination

Funding: Swiss National Science Foundation, "R'EQUIP" program for new equipment.
Contact: Prof. Hervé Bourlard
Description: New computer facilities (CPU server and real time audio-video acquisition with audio-video synchronization) for research into multi-modal audio-video processing.

The objective of the present R'EQUIP research grant is to provide IDIAP with the necessary computer resources to carry out multimodal signal processing research. This new equipment will support the realization of new projects in multimodal data processing and accelerate the accomplishment of existing projects in the audio-visual domain. These projects represent individual modules in the area of multimodal human computer interfaces.

3.3 Machine Learning Group

3.3.1 Research Focus

In its broad sense, machine learning means inference of a computational model from samples. The technologies studied, elaborated and experimented in our group, range from neural networks, to logical analysis of data, passing by classical AI methods (e.g. decision trees) and statistical approaches (e.g. Bayesian learning, hidden Markov chains). In its spirit, the activity of our group is always application driven, thus the learning techniques involved are not studied for themselves, but are always targeted to a specific application. Among the huge variety of potential applications of automated learning, those arising in the other two groups of IDIAP are favored:

- Classification problems characterized by a large number of classes and noisy data.
- Fusion of classifiers.

The main research directions not immediately connected to the other groups are:

- Computer assisted diagnosis.
- Time series prediction.

Classification among numerous classes
In perceptual AI, several problems are expressed as classification problems, involving a large number of classes: 10 digits, 26 letters, from 30 to 60 phonemes, N potential users of a system with access security, etc. Many classification methods are able to deal with more than two classes but quite often, their efficiency decreases rapidly with the number of classes. Other methods are restricted to discrimination between two classes only.

There is a very fruitful approach to handle problems with a large number of classes, consisting in decomposing the general problem into many subproblems involving a small number of classes or two classes only. The advantages of such approaches are numerous: each subproblem is quite simple; a large variety of learning methods can be used to solve the subproblems; since a class in a subproblem gathered usually several classes of the whole problem, the sample for each class are much larger; some redundancy among the subproblems provide some robustness to the global model; both in the training and in its usage, the process can easily be parallelized.

During the year 1997, we have been very active in this area and we proposed new methods for elaborating the decomposition into subproblems as well as strategies of reconstruction, turning the answers of all subproblems into one decision of class membership.

We experimented the strategy consisting in decomposing a problem into several subproblems to solve a text-dependent speaker verification problem and here again the adequate decomposition techniques have been shown to be very efficient.

Fusion of classifiers
When a classification problem is decomposed into several subproblems (see above), It has been demonstrated that it is often interesting to use learning techniques of different nature for the resolution of the subproblems. The same idea applies for simple classification problems. Even in the case of a two-classes discrimination problem, several subproblems can be created, either based on different data (see 3.2.1), or on a resampling of the training data (bagging, boosting, arcing), or learned by various learning techniques. The problem is then to fuse different decisions coming from each classifier into one final decision. In the most interesting case, the decision of each classifier is not just a class membership, but it also includes a confidence or a probability for this membership.

We carried out some common work with the Computer Vision group in multi-modal speaker identification systems along this line. Different classifiers based on different data (face, lip motion, voice) representing users of the system had to be fused into a single answer (see 3.2.1).

Computer assisted diagnosis

In applications related to human sciences (such as medical, social, economical issues), if it is desirable that the model provided by an automated learning system is reliable, it is even more important for this model to be easily understandable for a human expert of the field. Indeed, a medical doctor using computer assisted diagnosis system will be much more inclined to trust such a system if this one can give, for any of its answers, a rational justification that can be expressed in a language that makes sense to the doctor.

Logical Analysis of Data (LAD) is a new technique which can be used to learn from data with as essential feature this easiness of interpretation. Developing, improving and extending LAD is one of our main activity in the Machine Learning group. So far, we are still doing fundamental research along this line, but in a near future we are planing to apply it to very concrete and specific problems related to computer assisted diagnosis.

Time series prediction

As explained in Section 1.2, the general goal of the Machine Learning group is two-fold. First, it offers technical support for the two other groups, since most of the fundamental problems arising in perceptual AI are learning problems. Second, it also has a role of prospecting new fields of research, where the available technology could be successfully applied. Along this line, a new project in the Machine Learning group is initiated, dealing with time series prediction (TSP).

In this project, the usability for general TSP of well mastered technology in speech processing (hidden Markov models (HMM), hybrid HMM model and neural networks) will be evaluated. Mixture of experts of neural networks have been shown lately to be very efficient for certain types of time series (prevision of electrical demand). The use of these models will be further investigated for other types of time series (financial or environmental). Classically, in a mixture of expert model, each expert as well as the gate are neural networks. We will also investigate models where the experts and the gate are HMMs.

3.3.2 Research Grants

◇ Adaptive Optical Multilayer Neural Networks

Funding: Swiss National Science Foundation

Duration: October 95 – September 97

Partners: Institute of Microtechnology (IMT) of the University of Neuchâtel

Contact: Perry Moerland

Staff: Dr. Indu Saxena

Description: Optical neural networks provide a very promising alternative to electronic systems in the parallel handling of data, especially since currently electronic systems approach their physical (speed) limits. This is especially important in cases where real-time or massively parallel information processing is desired, such as computer vision and speech recognition.

This project involved the design and implementation of a large modular optical multilayer neural network. By using an optical device as non-linearity, a unique all-optical recall was established.

◇ Compact Hardware-Friendly Neural Networks

Funding: Swiss National Science Foundation, FN 21-45621.95

Duration: April 96 – March 99

Contact: Perry Moerland

Staff: Frederic Gobry, Tomas Lundin and Perry Moerland

Description: The aim of this project is the study and the optimization of artificial neural networks in order to ease their hardware implementation. According to the original plan, this improvement had to be done at three levels : the neuron (considering the neural function effectively realized by an analog implementation), the interconnection (understanding the influence of discrete weights with a low quantization level simulating a cheep numerical storage), and the topology (highly interconnected networks are not suitable for hardware representation).

The first two aspects have been considered during the first year of the project and now the focus has been placed on the third one. More precisely, the research effort is concentrated on the study of a modular neural network model, known as *mixtures of experts*. This model consists of a gating network that partitions the input space and expert networks that treat the resulting subproblems. It has a nice probabilistic interpretation as a mixture model which allows a principled approach to training of mixtures of experts. Research within this project is focusing on model selection (growing, pruning, and Bayesian methods), estimation of local error bars, and links and combinations with ensemble methods (bagging, boosting, arcing etc.).

◇ GLAD – Generalization of LAD

Funding: Swiss National Science Foundation, FN 21-46 974.96

Duration: November 96 – October 98

Partners: Swiss Federal Institute of Technology (EPFL)

Contact: Dr. Eddy Mayoraz

Staff: Miguel Moreira

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

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

3.3.3 Projects granted in 1997

◇ ZEPHYR – Time Series Prediction with Hybrid Markov Models

Funding: Swiss National Science Foundation, FN 21-50744.97
Duration: January 98 – December 99
Partners: Swiss Federal Institute of Technology (EPFL)
Contact: Dr. Eddy Mayoraz
Staff: Frédéric Gobry
Description: Hidden Markov Models (HMM) have been used extensively and very successfully for speech processing for the past 20 years. Hybrid models mixing adequately HMMs and artificial neural networks are powerful tools for speech recognition. The aim of this project is to study in what extend HMMs and hybrid models can be used for time series prediction. The mixture of experts (MEs) models will also be considered, as well as hybrid models involving MEs.

This study will target different types of time series, from very chaotic ones (financial series) to more structured ones (economical series such as the prevision of electrical demand), including series where the phenomena to be triggered are sparse (floods or avalanches forecasting). For each of these applications, we will consider the most adequate model, combining HMM or mixture of experts models with some appropriate machine learning methods (ARMA models, neural networks, decision trees, LAD).

4 Educational Activities

4.1 Ongoing Ph.D. Programs

- **Ph.D. Candidate:** Johan Myhre Anderson
  Supervisor: Prof. H. Bourlard
  Research topic: Robust Speech Recognition
  University: EPFL, Lausanne
- **Ph.D. Candidate:** Giulia Bernardis
  Supervisor: Prof. H. Bourlard
  Research topic: Speech Recognition
  University: EPFL, Lausanne
- **Ph.D. Candidate:** Gilles Caloz
  Supervisor: Prof. H. Bourlard
  Research topic: Speaker Verification
  University: EPFL, Lausanne
- **Ph.D. Candidate:** Dan Andrei Constantinescu
  Supervisor: Dr. Gérard Chollet
  Research topic: Towards Language Independent Speech Recognition For Multilingual Interactive Voice Servers
  University: ENST, Paris
- **Ph.D. Candidate:** Dominique Genoud
  Supervisor: Prof. Martin Hasler, Dr. Gérard Chollet
  Research topic: Enhanced Automatic Speaker Recognition in Telephony
  University: EPFL, Lausanne
- **Ph.D. Candidate:** Hervé Glotin
  Supervisor: Prof. H. Bourlard
  Research topic: Speech Recognition Based on Psycho-Acoustic Properties
  University: EPFL, Lausanne and INPG, Grenoble
- **Ph.D. Candidate:** Frédéric Gobry
  Supervisor: Dr. Eddy Mayoraz
  Research topic: Time Series Prediction with Hybrid Markov Models
  University: EPFL, Lausanne
- **Ph.D. Candidate:** Astrid Hagen
  Supervisor: Prof. Hervé Bourlard
  Research topic: Multistream Speech Recognition
  University: EPFL, Lausanne
- **Ph.D. Candidate:** Sacha Krstulović
  Supervisor: Prof. M. Hasler
  Research topic: Using Articulatory Features for Speech Recognition / Speaker Verification
  University: EPFL, Lausanne
- **Ph.D. Candidate:** Perry Moerland
  Supervisor: Prof. W. Gerstner
  Research topic: Mixtures of experts
  University: EPFL, Lausanne
- **Ph.D. Candidate:** Miguel Moreira
  Supervisor: Prof. A. Hertz
  Research topic: GLAD – Generalization of LAD
  University: EPFL, Lausanne

4.2 Ph.D. exams

- **Ph.D. candidate:** Jürgen Lüttin
  Supervisor: Dr. Steve Renals, Dr. Neil Thacker (University of Sheffield)
  Examiners: Prof. D. Hogg and Dr. M. Cooke
  University: University of Sheffield, U.K.
  Date: June 25
  Title: Visual Speech and Speaker Recognition [o-Lue97]
  Short summary:

  This thesis presents a learning based approach to speech recognition and person recognition from image sequences.

  An appearance based model of the articulators is learned from example images and is used to locate, track, and recover visual speech features. A major difficulty in model based approaches is to develop a scheme which is general enough to account for the large appearance variability of objects but which does not lack in specificity. The method described here decomposes the lip shape and the intensities in the mouth region into weighted sums of basis shapes and basis intensities, respectively, using a Karhunen-Loéve expansion. The intensities deform with the shape model to provide shape independent intensity information. This information is used in image search, which is based on a similarity measure between the model and the image.

  Visual speech features can be recovered from the tracking results and represent shape and intensity information. A speechreading (lip-reading) system is presented which models these features by Gaussian distributions and their temporal dependencies by hidden Markov models. The models are trained using the EM-algorithm and speech recognition is performed based on maximum posterior probability classification.

  It is shown that, besides speech information, the recovered model parameters also contain person dependent information and a novel method for person recognition is presented which is based on these features. Talking persons are represented by spatio-temporal models which describe the appearance of the articulators and their temporal changes during speech production. Two different topologies for speaker models are described: Gaussian mixture models and hidden Markov models.

  The proposed methods were evaluated for lip localisation, lip tracking, speech recognition, and speaker recognition on an isolated digit database of 12 subjects, and on a continuous digit database of 37 subjects. The techniques were found to achieve good performance for all tasks listed above. For an isolated digit recognition task, the speechreading system outperformed previously reported systems and performed slightly better than untrained human speechreaders.

- **Ph.D. candidate:** Georg Thimm
  Jury members: Prof. A. Skrivervik (President), Prof. M. Kunt (Director), Dr. E. Fiesler (internal Director), Prof. C. Pelligrini, Prof. W. Gerstner, Prof. L.C. Jain
  University: Swiss Federal Institute of Technology, Lausanne, Signal Processing Laboratory
  Date: May 7 (exam) and June 20 (public defense)
  Title: Optimization of High Order Perceptrons [o-Thi97]
  Short summary:

  Neural networks are widely applied in research and industry. However, their broader application is hampered by various technical details. Among these details are several training parameters and the choice of the topology of the network. The subject of this dissertation is therefore the elimination and determination of usually user specified learning parameters. Furthermore, suitable application domains for neural networks are discussed.

Among all training parameters, special attention is given to the learning rate, the gain of the sigmoidal function, and the initial weight range. A theorem is proven which permits the elimination of one of these parameters. Furthermore, it is shown that for high order perceptrons, very small random initial weights are usually optimal in terms of training time and generalization.

Another important problem in the application of neural networks is to find a network topology that suits a given data set. This favors high order perceptrons over several other neural network architectures, as they do not require layers of hidden neurons. However, the order and the connectivity of a network have to be determined, which is possible by two approaches. The first is to remove connections from an initially big network while training it. The other approach is to increase gradually the network size. Both types of approaches are studied, corresponding algorithms are developed, and applied to high order perceptrons. The (dis-)advantages of both approaches are gone into and their performance experimentally compared.

Then, an outlook on future research on the interpretation and analysis of high order perceptrons and their feasibility is given.

Finally, high order perceptrons and the developed algorithms are applied to a number of real world applications, and, in order to show their efficiency, the obtained performances are compared to those of other approaches.

4.3 Student Projects

- **Trainee:** Jean-Luc Beuchat
  School: EPFL
  Formation: diploma thesis
  Subject: Handprinted Character Recognition with Neural Networks
  Duration: October 1996 – February 1997
  Responsibles: Dr. Gilbert Maître and Dr. Georg Thimm

- **Trainee:** Alain Dannaoui
  School: EPFL, Lausanne
  Formation: semester project
  Subject: Analysis of Geo-Statistical Data with Neural Networks
  Duration: October 1997 – February 1998
  Responsibles: Perry Moerland and Prof. Wulfram Gerstner (EPFL)

- **Trainee:** Rachel Fournier
  School: University of Fribourg
  Formation: diploma thesis
  Subject: Étude de paramètres prosodiques en reconnaissance automatique de l'allemand
  Duration: October 1995 – March 1997
  Responsible: Dr. Jean-Luc Cochard

- **Trainee:** Tomas Lundin
  School: Chalmers University of Technology, Göteborg, Sweden
  Formation: diploma thesis
  Subject: Ontogenic Neural Network Quantization
  Duration: July 1996 – February 1997
  Responsibles: Dr. Emile Fiesler and P. Moerland

- **Trainee:** Ana Merchan
  School: University of Extremadura
  Formation: diploma thesis
  Subject: A posteriori reconstruction of decomposed classification methods
  Duration: October 1997 – March 1998
  Responsible: Dr. Eddy Mayoraz

- **Trainee:** Robert Ribnitz
  School: University of Fribourg
  Formation: diploma work
  Subject: Majordome vocal interactif
  Duration: Oct. 1996 – June 1997
  Responsible: Dr. Jean-Luc Cochard

- **Trainee:** Cédric Roserens
  School: University of Lausanne
  Formation: applied work
  Subject: Data binarization using a self-organizing hypercube
  Duration: July – August 1997
  Responsible: Dr. Eddy Mayoraz

- **Trainee:** Michael Schmal
  School: Fachhochschule Trier, Germany
  Formation: applied work
  Subject: Text dependent speaker verification
  Duration: July – September 1997
  Responsibles: Dominique Genoud, Dr. Eddy Mayoraz and Miguel Moreira

- **Trainee:** Samuel Vannay
  School: EURECOM/EPFL
  Formation: applied work
  Subject: Réalisation d'un majordome vocal
  Duration: July 97 – January 98
  Responsible: Olivier Bornet and Prof. Giovanni Coray (EPFL)

- **Trainee:** Robbert Visscher
  School: University of Groningen, The Netherlands
  Formation: diploma thesis
  Subject: Superceptron Growing
  Duration: May 96 – February 97
  Responsible: Dr. Emile Fiesler

4.4 Lectures and seminars

- **Title:** Reconnaissance automatique de la langue parlée
  Speaker: Dr. Jean-Luc Cochard
  School: University of Fribourg
  Duration: academic year 1996–97, (3 hours/week)
  Audience: 3rd and 4th year optional course for students in computer science

- **Title:** Logical Analysis : A new Methodology for Data Knowledge Extraction
  Speaker: Dr. Eddy Mayoraz
  Location: Institut de Recherche en Ophtalmologie (IRO), Sion
  Date: March 14, 1997

- **Title:** Optimization of Higher Order Perceptrons
  Speaker: Dr. Georg Thimm
  Location: SSC Seminar, EPFL, Lausanne
  Date: May 1, 1997

- **Title:** Markov Models, Hidden Markov Models and Hybrid Systems for Time Series Processing
  Speaker: Prof. Hervé Bourlard
  Location: Computer Science Department, EPFL, Lausanne
  Date: May 6, 1997
- **Title:** Neural networks and conventional algorithms – Applications to speech recognition
  Speaker: Prof. Hervé Bourlard
  Location: Cambridge Neural Networks Summer School, Cambridge (UK)
  Date: September 22–24, 1997
- **Title:** Mixtures of Experts: A Divide and Conquer Approach to Pattern Recognition
  Speaker: Perry Moerland
  Location: MANTRA seminar, EPFL, Lausanne
  Date: September 29, 1997

4.5 Examinations

- **School:** EPFL/Eurecomm
  Subject: PhD thesis committee
  Expert: Prof. Hervé Bourlard
  Candidate: Philippe Gelin
  Title: Détection de mots clés dans un flux de parole: Application à l'indexation de documents multimédia
  Date: April 30, 1997
- **School:** University of Evry Val D'Essonne (Paris)
  Subject: PhD thesis committee
  Expert: Prof. Hervé Bourlard
  Candidate: Vincent Vigneron
  Title: Méthodes d'apprentissage statistiques et problèmes inverses. Application à la spectrographie
  Date: May 5, 1997
- **School:** Faculté Polytechnique de Mons (Belgium),
  Subject: PhD thesis committee
  Expert: Prof. Hervé Bourlard
  Candidate: Christophe Couvreur
  Title: Environmental sound recognition: a statistical approach
  Date: June 9, 1997
- **School:** École d'Ingénieur du Valais
  Teacher: Dr. François Corthay
  Course: Logic systems
  Expert: Dr. Gilbert Maître
  Date: October, 1997
- **School:** École d'Ingénieur du Valais
  Subject: Diploma project
  Expert: Dr. Eddy Mayoraz
  Candidate: Johnny Mariéthoz
  Date: December 15, 1997

5 Other Scientific Activities

5.1 Editorship

- **Name:** Prof. Hervé Bourlard
  Function: Editor-in-Chief
  Journal: Speech Communication
- **Name:** Dr. Emile Fiesler
  Function: Editor-in-Chief
  Book title: Handbook of Neural Computation

5.2 Scientific Committees Membership

- **Name:** Prof. Hervé Bourlard
  Function: Member of the Scientific Committee
  Conference: European Symposium of Artificial Neural Networks (ESANN)
- **Name:** Prof. Hervé Bourlard
  Function: Member of the Scientific Committee
  Conference: ESCA Workshop on Robust Speech Recognition for Unknown Communication Channels
- **Name:** Prof. Hervé Bourlard
  Function: Member of the Scientific Committee
  Conference: ESCA Workshop on Modeling Pronunciation Variation for Automatic Speech Recognition
- **Name:** Prof. Hervé Bourlard
  Function: Member of the Scientific Committee
  Society: International Association for Cybernetics
- **Name:** Prof. Hervé Bourlard
  Function: Member of the Advisory Board
  Conference: International Conference on Neural Networks, ICANN'97
- **Name:** Prof. Hervé Bourlard
  Function: Member of the Scientific Committee
  Conference: Neural Information Processing Systems (NIPS)
- **Name:** Dr. Jean-Luc Cochard
  Function: Member of the Scientific Committee
  Conference: XXIème Journées d'Étude sur la Parole, Avignon
- **Name:** Dr. Eddy Mayoraz
  Function: Member of the Editorial Board
  Conference: European Symposium of Artificial Neural Networks (ESANN)
- **Name:** Dr. Eddy Mayoraz
  Function: Member of the Program Committee and organizer of 2 special sessions on artificial neural networks
  Conference: 5th International Symposium on Artificial Intelligence and Mathematics
- **Name:** Dr. Eddy Mayoraz
  Function: Member of the Technical Committee
  Conference: International Conference on Neural Networks, ICANN'97
- **Name:** Dr. Georg Thimm
  Function: Current Events Editor
  Journal: Neurocomputing

5.3 Organization of Conferences

- **Title:** First International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA'97)
  Location: Crans-Montana, Congress and Exhibition Centre "Le Régent"
  Date: March 12 – 14, 1997
  Conference Board: *General Chairs*: Josef Bigün (EPFL, Switzerland), Gérard Chollet (ENST, France & IDIAP, Switzerland); *Publications*: Gunilla Borgefors (Sweden); *Local Arrangements*: Gilbert Maître (Switzerland); *Publicity*: Stéphane Pigeon (Belgium), Gabriella Sanniti di Baja (Italy), Luc Vandendorpe (Belgium); *Registration*: Benoît Duc (Switzerland), Stefan Fischer (Switzerland), Dijana Petrovska-Delacrétaz (Switzerland); *Industry and Research Liaison*: Eric Badiqué (European Union), Philip Lockwood (France)
  Organisers: EPFL and IDIAP
  Sponsors: IAPR, SGAICO, Ascom, Matra Communication, Sodeval S.A., Swisscom
  Technical Programme Board: Marc Acheroy (Belgium), Eric Badiqué (European Union), Martin Bichsel (Switzerland), Josef Bigün (Switzerland), Frédéric Bimbot (France), Mats Blomberg (Sweden), Gunilla Borgefors (Sweden), Louis Boves (The Netherlands), Roberto Brunelli (Italy), Hans du Buf (Portugal), Gérard Chollet (France & Switzerland), Paloma Domingo (Spain), Benoît Duc (Switzerland), Daniele Falavigna (Italy), Stefan Fischer (Switzerland), Gösta Granlund (Sweden), Jean-Paul Haton (France), Thomas S. Huang (USA), Takeo Kanade (USA), Josef Kittler (UK), Jens-Peter Köster (Germany), Philip Lockwood (France), Gilbert Maître (Switzerland), Christoph von der Malsburg (Germany), Henri Méloni (France), Bruce Millar (Australia), Andrea Paoloni (Italy), Alex Pentland (USA), Ioannis Pitas (Greece), Douglas A. Reynolds (USA), Gabriella Sanniti di Baja (Italy), Massimo Tistarelli (Italy), Saburo Tsuji (Japan), Luc Vandendorpe (Belgium), Richard Winski (UK), Hezy Yeshurun (Israel)

- **Title:** XXII^{emes} Journées D'Étude sur la Parole (JEP'98)
  Location: Martigny, Centre du PARC, Valais, Switzerland
  Date: June 15 – 19, 1998
  Conference Board: *General Chair*: Hervé Bourlard (IDIAP, Switzerland); *Secretary*: Chantal Pillet (Switzerland); *Organizers*: Olivier Bornet (IDIAP, Switzerland), Gilles Caloz (IDIAP, Switzerland), Frank Formaz (IDIAP, Switzerland), Dominique Genoud (IDIAP, Switzerland), Cédric Jaboulet (UBILAB, Switzerland), Sacha Krstulovic (IDIAP, Switzerland);
  Organisers: IDIAP and GFCP (Groupe Francophone de la Communication Parlée)
  Sponsors: SFA (Société Française d'Acoustique), ESCA (European Speech Communication Association), Swisscom, SNSF (Swiss National Science Foundation) and the City of Martigny
  Scientific Committee: M. Adda-Decker (France), R. André-Obrecht (France), F. Bimbot (France), J.-F. Bonastre (France), H. Bourlard (Switzerland), J.-L. Cochard (Switzerland), P. Dupont (France), P. Deléglise (France), J.-M. Hombert (France), Y. Laprie (France), R.K. Moore (UK), C. Montacié (France), P. Perrier (France), J. Schoentgen (Belgium), R. Sock (France), B. Teston (France), J. Vaissière (France)

5.4 Short term visits

- **Location:** RUTCOR, Rutgers University, NJ, USA
  Visitor: Dr. Eddy Mayoraz
  Date: June 24 – July 26, 1997

- **Location:** Johns Hopkins University, MD, USA
  Visitor: Dr. Jürgen Lüttin
  Date: July 14 - August 22, 1997

- **Location:** International Computer Science Institute (ICSI), Berkeley, CA, USA
  Visitor: Prof. Hervé Bourlard
  Date: March 24 – 31, July 15 – August 15, and November 14 – 23, 1997

6 Events and Presentations

6.1 Awards

In April 1997, at the International Conference on Acoustics, Speech, and Signal Processing (ICASSP'97, Munich), Prof. Hervé Bourlard received the IEEE Award for the best 1995 IEEE Signal Processing Magazine paper entitled "Continious Speech Recognition — an Introduction to the Hybrid Hidden Markov Model / Connectionist Approach" (vol. **12**-3).

6.2 Scientific Presentations

- **Event:** European Forum on Face and Voice Related Signal Processing, Crans Montana, Switzerland, March 11, 1997
  Speaker: Jürgen Lüttin
  Title: Grey-level based lip-tracking within the M2VTS project

- **Event:** ESCA-NATO Workshop on Robust Speech Recognition for Unknown Communication Channels, April 17–18, 1997
  Speaker: Hervé Bourlard [p-DBR97]

- **Event:** International Conference on Acoustics, Speech, and Signal Processing, ICASSP'97, Munich, April 21–24, 1997
  Speaker: Hervé Bourlard [p-FB97, p-BD97, p-DBD+97]

- **Event:** Workshop on SNHC - Synthetic/Natural Hyprid Coding, Milano, Italy, May 20, 1997
  Speaker: Jürgen Lüttin
  Title: Lip-Tracking for Speech-Assisted Video Processing

- **Event:** The Fourteenth International Conference on Machine Learning, ICML'97, Nashville, TN, USA, July 8–12, 1997
  Speaker: Miguel Moreira [p-MM97]

- **Event:** Research Workshop on Inovative Techniques for Large Vocabulary Conversational Speech Recognition, Baltimore, USA, July 14 – August 22, 1997
  Speaker: Jürgen Lüttin
  Title: RASTA-HMM: Relative Spectra Hidden Markov Models

- **Event:** International Symposium on Mathematical Programming, ISMP'97, Lausanne, August 24-29, 1997
  Speaker: Eddy Mayoraz
  Title: Logical Analysis of Data for Multi-Class Learning

- **Event:** NATO workshop on Adaptive Processing of Temporal Information, Salerno (Italy), September 8–17, 1997
  Speaker: Hervé Bourlard (invited keynote speaker) [p-BM97]

- **Event:** Eurospeech, Rhodos, Greece, September 22–25, 1997
  Speaker: Hervé Bourlard [p-HRBR97, p-DB97]

- **Event:** Eurospeech, Rhodos, Greece, September 22–25, 1997
  Speaker: Dominique Genoud [p-JLGW97b]

- **Event:** Eurospeech, Rhodos, Greece, September 22–25, 1997
  Speaker: Jürgen Lüttin [p-Lue97]

- **Event:** International Conference on Artificial Neural Networks, ICANN'97, Lausanne, October 7–10, 1997
  Speaker: Hervé Bourlard (invited keynote speaker) [p-Bou97]
- **Event:** International Conference on Artificial Neural Networks, ICANN'97, Lausanne, October 7–10, 1997
  Speaker: Eddy Mayoraz [p-May97]
- **Event:** International Conference on Artificial Neural Networks, ICANN'97, Lausanne, October 7–10, 1997
  Speaker: Perry Moerland [p-Moe97]
- **Event:** International Conference on Artificial Neural Networks, ICANN'97, Lausanne, October 7–10, 1997
  Speaker: Indu Saxena [p-SMFP97]
- **Event:** 1997 NIST evaluation, EPFL, Lausanne, October 13, 1997
  Speaker: Dominique Genoud

6.3 Regional Presentations

In 1997, IDIAP's activities were presented to the broad public at two regional fairs:

- La Foire du Valais, Martigny, October 6, 97;
- Le Festival International des Métiers de Montagne, Martigny, December 3 – 6, 97.

7 Publications

7.1 Books and Book Chapters

[b-Fie97] *CRC Comprehensive Dictionary of Electrical Engineering.* CRC Press, Boca Raton, Florida, 1997. Contributing Author: E. Fiesler.

[b-MF97] P. D. Moerland and E. Fiesler. Neural network adaptations to hardware implementations. In E. Fiesler and R. Beale, editors, *Handbook of Neural Computation*, pages E1.2:1–13. Institute of Physics Publishing and Oxford University Publishing, New York, 1997. IDIAP-RR 97-17.

[b-Sax97] I. Saxena. Ellipsometry. In P. K. Rastogi, editor, *Optical Metrology.* Artech House, 1997.

7.2 Articles in International Journals

[a-BKT97] M. Bader, W. E. Klee, and G. Thimm. The 3-regular nets with 4 and 6 vertices per unit cell. *Zeitschrift für Kristallographie*, 212:553–558, 1997.

[a-DBB+97] Benoît Duc, Elizabeth Saers Bigün, Josef Bigün, Gilbert Maître, and Stefan Fischer. Fusion of audio and video information for multi modal person authentication. *Pattern Recognition Letters*, 18(9):835–843, 1997.

[a-JLGW97] P. Jourlin, J. Luettin, D. Genoud, and H. Wassner. Acoustic-labial speaker verification. *Pattern Recognition Letters*, 18(9):853–858, 1997. IDIAP-RR 97-13.

[a-LT97] J. Luettin and N. A. Thacker. Speechreading using probabilistic models. *Computer Vision and Image Understanding*, 65(2):163–178, 1997. IDIAP-RR 97-12.

[a-MFS98] P. D. Moerland, E. Fiesler, and I. Saxena. Discrete all-positive multilayer perceptrons for optical implementation. *Optical Engineering*, 37(4), April 1998. IDIAP-RR 97-02.

[a-TF97a] Georg Thimm and Emile Fiesler. High order and multilayer perceptron initialization. *IEEE Transactions on Neural Networks*, 8(2), 1997.

[a-TF97b] Georg Thimm and Emile Fiesler. Two neural network construction methods. *Neural Processing Letters*, 6(1), 1997.

[a-Thi97] Georg Thimm. Calendar of meetings (several issues). *Neurocomputing*, 1997. published since 1995.

[a-TK97] G. Thimm and W. E. Klee. Zeolite cycle sequences. *Zeolites*, 19:422–424, 1997.

7.3 Articles in Conference Proceedings

[p-BD97] H. Bourlard and S. Dupont. Subband-based speech recognition. In *IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing*, pages 1251–1254, April 1997.

[p-BG97] F. Bimbot and D. Genoud. Likelihood ratio adjustment for the compensation of model mismatch in speaker verification. In *Eurospeech 97*, 1997. IDIAP-RR 97-05.

[p-BHJ+97] F. Bimbot, H-P. Hutter, C. Jaboulet, J. Koolwaaij, J. Lindberg, and J-B. Pierrot. Speaker verification in the telephone network : Research activities in the CAVE project. In *Proceedings of the European Conference on Speech Communication and Technology (EUROSPEECH'97)*, 1997.

[p-BM97] Hervé Bourlard and Nelson Morgan. Hybrid HMM/ANN systems for speech recognition: Overview and new research directions. In *International School on Neural Nets: Adaptive Processing of Temporal Information*. Springer Verlag, 1997. To appear.

[p-Bou97] Hervé Bourlard. State-of-the-art and recent progress in hybrid HMM/ANN speech recognition. In W. Gerstner, A. Germond, M. Hasler, and J.-D. Nicoud, editors, *Proceedings of the International Conference on Artificial Neural Networks (ICANN'97)*, number 1327 in Lecture Notes in Computer Science, pages 875–884. Springer-Verlag, 1997.

[p-DB97] S. Dupont and H. Bourlard. Using multiple time scales in a multi-stream speech recognition system. In *EUROSPEECH'97*, pages 3–6, September 1997.

[p-DBD+97] S. Dupont, H. Bourlard, O. Deroo, V. Fontaine, and J.-M. Boite. Hybrid HMM/ANN systems for training independent tasks: Experiments on 'phonebook' and related improvements. In *IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing*, pages 1767–1770, April 1997.

[p-DBR97] Stéphane Dupont, Hervé Bourlard, and Christophe Ris. Robust speech recognition based on multi-stream features. In *Proc. of the ESCA-NATO Workshop on Robust Speech Recognition for Unknown Communication Channels*, pages 95–98, April 1997. IDIAP-RR 97-01.

[p-DMFB97] Benoît Duc, Gilbert Maître, Stefan Fischer, and Josef Bigün. Person authentication by fusing face and speech information. In *Proceedings of the First International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA'97)*, Lecture Notes in Computer Science, pages 311–318. Springer Verlag, 1997.

[p-FB97] V. Fontaine and H. Bourlard. Speaker-dependent speech recognition based on phone-like unit model – application to voice dialing. In *IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing*, pages 1527–1530, April 1997.

[p-FHC+98] Bimbot F., Hutter H.P., Jaboulet C., Koolwaaij J., Lindberg J., and Pierrot J.B. An overview of the cave project research activities in speaker verification. In *Reconnaissance du locuteur et ses applications commerciales et criminalistiques*, 1998. To appear.

[p-FM97] E. Fiesler and M. Maignan. A connectionist system for two-dimensional representation of multivariate location data. In *Proceedings of the Fifth International Workshop on Artificial Intelligence for High Energy Physics*, Amsterdam, The Netherlands, 1997. AIHENP, Elsevier Science.

[p-GMM98] Dominique Genoud, Miguel Moreira, and Eddy Mayoraz. Text dependent speaker verification using binary classifiers. In *Proceedings of the International Conference on Automatic Speech and Signal Processing*, 1998. IDIAP-RR 97-08.

[p-HRBR97] J. Hennebert, C. Ris, H. Bourlard, and S. Renals. Estimation of global posteriors and forward-backward training of hybrid HMM/ANN systems. In *EUROSPEECH'97*, pages 1951–1954, September 1997.

[p-JJJ+98] Pierrot J.B., Lindberg J., Koolwaaij J., Hutter H.P., Genoud D., Blomberg M., and Bimbot F. A comparison of a priori threshold setting procedures for speaker verification in the cave project. In *ICASSP 98*, 1998. To appear.

[p-JLGW97a] Pierre Jourlin, Juergen Luettin, Dominique Genoud, and Hubert Wassner. Acoustic-labial speaker verification. In *Proceedings of the First International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA'97)*, Lecture Notes in Computer Science, pages 319–326. Springer Verlag, 1997. IDIAP-RR 97-13.

[p-JLGW97b] P Jourlin, J. Luettin, D. Genoud, and H. Wassner. Integrating acoustic and labial information for speaker identification and verification. In *Proceedings of the European Conference on Speech Communication and Technology*, pages 1603–1606, 1997.

[p-Lue97] Juergen Luettin. Towards speaker independent continuous speechreading. In *Proceedings of the European Conference on Speech Communication and Technology*, pages 1991–1994, 1997.

[p-May97] Eddy Mayoraz. On the complexity of recognizing iterated differences of polyhedra. In W. Gerstner, A. Germond, M. Hasler, and J.-D. Nicoud, editors, *Proceedings of the International Conference on Artificial Neural Networks (ICANN'97)*, number 1327 in Lecture Notes in Computer Science, pages 475–480. Springer-Verlag, 1997. IDIAP-RR 97-10.

[p-MM97] Eddy Mayoraz and Miguel Moreira. On the decomposition of polychotomies into dichotomies. In *Proceedings of The Fourteenth International Conference on Machine Learning*. Morgan Kaufmann, 1997. IDIAP-RR 96-08.

[p-MM98] Miguel Moreira and Eddy Mayoraz. Improved pairwise coupling classification with correcting classifiers. In *Proceedings of the European Conference on Machine Learning*, 1998. IDIAP-RR 97-09.

[p-Moe97] Perry Moerland. Mixtures of experts estimate a posteriori probabilities. In W. Gerstner, A. Germond, M. Hasler, and J.-D. Nicoud, editors, *Proceedings of the International Conference on Artificial Neural Networks (ICANN'97)*, number 1327 in Lecture Notes in Computer Science, pages 1253–1258, Berlin, 1997. Springer-Verlag. IDIAP-RR 97-07.

[p-PDHGC96] D. Petrovska-Delacretaz, J. Hennebert, D. Genoud, and G. Chollet. Semi-automatic hmm-based annotation of the polycost database. In *Application of speaker recognition techniques in telephony*. COST250, 1996.

[p-PHMG98] D. Petrovska, J. Hennebert, H. Melin, and D. Genoud. Polycost: a telephone-speech database for speaker recognition. In *Reconnaissance du locuteur et ses applications commerciales et criminalistiques*, 1998. To appear.

[p-SMF+97] I. Saxena, P. Moerland, E. Fiesler, A. R. Pourzand, and N. Collings. An optical thresholding perceptron. In *Proceedings of the Workshop on Optics and Computer Science*, 1997. IDIAP-RR 97-16.

[p-SMFP97] I. Saxena, P. Moerland, E. Fiesler, and A. Pourzand. Handwritten digit recognition with binary optical perceptron. In W. Gerstner, A. Germond, M. Hasler, and J.-D. Nicoud, editors, *Proceedings of the International Conference on Artificial Neural Networks (ICANN'97)*, number 1327 in Lecture Notes in Computer Science, pages 1253–1258, Berlin, 1997. Springer-Verlag. IDIAP-RR 97-15.

[p-TEFK97] M. Tajine, D. Elizondo, E. Fiesler, and J. Korczak. Adapting the 2-class recursive deterministic perceptron neural network to m classes. In *Proceedings of the International Conference on Neural Networks*. IEEE, IEEE, 1997.

7.4 IDIAP Research Reports

[r-Beu97] Jean-Luc Beuchat. Reconnaissance de caractères manuscrits à l'aide de réseaux neuromimétiques. IDIAP-RR 18, IDIAP, 1997.

[r-BY97] Souheil Ben-Yacoub. Fast object detection using mlp and fft. IDIAP-RR 11, IDIAP, 1997.

[r-DL97] S. Dupont and J. Luettin. Using the multi-stream approach for continuous audio-visual speech recognition. IDIAP-RR 14, IDIAP, 1997.

[r-Krs97] Sacha Krstulović. Investigation of a possible process identity between drm and linear filtering. IDIAP-RR 19, IDIAP, 1997.

[r-TF97a] G. Thimm and E. Fiesler. Optimal setting of weights, learning rate, and gain. IDIAP-RR 04, IDIAP, 1997.

[r-TF97b] G. Thimm and E. Fiesler. Pruning of neural networks. IDIAP-RR 03, IDIAP, 1997.

[r-VMM97] Patrick Verlinde, Gilbert Maître, and Eddy Mayoraz. Decision fusion using a multiple linear classifier. IDIAP-RR 6, IDIAP, 1997.

7.5 IDIAP Communications

[c-ACB97] Johan M. Andersen, Gilles Caloz, and Hervé Bourlard. Swisscom "AVIS" project (no. 392) advanced vocal interfaces services. IDIAP-COM 6, IDIAP, 1997.

[c-BCF+97] Hervé Bourlard, Jean-Luc Cochard, Emile Fiesler, Gilbert Maitre, and Eddy Mayoraz. Activity report 1996. IDIAP-Com 1, IDIAP, 1997.

[c-GC97] D. Genoud and G. Caloz. 1997 NIST evaluation: Text independent speaker detection (verification). IDIAP-Com 3, IDIAP, 1997.

[c-LM97] Tomas Lundin and Perry Moerland. Quantization and pruning of multilayer perceptrons: Towards compact neural networks. IDIAP-Com 2, IDIAP, March 1997.

[c-Moe97] Perry Moerland. Some methods for training mixtures of experts. IDIAP-Com 5, IDIAP, November 1997.

[c-Sch97] Michael Schmal. Speaker verification by pairwise coupling. IDIAP-Com 7, IDIAP, October 1997.

[c-Van97] Samuel Vannay. Réalisation d'un majordome vocal. IDIAP-Com 4, EPFL / IDIAP, 1997.

7.6 Other Documents

[o-Lue97] Juergen Luettin. *Visual Speech and Speaker Recognition.* PhD thesis, University of Sheffield, 1997.

[o-Thi97] Georg Thimm. *Optimization of high order perceptrons.* PhD thesis, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, June 1997. Dissertation number 1633.