IDIAP COMMUNICATION

IDIAP

Martigny - Valais - Suisse

Activity Report 1996

IDIAP–Com 97-01

February 97

Dalle Molle Institute for Perceptive 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

Foreword

The Dalle Molle Institute for Perceptive Artificial Intelligence (IDIAP) is a non-profit research institute which was founded in 1991 on the occasion of the 20^{th} anniversary of the Dalle Molle Foundation. This foundation was originally focusing on studies in algorithmic linguistics and became, in 1985, the *Dalle Molle Foundation for the Quality of Life*. IDIAP is one of the three semi-public research centers founded by the Dalle Molle Foundation, after ISSCO in Geneva and IDSIA in Lugano, primarily concentrating on automatic translation and fundamental artificial intelligence.

IDIAP is primarily funded by long-term support from the City of Martigny, the Canton of Valais, the Swiss Confederation and the Swiss Telecom PTT. The "Loterie Romande" is also supporting our research efforts with an annual grant. In addition, IDIAP receives substantial research grants from the Swiss National Science Foundation (SNSF) for national projects and the Swiss Office for Science and Education (SOSE) for European projects. Today there is an average of about 25 scientists in residence at IDIAP including permanent staff, postdoctoral fellows, PhD students, and industrial visitors.

IDIAP research activities are of both theoretical and applied nature. Focusing on a few, well defined, research axes, IDIAP carries out fundamental research and develops prototype systems to validate its models against the reality of applications. In this framework, IDIAP is thus very active in the formation of researchers and engineers and in industrial collaboration, and could eventually play an important role in promoting the economic development of the Canton of Valais.

The main research and development activities of IDIAP are centered on the general issues of perception, cognition and pattern analysis (mainly to respond to the present and future needs related to user-machine interaction). In 1996, these activities were split up along three research groups :

- Machine learning, including artificial neural networks, data analysis and data knowledge extraction.
- Speech processing, including automatic speech recognition and speaker verification.
- Vision, including visual speech/speaker recognition and handwriting recognition.

The present Activity Report gives a short overview of the activities which have been carried out in each of these groups. For a technical description, we refer the reader to the numerous publications, IDIAP Research Reports, and IDIAP Communications listed at the end of this report.

In 1996, IDIAP mainly worked under the supervision of a Scientific and Management Committee consisting of Prof. Giovanni Coray (EPFL), Prof. Jean-Paul Haton (University of Nancy, France), Prof. Christian Pellegrini (Geneva University), and Mr. Jean-Pierre Rausis (Secretary of the Dalle Molle Foundation). The Management Committee delegated the supervision of the daily activities of the institute to an internal committee composed of Drs. Gérard Chollet, Jean-Luc Cochard, Emile Fiesler, Gilbert Maître and Eddy Mayoraz. This situation lasted until November 1996 when I was appointed as the new Director of IDIAP.

In many respects, 1996 has been an important year for IDIAP since the appointment of its new director also coincided with its new legal status. Indeed, in November 1996, IDIAP acquired the status of Research Foundation (IDIAP Foundation), now independent of the Dalle Molle Foundation, but in which the initial members (the City of Martigny, the Canton of Valais and the Swiss Confederation) were joined by the Swiss Telecom PTT, the Swiss Federal Institute of Technology in Lausanne (EPFL) and the University of Geneva. In 1996, IDIAP also reinforced its synergy with the industry, the best example being the active collaboration with the Swiss Telecom PTT to develop advanced interactive voice servers.

The scientific structure of IDIAP has slightly evolved since November 1996, looking for a better balance among the various groups, focusing the activities on speech processing while reinforcing the activities in neural networks and vision. In this framework, the scientific axes of the Neural Networks and Vision groups were slightly redefined to broaden their activities and improve their industrial relevance. The Neural Networks group, renamed Machine Learning group, works now on several approaches (including neural networks) towards data knowledge extraction and representation. One of

the goals of this shift is to bias the research towards the problems and the possible approaches to address them instead of being driven by one single research paradigm. In this way, the research carried out in this group should ultimately serve as a test-bed for new technologies on various application domains (such as, e.g., time series prediction and medical diagnosis assistance), while also identifying new promising research directions. The Vision group will expand its activity in automatic character and handwriting recognition, also exploring techniques developed in the Speech Processing group. Research on speech/speaker recognition should broaden to other applications of visual motion analysis in human computer interaction. Finally, it is believed that this new organization should foster communication between the three groups. With its new status and new organization, it is expected that the participation of IDIAP in the national and international research networks, as well as its synergy with the industry, will be reinforced.

Dr. Hervé Bourlard
Director

Préface

L'Institut Dalle Molle d'Intelligence Artificielle Perceptive (IDIAP) est un institut de recherche à but non lucratif qui a été fondé en 1991 à l'occasion du vingtième anniversaire de la Fondation Dalle Molle. Cette Fondation qui était, à l'origine, centrée sur les études en linguistique algorithmique, est devenue en 1985 la *Fondation Dalle Molle pour la qualité de la vie*. L'IDIAP est l'un des trois instituts fondés par la Fondation Dalle Molle, après l'ISSCO à Genève et l'IDSIA à Lugano dont les thématiques de recherche sont respectivement la traduction assistée par ordinateur et l'intelligence artificielle fondamentale.

L'IDIAP bénéficie d'un soutien financier stable de la Ville de Martigny, du Canton du Valais et de la Confédération suisse ainsi que des Télécom PTT. La Loterie Romande verse actuellement une subvention qui vient compléter le financement de base de l'institut. Par ailleurs, l'institut recoit de nombreux subsides de recherche de la part du Fonds national suisse pour la recherche scientifique, pour des projets nationaux, et de la part de l'Office fédéral de l'éducation et de la science (OFES), pour notre participation à des projets européens. Actuellement, le personnel de l'IDIAP est constitué d'environ 25 scientifiques parmi lesquels figurent des chercheurs permanents, des doctorants et des chercheurs invités provenant des milieux industriels.

Les activités scientifiques de l'IDIAP sont à la fois de nature théorique et appliquée. En se concentrant sur quelques axes de recherche bien définis, l'IDIAP effectue des recherches fondamentales et développe des systèmes prototypes qui servent à valider les modèles théoriques sur des applications concrètes. Dans ce contexte, l'IDIAP est très active dans la formation de chercheurs et d'ingénieurs ainsi que dans des collaborations industrielles, et devrait donc jouer un rôle important dans la promotion du développement économique du Canton du Valais.

Les activités majeures de l'IDIAP en recherche et développement sont centrés sur les thématiques générales de la perception, de la cognition et de la reconnaissance de formes (principalement pour répondre aux besoins actuels et futurs en matière d'interaction homme-machine). En 1996, ces activités se partageaient en trois groupes de recherche :

- l'apprentissage automatique, incluant les réseaux de neurones artificiels, l'analyse de données et l'extraction de connaissances de bases de données;
- le traitement de la parole, incluant la reconnaissance automatique de la parole et l'identification du locuteur;
- la vision, avec la reconnaissance visuelle de la parole et de locuteur et la reconnaissance de caractères manuscrits.

Le présent rapport donne un aperçu des différentes activités effectuées par chacun de ces groupes. Pour les descriptions plus techniques, nous renvoyons le lecteur à la liste des publications, rapports de recherche (IDIAP–RR) et communications (IDIAP–Com) qui figurent à la fin de ce document.

En 1996, l'IDIAP a principalement fonctionné sous la responsabilité d'un comité scientifique et de direction formé des Professeurs Giovanni Coray (EPFL), Jean-Paul Haton (Université de Nancy, France), Christian Pellegrini (Université de Genève) ainsi que de Monsieur Jean-Pierre Rausis (Secrétaire de la Fondation Dalle Molle). Ce comité de direction a délégué la gestion des activités quotidiennnes de l'institut à un comité interne formé des Docteurs Gérard Chollet, Jean-Luc Cochard, Emile Fiesler, Gilbert Maître et Eddy Mayoraz. Cette situation a duré jusqu'au début du mois de novembre, date à laquelle j'ai été nommé directeur de l'IDIAP.

L'année 1996 a été une année importante pour l'IDIAP à plus d'un titre, puisque la nomination du nouveau directeur a coïncidé avec la mise sur pied d'un nouveau statut légal. En effet, la *Fondation IDIAP* a vu le jour en novembre 1996, en tant qu'entité indépendante de la Fondation Dalle Molle, constituée des mêmes partenaires, à savoir la Ville de Martigny, le Canton du Valais et la Confédération suisse, auxquels se sont joints les Télécom PTT, l'École polytechnique fédérale de Lausanne et l'Université de Genève. En 1996, l'IDIAP a aussi consolidé ses contacts industriels no-

tamment par sa collaboration active avec les Télécom PTT pour le développement de serveurs vocaux interactifs.

La structure scientifique de l'IDIAP a aussi légèrement évolué depuis le mois de novembre, afin de tendre vers un meilleur équilibre des différents groupes en concentrant les activités du groupe Traitement de la Parole et en renforçant celles en réseaux de neurones artificiels et celles en vision. Ainsi, les axes scientifiques des groupes Réseaux de Neurones et Vision ont été légèrement redéfinis afin d'élargir leurs activités et d'améliorer leur pertinence industrielle. Le groupe Réseaux de Neurones, renommé groupe Apprentissage Automatique, travaille actuellement sur plusieurs approches (incluant les réseaux de neurones) de traitement de données en vue de l'extraction et de la représentation de connaissances. L'un des buts poursuivis par cette évolution est de définir les projets de recherche en fonction de problèmes pratiques et des approches possibles pour les résoudre, et non pas en fonction de problèmes théoriques définis *a priori*. Les recherches effectuées par ce groupe devraient ainsi être un banc d'essai pour de nouvelles technologies dans différents domaines d'application (tels que la prédiction de séries temporelles et l'assistance au diagnostic médical), tout en identifiant de nouvelles directions de recherche prometteuses. Le groupe Vision devrait étendre ses activités en reconnaissance automatique de caractères imprimés ou manuscrits, en explorant également des techniques développées dans le groupe Traitement de la Parole. À travers son nouveau status et sa nouvelle organistaion interne, nous espérons que la participation de l'IDIAP au tissu de recherche national et international, ainsi que ses synergies industrielles, s'en trouveront renforcées.

Dr Hervé Bourlard
Directeur

Vorwort

Das Dalle Molle Institut für Perzeptive Künstliche Intelligenz (IDIAP) ist ein gemeinnütziges Forschungsinstitut, das im Jahre 1991 anlässlich des 20-jährigen Jubiläums der Dalle Molle Stiftung gegründet wurde. Die Dalle Molle Stiftung, die sich ursprünglich mit algorithmischer Linguistik beschäftigte und 1985 in *Fondation Dalle Molle für Lebensqualität* umbenannt wurde, hat neben IDIAP zwei weitere halb-öffentliche Institute gegründet: ISSCO in Genf und IDSIA in Lugano, deren Forschungsschwerpunkte die automatische Übersetzung bzw. fundamentale künstliche Intelligenz sind.

IDIAP wird hauptsächlich finanziert durch langfristige Unterstützung der Stadt Martigny, dem Kanton Wallis, der Schweizerischen Eidgenossenschaft sowie der Schweizer Telekom. Unsere Forschungsaktivität wird ausserdem unterstützt durch Einnahmen der "Loterie Romande". IDIAP erhält zusätzlich wesentliche finanzielle Unterstützung für nationale Projekte fom Schweizerischen Nationalfond und für europäische Projekte vom Bundesamt für Bildung und Wissenschaft. Zur Zeit sinc ca. 25 Wissenschaftler bei IDIAP tätig die sich aus fest angestellten Wissenschaftlern, Forschungs-Assistenten, Doktoranden und Gastwissenschaftlern aus der Industrie zusammensetzten.

IDIAP ist in theoretischer und angewandter Forschung tätig die auf wenige, genau definierte Forschungsbereiche konzentriert ist. In diesen Bereichen wird fundamentale Forschung betrieben die zu prototypichen Systemen führt, welche die theoretischen Modelle in ihrer konkreten Anwendung für gültig erklären. IDIAP ist daher sehr aktiv in der Ausbildung von Wissenschaftlern und Ingenieuren und in der Zusammenarbeit mit der Industrie, was eine Wichtige Rolle spielen wird in der wirtschaftlichen Entwicklung des Kantons Wallis.

Die meisten Forschungs- und Entwicklungsprojekte bei IDIAP betreffen Wahrnehmung, Kognition und Mustererkennung (besonders um den heutigen und zukünftigen Bedürfnissen im Bereich der Interaktion zwischen Mensch und Computer nachzukommen).

Im Jahre 1996 wurden the Forschungsaktivitäten auf drei Forschungsgruppen verteilt :

- Machine learning, einschliesslich künstlicher neuronaler Netze, Datenanalyse, und Wissensextraction von Daten.
- Neuronale Netze, Optimierung, Implementierung neuronaler Netze mit optischen Bauelementen und Anwendungen für Mustererkennung.
- Automatische Verarbeitung der gesprochenen Sprache : Spracherkennung und Sprechererkennung
- Vision : Visuelle Spracherkennung, Sprechererkennung und Handschrifterkennung.

Dieser Forschungsbericht gibt eine kurze Übersicht über die verschiedenen Tätigkeiten, die von den einzelnen Gruppen durchgeführt worden sind. Für technische Einzelheiten ist der Leser auf die Liste der Veröffentlichungen, Forschungsberichte (IDIAP–RR) und Kurzberichte (IDIAP–Com) hingewiesen.

Die wissenschaftliche und leitende Verantwortung des Instituts wurde 1996 von einem Aufsichtsrat ausgeübt, bestehend auf Prof. Giovanni Coray (EPFL), Prof. Jean-Paul Haton (Universität Nancy, Frankreich), Prof. Christian Pellegrini (Universität Genf) sowie von Herrn Jean-Pierre Rausis (Sekretär der Dalle Molle Stiftung). Die Leitung vor Ort wurde von einem internen Komitee, bestehend aus Dr. Gérard Chollet, Dr. Jean-Luc Cochard, Dr. Emile Fiesler, Dr. Gilbert Maître und Dr. Eddy Mayoraz, übernommen. Diese Struktur wurde bis Anfang November 1996 beibehalten, als ich zum Direktor von IDIAP ernannt wurde.

1996 war für IDIAP ein bedeutendes Jahr, da gleichzeitig mit der Ernennung des neuen Direktors IDIAP eine neue Rechtsstellung erhalten hat. Die "Stiftung IDIAP" ist im November 1996 als selbständige Einheit der Stiftung Dalle Molle entstanden, sie besteht noch aus denselben Partnern: der Stadt Martigny, dem Kanton Wallis und der Schweizerische Eidgenossenschaft, denen sich die Schweizer Telekom, die Eidgenössiche Technische Hochschule Lausanne (EPFL) und die Universität Genf angeschlossen haben. Daneben hat IDIAP im Jahre 1996 ihre Synergie im industriellen Sektor

verstärkt, insbesondere durch ihre aktive Mitarbeit mit den Schweizer Telekom bei der Entwicklung von interaktiven sprachgesteuerten Auskunftssystemen.

Um ein besseres Gleichgewicht der verschiedenen Forschungsgruppen herzustellen, hat sich die wissenschaftliche Struktur von IDIAP seit Anfang November ebenfalls verändert. Die Tätigkeiten in der Sprachgruppe wurden auf bestimmte Bereiche konzentriert und die Tätigkeiten in den beiden anderen Gruppen wurden verstärkt. So wurden die wissenschaftlichen Tätigkeitsgebiete der Neuronalen Netzwerk Gruppe und der Vision Gruppe leicht verändert, um ihre Aktivitäten zu erweitern und um ihre industrielle Relevanz zu verbessern. Die Neuronale Netzwerk Gruppe wurde in die Machine Learning Gruppe umbenannt und befasst sich zur Zeit mit mehreren Möglichkeiten (einschliesslich neuronaler Netze) zur Verarbeitung von Daten für 'Knowledge Extraction and Representation'. Ein Ziel dieser Vorgehensweise ist es, die Forschungsprojekte in Bezug auf praktische Probleme und auf die möglichen Lösungen zu definieren und nicht etwa in Bezug auf a priori festgelegte theoretische Probleme. Dadurch kann die von dieser Gruppe durchgeführte Forschungen als Testsystem eingesetzt werden für neue Technologien in verschiedenen Anwendungsbereichen (z. B. die Vorhersage von zeitlichen Folgen oder Unterstützung bei ärzlichen Diagnosen). Die Vision Gruppe wird ihre Tätigkeit in der automatischen Erkennung von gedruckten oder handgeschriebener Schrift vertiefen, indem unter anderem von der Sprachverarbeitungsgruppe entwickelte Techniken angewendet werden. Die Forschung in der visuellen Erkennung von Sprache und Sprecher wird erweitert auf andere Anwendungen visueller Bewegungsanalyse, insbesondere der Interaction von Mensch und Computer. Durch seinen neuen Status und seine neue interne Organisation hoffen wir, dass sich die Tätigeit von IDIAP in nationalen und internationalen Forschungsbereichen sowie die Kontakte im industriellen Bereich verstärken werden.

Dr. Hervé Bourlard
Director

Contents

1 Staff

The staff members currently at IDIAP share the following addresses :

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

In the following subsections, we list the persons who worked at IDIAP in 1996.

1.1 Scientific staff

Mr.	Olivier Bornet Olivier.Bornet@idiap.ch	engineer +41 – 27 – 721 77 40	
Dr.	Hervé Bourlard Herve.Bourlard@idiap.ch	director +41 – 27 – 721 77 20	from 01.11.96
Mr.	Gilles Caloz Gilles.Caloz@idiap.ch	engineer +41 – 27 – 721 77 23	
Dr.	Gérard Chollet Gerard.Chollet@idiap.ch	consultant +41 – 27 – 721 77 25	
Dr.	Jean-Luc Cochard Jean-Luc.Cochard@idiap.ch	researcher, speech processing group leader +41 – 27 – 721 77 28	
Mr.	Andrei Constantinescu Andrei.Constantinescu@idiap.ch	researcher, Ph.D. student +41 – 27 – 721 77 37	
Dr.	Emile Fiesler Emile.Fiesler@idiap.ch	researcher, machine learning group leader +41 – 27 – 721 77 35	
Mr.	Philippe Fu Philippe.Fu@idiap.ch	researcher, Ph.D. student +41 – 27 – 721 77 31	from 01.12.96
Mr.	Arnaud Gaudinat Arnaud.Gaudinat@idiap.ch	researcher +41 – 27 – 721 77 11	from 15.02.96
Mr.	Dominique Genoud Dominique.Genoud@idiap.ch	researcher, Ph.D. student +41 – 27 – 721 77 26	
Mr.	Hervé Glotin Herve.Glotin@idiap.ch	researcher, Ph.D. student +41 – 27 – 721 77 33	from 01.07.96
Mr.	Cédric Jaboulet Cedric.Jaboulet@idiap.ch	engineer +41 – 27 – 721 77 36	
Mr.	Pierre Jourlin from 15.07.96 to 30.09.96	visiting researcher	

Mr.	Sacha Krstulović Sacha.Krstulovic@idiap.ch	researcher +41 – 27 – 721 77 43	from 01.10.96
Mr.	Jürgen Lüttin Juergen.Luettin@idiap.ch	researcher, Ph.D. student +41 – 27 – 721 77 27	from 01.02.96
Dr.	Gilbert Maître Gilbert.Maitre@idiap.ch	researcher, vision group leader +41 – 27 – 721 77 30	
Dr.	Eddy Mayoraz Eddy.Mayoraz@idiap.ch	researcher +41 – 27 – 721 77 29	
Mr.	Guillaume Melin Guillaume.Melin@idiap.ch	engineer +41 – 27 – 721 77 38	from 01.02.96
Mr.	Perry Moerland Perry.Moerland@idiap.ch	researcher, Ph.D. student +41 – 27 – 721 77 32	
Mr.	Miguel Moreira Miguel.Moreira@idiap.ch	researcher, Ph.D. student +41 – 27 – 721 77 45	until 15.01.96, from 1.11.96
Dr.	Indu Saxena Indu.Saxena@idiap.ch	researcher +41 – 27 – 721 77 22	
Mr.	Georg Thimm Georg.Thimm@idiap.ch	researcher, Ph.D. student +41 – 27 – 721 77 39	
Mrs.	Murielle Vial	researcher	until 30.09.96
Mr.	Hubert Wassner	researcher	from 01.01.96, until 15.08.96

1.2 Non-scientific staff

Mrs.	Catherine Delattre-Locher	secretary	until 30.09.96
Mrs.	Sandrine Dély-Schindler Sandrine.Dely@idiap.ch	annotator +41 – 27 – 721 77 37	
Mrs.	Sylvie Millius Sylvie.Millius@idiap.ch	secretary +41 – 27 – 721 77 21	from 01.10.96
Mr.	Serge Meunier	graphic designer	from 01.02.96, until 31.07.96

1.3 Scholars

Several students have accomplished their diploma work or a traineeship at IDIAP. Their names are listed in Section 3.4.

2 Research Activities

2.1 Machine Learning Group

By the end of 1996, it was decided to rename the Neural Networks Group into Machine Learning Group to reflect the broadening of its activities. Indeed, it was observed that the group had a natural tendency to extend its research axes towards different approaches related to data knowledge extraction and representation. Towards this goal, artificial neural networks are then just one possible tool among others that will be investigated here. This has the advantage of biasing the research towards the problems and the alternative approaches to address them instead of being driven by one single theoretical paradigm. This should also allow the group to serve as a test-bed for new technologies on various application domains (such as, e.g., time series prediction and medical diagnosis assistance).

2.1.1 Research Directions

In spite of the paradigm shift briefly discussed above (and decided only by the end of 1996), the activities of the former Neural Networks group at IDIAP were mainly a logical continuation of those of 1995, with some extensions towards data mining or data knowledge extraction techniques.

Even though neural networks have proven their usefulness in a broad range of application domains, their wide-scale acceptance and use is hampered by their user-unfriendliness. That is, a large amount of expertise and training overhead is required for the selection of their topology and their multitude of learning rule parameters. Another crucial problem is that the intrinsic power of neural networks, which lies in their massive parallelism and distributed nature, is not exploited since they are usually only simulated on (sequential) computers. This exploitation by parallel and distributed hardware implementations, whether in electronics or optics, is restricted by the lack of existing neural network adaptations to hardware requirements. The main advantage of optical systems is that massively parallel computation can be performed, and that the number of processing elements can be scaled without compromising speed appreciably. In realizing the great potential of optical computing, and to exploit the full capabilities of optics by taking advantage of the ultimate speed, the speed of light, optical neural networks are to be developed. Optical neural networks provide a promising alternative to electronic systems in the parallel handling of data, offering up to the order of 10 billion interconnections per optical component. Having no crosstalk problems and a very low power consumption are additional benefits of optics as compared to electronics; the capabilities of the latter technology being almost saturated. The promise of optics can only be realized if fully optical modules, building blocks for all-optical neural networks, are developed. The newly developed systems are to be applied to a variety of domains, in particular to pattern recognition.

Besides the above research on artificial neural networks, we started exploring other aspects of learning and started studying alternative approaches towards data mining and pattern recognition. Some of the applications targeted so far include pattern classification (as support to the speech and vision groups), time series prediction and diagnosis assistance. In 1996, the main non-neural were essentially related to data mining and more specifically to a new methodology named *logical analysis of data* for data knowledge discovery.

The research activities of the Machine Learning group in 1996 was thus concerned with the following themes :

- Compact User-Friendly Neural Networks
- Optical Neural Network Design
- Neural Network Adaptation for Hardware Implementation
- Applications
- Logical Analysis of Data

Some of our most important research and development activities related to these five themes are briefly discussed below. For more technical details, we refer the reader to our publications listed in Section 6.

Compact User-Friendly Neural Networks

Our major research goal is the development of a novel family of powerful multi-functional neural networks that are both compact and user-friendly. This ambitious research has a strong foundation as it is based on the integration of the potential and latest results of IDIAP's research activities in this area.

Compactness will be obtained by complexity reduction strategies based on sparse connectivity and ontogenic training methods that automatically modify the network during the training process. Parameter reduction is established by careful analysis of the networks' functioning and by exploiting the key aspects of neural networks, namely their capability to learn and self-organize.

These easy to use and yet multi-purpose neural networks, with their potential to outperform present-day multilayer networks, will also be extremely compact and therefore very suitable for a wide range of hardware implementations.

For testing our neural networks we are using the large corpus of data available at IDIAP (as briefly discussed in the section on Applications below).

Our recent results include:

- Work on neural network initialization and its extension by an in-depth and higher dimensional exploration of its relationship with other network parameters [a-TF97].

- Possibility of eliminating a key parameter of the network's nonlinearity by proving its relationship to learning rate and initial conditions, which simplifies the neural network design process [a-TMF96].

Optical Neural Network Design

In 1996, we have been continuing the development of our unique multilayer optical neural network (ONN), having all-optical recall, which is modular and extendible by adding additional neuron layers [a-SF95]. This work is performed in collaboration with the Institute of Microtechnology of the University of Neuchâtel. The challenge of multilayer ONNs is to incorporate not only recall, but 'training' as well. In the optical implementation of neural networks typically performed by incoherent techniques, the representation of negative values is precluded. This necessitated a solution which allows optical methods to implement neural networks efficiently. More precisely, we found that subtraction stage is essential for efficient optical implementations. A novel technique for performing the subtraction has been developed and integrated into our multilayer ONN.

In 1996, we completed the first module of the implementation of our large modular optical multilayer neural network with all-optical recall and optically implemented a Perceptron based on 70,000 information carrying light beams [p-SMF+97].

Neural Network Adaptation for Hardware Implementation

It is foreseen that progress in optics, as well as in analog and digital (VLSI) electronics, will increase demands on the development of dedicated neural networks for hardware implementations. Our goal is therefore to develop adaptations of neural networks for hardware implementations by working in close cooperation with groups active in the hardware developmental aspects of it. In this context we are continuing to closely follow the progress in Optical Neural Network research (described above) by extending our techniques to efficiently train modular optical multilayer neural networks with all-optical forward propagation.

The most important goal from the neural network side is to concentrate our activities on both weight discretization and transfer curve adaptation techniques by improving them in applicability, flexibility, and speed. These new weight discretization and network optimization results will also be

applied to complexity reduction and integrated into the Compact User-Friendly Neural Networks research.

For reference, an informative survey of neural network adaptations for hardware implementation can be found in [p-MF96].

Applications

The neural network models that we have developed and improved by our research efforts have been successfully applied to various problems including handwritten character recognition and pattern recognition in general, classification problems, and quality control.

In 1996 we successfully finished the first phase of a collaborative project with GPIL in Martigny (see Section 4.3) on the quality control of wrist watches, from a well-known Swiss watch company, by neural networks. This project is based on neural pattern recognition of the hands of the wrist watches. The results show an efficient and fast solution with a maximal performance, requiring a minimal training set.

In 1996 we also initiated two new collaborations with the University of Lausanne focusing on specific applications:

- Applying neural network techniques to cartographical mapping problems (Professor Maignan): see "Prospective Research Grants" below.
- Automatic drug analysis based on visual, physical and chemical features (Forensic Laboratory, Mr. Ribaux). This involves problems such as classification into drug classes and identification of links between drug classes (e.g., same powder, same manufacturer, same batch, etc.).

Logical Analysis of Data (LAD)

LAD is a new methodology for extracting knowledge from databases by using only logical features of the data. It has been proposed and originally developed in RUTCOR, Rutgers University, New Jersey. Like other tools such as regression trees, this approach presents the advantage of being very easily understandable and interpretable by human experts. It has been proved to be very efficient as a learning and classification technique and to compare favorably to the best existing methods either from statistics, or from neural networks, or from artificial intelligence [r-BHI+96]. However, in its original form it can only handle two-class problems. Therefore, our particular interest in the research carried on at IDIAP is to generalize LAD to multi-class learning problems.

In a preliminary study, we focused on general methods of decomposition of a multi-class classification problem into a series of problems composed of two classes (dichotomies), independently of the learning method used to solve each dichotomy. In this framework, we developed a novel approach [r-MM96] enhancing a method recently proposed and which is based on error-correcting codes. The expertise acquired during this first phase will allow us to design performant decomposition methods specific the LAD.

2.1.2 Research Grants

♢ Adaptive Optical Multilayer Neural Networks

Funding Swiss National Science Foundation
Duration October 1995 – September 1997
Partners Institute of Microtechnology (IMT) of the University of Neuchâtel
Principal investigator Dr. Emile Fiesler
Other internal 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 involves 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 is established.

♢ Compact Hardware-Friendly Neural Networks

Funding Swiss National Science Foundation, FN 21-45621.95

Duration April 1996 – March 1999

Principal investigator Dr. E. Fiesler

Other IDIAP staff P. Moerland and T. Lundin

Description To fully profit from the inherent parallelism of neural networks and to enable real-time processing, hardware implementations are essential. However, the mapping of existing neural network algorithms onto fast, compact, and reliable hardware is a difficult task and any kind of implementation of neural networks is subject to various constraints on the three basic elements of a neural network:

Topology of the network One of the main drawbacks of neural network learning algorithms is the difficulty of determining their topology. This leads to over-sized networks that have much redundancy and are not suitable for hardware implementation.

Interconnection weights The accuracy that is needed for representing the weights of a neural network is too area consuming in digital implementations and is incompatible with the system noise in analog implementations. Hence, the number of different weight values of the network should be as small as possible in order to permit efficient implementations.

Neurons In analog implementations of neural networks the functional units, such as the multipliers and non-linear activation functions, are not ideal and suffer from non-uniformities between components and other inaccuracies.

It is the goal of this project to develop a unified solution to the optimization of these constraints. In order to obtain neural networks that are compact in size and robust against inaccuracy, a combination of ontogenic algorithms and an efficient discretization rule will be developed. The resulting compact solution will be augmented to handle inaccurate neurons with non-standard activation functions and spatial non-uniformities.

♢ Pattern Recognition by Neural Networks for Quality Control of Watches

Funding Industrial project

Duration September 1995 – January 1996 (phase I)

Partners GPIL (Gianni Pante Ingenierie du Logiciel)

Principal investigator Dr. Emile Fiesler

Other IDIAP staff Mr. Miguel Moreira

Description This project concerns the quality control of Swiss watches by neural networks. The challenge of this project lies in minimizing the number of watch images needed to train the system, in order to optimize the robotized control process, while retaining maximum quality.

♢ GLAD – Generalization of LAD

Funding Swiss National Science Foundation, FN 21-46974.96
Duration November 1996 – October 1998
Partners Swiss Federal Insitute of Technology (EPFL)
Principal investigator Dr. Emile Mayoraz
Internal staff Mr. Miguel Moreira
Description This project is about the generalization of Logical Analysis of Data into a method capable of handling classification problems with a large number of classes.

2.1.3 Prospective Research Grants

♢ NEVROTOPOS

Funding Swiss National Science Foundation
Duration March 1997 – August 1998 (proposed)
Partners University of Lausanne, University of Geneva, and the "Commission Internationale Protection des Eaux du Léman"
Principal investigator Prof. M. Maignan (University of Lausanne)
IDIAP responsible Dr. Emile Fiesler
Description This project concerns the quality control of the water in Lake Geneva using neural networks to create maps from multivariate sparse data.

2.2 Speech Processing Group

2.2.1 Research Directions

The Speech Processing group is working on all aspects of automatic speech recognition and speaker verification, including fundamental research and development of (non real-time and real-time) prototype systems. Its research activities can be described along three main axes:

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

Some of our most recent research and development activities related to these three themes are briefly discussed below. For more technical details, we refer the reader to our publications listed in Section 6.

Automatic Speech Recognition (ASR)

In 1996, work carried on in the framework of ASR mainly focused on (1) improving speech unit models, (2) multi-agent systems, and (3) prototype development and technology transfer. More specifically :

1. Speech unit models — Development of better phonetic models based on hidden Markov models and/or artificial neural networks 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.

2. Multi-agent systems — Besides being a fundamental research issue in computer science, it is investigated here how the concept of multi-agent systems could be used to include multiple (statistical and non-statistical) heterogeneous knowledge sources in state-of-the-art speech recognition systems.

3. Technology transfer — In collaboration with the Swiss Telecom PTT, IDIAP defines and develops application prototypes involving speech recognition (and speaker verification) technology. For example, *interactive voice server* (IVS) prototypes are being developed to assess the technology as well as the possible applications.

Most of these developments were performed in the framework of research grants briefly described in Section 2.2.2.

Speaker Recognition (SR) and Speaker Verification (SV)

In 1996, the SR/SV effort was 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 project and two European projects (CAVE and COST 250) described below. Most of the research is performed in the context of an existing in-house large database of speech samples designed especially to address the problem of inter- and intra-speaker variability.

Speech Databases

One major auxiliary problem related to ASR and SR/SV research is the availability of large speech databases. Consequently, since 1992, IDIAP has been recording large speech databases over telephone lines. To manage and use these large databases, a set of software packages has been developed.

As this recording process is still a currently running activity in collaboration with other research centers in Switzerland or abroad, adaptation and dissemination of our software packages is performed.

2.2.2 Research Grants

♢ CAVE – Speaker Verification in Banking and Telecommunication

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

Duration Dec. 1, 1995 – May 31, 1997

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

Principal investigator Dr. Gérard Chollet

Internal staff Dr. Jean-Luc Cochard (task manager), 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 a 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 Mar. 1, 1996 – Feb. 28, 1998
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), Swiss Telecom PTT (CH), Telenor R&D (N), Univ. of Maribor (SL), Univ. München (D), Univ. of Patras (GR), Univ. Politecnica de Catalunya (E), Vocalis Ltd (UK)
Principal investigator Dr. Gérard Chollet
Internal staff Dan Andrei Constantinescu, Gilles Caloz
Description The SpeechDat project is a CEC-funded initiative that addresses the fields of production, standardization, evaluation and dissemination of Spoken Language Resource (SLR).

The goal of the present project is to provide guidelines and recommendations issued from the work achieved in SpeechDat (focusing on telephone like applications) as well as from other relevant current or past initiatives (covering other applications).

◊ Automatic Speaker Recognition over the Telephone Network

Funding European project, COST 250 action, supported by OFES
Duration Oct. 1, 1995 – Sep. 30, 1998
Partners France, Italy, United Kingdom, Sweden, The Netherlands, Spain, Portugal, Ireland, Denmark, Greece, Slovenia, Switzerland, Turkey
Principal investigator Gérard Chollet
Internal staff Dominique Genoud, Dr. Gérard Chollet (coordinator)
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 record, process, annotation and distribution of the Database was made by IDIAP, EPFL and KTH. This database was distributed for free to each partner of the project.

◊ VTC: Voice Telecom Card: Speech and speaker recognition application

Funding Swiss Telecom
Duration Jul. 1, 1996 – October. 30, 1996
Partners Swiss Telecom, aComm, Link
Principal investigator Dr. Jean-Luc Cochard
Internal staff Olivier Bornet, Guillaume Melin, Gilles Caloz, Dominique Genoud
Description VTC is a speech/speaker recognition application built for the Swiss Telecom PTT. The goal of VTC was to perform a field-test in calling-card users to assess the acceptance rate for new speech-based applications. This project was conducted by Link, an independent marketing institute.

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

Funding Swiss Telecom
Duration July. 1st, 1996 – March. 30th 1997
Principal investigator Dr. Jean-Luc Cochard
Internal staff Dominique Genoud.
Partners Swiss Telecom, Frédéric Bimbot (ENST)
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.

♢ $\mathsf{ETC}_{\mathsf{reco}}$: A system of automatic speech recognition over the telephone

Funding European project, COST 249 action, supported by OFES
Duration Oct. 1, 1995 – Sep. 30, 1998
Partners Belgium, Switzerland, Czech Republic, Germany, Denmark, Spain, France, Greece, Hungary, Italy, Lithuania, Norway, The Netherlands, Portugal, Poland, Sweden, Slovakia, Slovania, United Kingdom
Principal investigator Dr. Jean-Luc Cochard
Internal staff Dan Andrei Constantinescu, Hervé Glotin
Description $\mathsf{ETC}_{\mathsf{reco}}$ has to be understood as an extension of $\mathsf{ETC}_{\mathsf{vérif}}$ in many directions (see below). First, its dedicated objective is to implement an effective continuous spoken language recognition system. Second, the set of knowledge source is larger than the above ones. Some efforts will be directed towards the integration of linguistic knowledge, namely statistical and rule-based ones. A third improvement will concern a better decomposition between the kernel of the system ETC and its surrounding modules. We hope this work to give rise to a real application independent multi-agent approach.

♢ Speech and speaker recognition over the telephone on SwissNet

Funding Swiss project, CTI (formerly CERS)
Duration Sep. 1, 1995 – Aug. 31, 1997
Partners Sun Microsystems (Switzerland), aComm, Telecom PTT, EPFL, IDIAP
Principal investigator Dr. Andrzej Drygajlo (EPFL)
Internal staff Olivier Bornet, Dr. Jean-Luc Cochard (co-ordinator)
Description The goal of this project is to make speech/speaker recognition systems available on multi-processor workstations and SwissNet platform to industrial partners, and particularly to Swiss industry for Swiss French. The research addresses among others the problem of phonetic models for a speaker independent, vocabulary independent speech recognition system for telephone applications. The results are likely to be exploited in a new Interactive Voice Server application to be configured by the Swiss Telecom PTT.

◊ $\mathsf{ETC}_{\mathsf{vérif}}$: a system to help labelling of speech corpora

Funding Swiss National Science Foundation, FN 20-43494.95, follow-up of FN 21-37 467.93

Duration Nov. 1, 1995 – Oct. 31, 1997

Principal investigator Dr. Jean-Luc Cochard

Internal staff Philippe Fu, Arnaud Gaudinat, Hervé Glotin, Murielle Vial

Description $\mathsf{ETC}_{\mathsf{vérif}}$ is a prototype of a continuous spoken language recognition system (CSLR). This work stems from the strong intuition that a probable solution to the general problem of speech understanding lies in the development of a system able to deal with a large set of distinct, partial and even unreliable problem solvers, namely HMMs (Hidden Markov Models), GTP (Graphemes To Phonemes) agents, prosodic analysers and even higher order agents processing syntactic and semantic knowledge.

The system $\mathsf{ETC}_{\mathsf{vérif}}$ under development 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"). The purpose of decomposing a continuous spoken language recognition system into two layers: a kernel that is application-independent and a periphery that bears all the knowledge of the application domains, will improve the design of a flexible and adjustable system.

$\mathsf{ETC}_{\mathsf{vérif}}$ is addressing a simplified problem of CSLR, namely verification of speech utterances. This concretely means that the input data of the system is twofold: the signal sample, on one side, and the text that had to be uttered, on the other side. This simplified context greatly reduces the required number of agents and the internal complexity of some agents that have to be developed. Nevertheless, many experiments can be conducted in this case that give us some valuable information on how to setup an entire and efficient CSLR system.

◊ Enhanced automatic speaker recognition in telephony

Funding Swiss National Science Foundation, FN 21-45624.95

Duration Apr. 1, 1996 – Mar. 31, 1998

Internal staff Dominique Genoud

Description The research topics under consideration in this project aim at gaining more knowledge in speaker recognition by performing:

- 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 token.
- Evaluation of recognition speaker technology with regards to or in synergy with other biometrical technologies.

♢ ESA-SPACT

Funding European Space Agency (ESA), Noordwijk (NL)
Duration Jan. 1, 1996 – Aug. 31, 1997.
Partners ESTEC (NL), NLR (NL), ORIGIN/BSO (NL), TNO-TM (NL), TCD (IR), IDIAP (CH)
Principal investigator Dr. Jean-Luc Cochard
Internal staff Gilles Caloz, Dan Andrei Constantinescu, Luis Miguel Moreira
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 synthesised 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 Swiss Telecom-PTT
Duration Dec. 1, 1995 – Apr. 30, 1997
Partner TIK/ETH Zürich
Principal investigator Dr. Jean-Luc Cochard
Internal staff Gilles Caloz, Dan Andrei Constantinescu (coordinator)
Description Since IDIAP has already gained valuable experience in the Swiss-French Polyphone project, the huge amount of work for content specification, sheet preparation, speaker recruiting, speech recording, and annotation will be split between TIK and IDIAP. The tasks dedicated to IDIAP are the following ones:

Preparation of the prompting sheets: A lot of the addressed speakers finally don't call the automatic recording machine. In order to get still maximum statistically balanced speech material this has to be taken into account, i.e., the sheets have to be produced continuously. This task is done at IDIAP.

Tools: IDIAP will provide their tools used for text preparation and for speech annotation. In particular, IDIAP will support TIK to adapt the tools to the new requirements.

♢ PolyVar – Recording and annotation of a Swiss-French corpus

Funding internally supported
Duration Feb. 1st, 1994 –Jul. 31st, 1996
Principal investigator Dr. Gérard Chollet
Internal staff Dan Andrei Constantinescu, Sandrine Dely (annotator), Cédric Jaboulet
Description The objective of PolyVar is to create a database of speech samples designed to capture intra-speaker variability. The content and recording conditions of PolyVar are very similar to the ones of Polyphone. The population is however very different. Instead of 5000 distinct persons, PolyVar initial goal was to have 100 recording sessions of 50 persons. This ideal distribution won't be reached even if we already have more than 3000 recording sessions, verified and ready to be put on CD-Rom.

♢ M2VTS (see Section 2.3.2)

2.2.3 Prospective Research Grants

♢ Using articulatory features for speech recognition / speaker verification

Funding Swiss National Science Foundation
Duration April 1997 – March 1999 (proposed)
Principal investigator Dr. Jean-Luc Cochard
Internal staff Sacha Krstulović, Jürgen Lüttin
Description This research project aims at using articulatory features in SR/SV applications. Such features are believed to lead to significant improvements of the 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 a cineradiographic database 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.

♢ 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).
Principal investigator Dr. Jean-Luc Cochard
Description The purpose of this project is to organize an advanced course that allows students to qualify for multidisciplinary teamworking in the language industries. Besides in depth knowledge of Speech Science, Natural Language Processing or Computer Science, that has been provided by undergraduate studies, through this Masters the student will obtain the contextual knowledge from the fields that are not his/her specialisation.

IDIAP and Institut Kurt Boesch applied for participation to this Masters with the objective to create a center of excellence in the domain of Speech Processing for graduated students in Wallis, that is part of a large European teaching network.

♢ THISL – THematic Indexing of Spoken Language

Funding European project, ESPRIT Program, Long Term Research supported by OFES
Duration 3 years (proposed, from 1997 to 2000)
Partners Sheffield University (UK), SoftSound Ltd (UK), Faculté Polytechnique de Mons (B), BBC (UK), Thomson (F), and Intl. Comp. Science Institute (Berkeley, CA).
Principal investigator Dr. Hervé Bourlard
Internal staff Jürgen Lüttin
Description Automatic indexation of high fidelity audio recordings and retrieval based on vocal access of thematic keywords.

♢ SPHEAR – SPeech, HEAring and Recognition

Funding OFES (European DGXII TMR Research Network)
Duration 3 years (proposed, from 1997 to 2000)
Partners Sheffield Univ. (UK), Univ. of Bochum (D), Ericsson (D), ICP Grenoble (F), IDIAP (CH), Univ. of Keele (UK), Univ. of Patras (GR).
Principal investigator Dr. Hervé Bourlard
Internal staff Open
Description The objectives of this project are to achieve better understanding of auditory information processing and to deploy this understanding for automatic speech recognition (ASR) in adverse conditions. The project content has four related themes. The first two of these are concerned with modelling auditory Scene Analysis: the ability of listeners to separate, and pay selective attention to, evidence from individual sound sources in the mixture which reaches the ears. In the third theme we intend to develop ASR techniques which will function in natural auditory scenes, where there are many streams of information, each of which may be incomplete or have contributions from more than one source. In the final theme we will deploy these ideas in commercially important ASR applications.

♢ C-STAR II – Consortium for Speech Translation Advanced Research

Duration 3 years
Partners ATR (JPN), CMU (USA), ETRI (KO), IRST (IT), Univ. of Karlsruhe (D), Siemens (D), and a Francophone sub-group constituted by IDIAP, Univ. of Lausanne, Univ. of Geneva, Univ. of Grenoble (F).
Principal investigator Dr. Jean-Luc Cochard
Internal staff Open
Description The purpose of this project is to develop a set of machine translation systems, using speech input and providing speech output. The translation domain under consideration is the one of touristic information services. The foreseen scenario is the following one: two persons (e.g., a French speaker, and a Korean one) want to communicate. The French speaker could be a touristic agent located in Martigny, and the Korean one, a tourist seeking for information in order to organize holidays in Switzerland, from his home in Korea.

2.3 Vision Group

2.3.1 Research Directions

In the wide spectrum of machine vision, the Vision group's research is focused on problems related to human computer interaction. These are:

- Document Analysis and Recognition: off-line handwriting recognition
- Shape Representation and Recovery: lip tracking, cineradiographic (X-ray) image sequence analysis
- Image Sequence Analysis and Visual Learning: visual speech recognition, audio-visual speaker authentication

Off-line Handwriting Recognition

Although it is a sub-area of machine vision research, we believe that off-line handwriting recognition faces many inherent problems of machine vision, for which there is still a lack of theory and which we express as :

- How to represent shape and how to measure similarity to achieve the best recognition rates?

- What is the optimal data/information reduction for knowledge?
- How to solve the segmentation/recognition dilemma?

Beside the theoretical interest just mentioned, we also consider applications of handwriting recognition. Our major interest is in recognizing ancient manuscripts. Indeed, we think that in the age of telematics, scientists and technicians should provide tools which allow a better access to the huge heritage received from our ancestors in the form of written documents.

Lip-tracking

Lip-tracking has become an important issue in both automatic speech processing and automatic face processing. Potential applications include :

- Audio-visual speech recognition
- Audio-visual person recognition
- Lip-synchronization
- Speech-driven talking heads
- Speech-based image coding

Most previous approaches for lip-tracking have simplified the problem by marking the subjects lips with color or a reflective marker, by locating the lips in the first image by hand, by performing experiments for one subject only, or by using very controlled lighting conditions.

We have developed a model-based approach for localising and tracking lips in grey-level image sequences. The model learns patterns of shape variability and image variability from a training set. This constrains the model during image search to only deform in ways similar to the training examples. The system has been tested on two different databases and been used for both, speech recognition and person authentication applications.

Visual Speech Recognition (Lipreading)

The performance of most state-of-the-art speech recognition systems drops considerably in the presence of noise. This limits their use in real world applications, which are basically all subject to some interference from noise. Attempts to reduce the effect of noise in the speech signal have only shown limited success, particularly when the noise was due to crosstalk (cocktail party effect).

Humans on the other hand use lipreading (speechreading) as supplementary information for speech perception, especially in noisy conditions. The main benefit of visual information stems from its complementarity to the acoustic signal, i.e. phonemes that are difficult to distinguish acoustically are easier to distinguish visually.

We have developed a speech-reading system which extracts visual speech information from the image sequence of a speaking person, which are used for speech recognition. The extracted parameters describe important features like the lip shape and the visibility of teeth and tongue. The features are modelled by Gaussian distributions and their temporal dependencies by Hidden Markov Models. The system achieved an accuracy of 90.6 % for a speaker-independent recognition task of digits, using only visual features. This performance is approximately equivalent to the performance of humans with no lipreading knowledge, who performed the same task.

Audio-Visual Speaker Authentication

Automatic verification of a person's identity is a difficult problem and has received considerable attention over the last decade. The ability of such a system to reject impostors, who claim a false identity, becomes a critical issue in security applications.

We have developed a novel approach for person recognition, based on spatio-temporal modelling of visual features, extracted from the talking face. We train models, specific to a person's speech articulators and the way that person speaks. Person identification is performed by tracking mouth

movements of the talking face and by estimating the likelihood of each model of having generated the observed sequence of features. The model with the highest likelihood is chosen as the recognised person. The system was tested on the Tulips1 database and had an error rate of 2.1%.

We have extended this approach to a multimodal person authentication system which consists of two classifiers, one using visual features and the other using acoustic features. Verification experiments were performed on the M2VTS database for the individual modalities and for the combined classifier. The performance of the integrated system outperformed each sub-system and reduced the false acceptance rate of the acoustic sub-system from 2.3% to 0.5%.

2.3.2 Research Grants

♢ Optical Character Recognition

Funding Swiss National Science Foundation, FN 21-39576.93
Duration April 94 – March 96
Principal investigator Dr. Thomas M. Breuel (now with the IBM Almaden Research Center)
Internal staff Dr. Thomas M. Breuel, Dr. Gilbert Maître
Description The goal of this project was to build a character recognition module that performs significantly better than existing systems in complex OCR tasks, i.e. in the presence of interfering strokes, or in writing styles that are difficult to segment.

We have developed new methods for matching efficiently under bounded error and the minimum Hausdorff distance. We have also evaluated and compared similarity measures that have been proposed both in the OCR and computer vision literature. Some of the empirical results of this work have influenced the design and implementation of a character recognition system for the US Census applications.

♢ M2VTS – Multimodal Verification for Teleservices and Security Applications

Funding European project AC 102, ACTS Program, supported by OFES
Duration Oct. 1st, 1995 – Sep. 30, 1998
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)
Principal investigator Dr. Gérard Chollet
Internal staff Pierre Jourlin, Jürgen Lüttin, Dr. Gilbert Maître, Hubert Wassner
Companion projects CAVE, COST 250 (see Section 2.2.2)
Description The primary goal of the M2VTS project is to address the issue of secured access to local and centralised services in a multi-media environment, developing automatic verification systems which combine multimodal strategies (secured access based on speech, image and other information)

The first year of the project was mainly devoted to the recording of an audio-visual database (37 persons), to the set-up of pilot demonstrators, to the choice of a platform for the final demonstrator, and to the development of basic algorithm components.

IDIAP has set up a demonstrator of audio-based speaker authentication based on algorithms developed in other projects (ATTACKS, COST 250) and developed basic algorithms for lip tracking, for video-based speaker authentication, and for combined audio- and video-based speaker authentication.

The video-based speaker authentication achieved a performance of about 85% to 90% of correct accept/reject decision on the database. The combination of video-based and audio-based speaker authentication achieved up to 99% of correct accept/reject decision.

2.3.3 Prospective Research Grants

♢ ADMIRE – Accessing anD Managing multimodal Information REsources

Submitted December 18th, 1996
Funding European project, ESPRIT Program, Long Term Research supported by OFES
Duration 3 years
Principal investigator Dr. Hervé Bourlard
Partners ENST (F), Aalborg University (DK), Faculté Polytechnique de Mons (B), A2iA (F), SLP InfoWare (F), EPFL (CH), Barclays Bank (UK), France Telecom (F), Swiss Telecom (CH)
Description The general objective of ADMIRE is to develop an intelligent multimodal information resource manager dedicated to every day life exchange of multimodal information and efficient management of large amounts of multimodal data. The underlying goal is to contribute to the emergence of a new shared information space resulting from the aggregation of various well-established communication networks. As discussed in the proposal, we think that this problem presents and effective context to address several important problems like: presentation of information for improved social interaction, management of inhabited (coexisting) information spaces, and further development and testing of related research issues.

If accepted, the main role of IDIAP in this project will be to :

1. Provide speech recognition algorithms (with no additional research) and adapt them to the application.
2. Adapt the existing speech recognition software to the recognition of written documents. There is indeed a growing interest in transfering speech recognition technology to character and handwriting recognition problems.

♢ Using articulatory features for speech recognition / speaker verification (see Section 2.2.3)

♢ ICR for libraries

Description Contacts have been established with the BAMBI project consortium (EC Telematics Program) and with the Swiss Federal Archives in order to start projects on the development of Intelligent Character Recognition systems for libraries.

3 Educational Activities

3.1 Ph.D. Programs

- **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** David Elizondo
  Supervisor Dr. Emile Fiesler
  Research topic Sparse Neural Networks
  University Louis Pasteur University, Strassbourg, France

- **Ph.D. Candidate** Dominique Genoud
  Supervisor Prof. Martin Hasler, Dr. Gérard Chollet
  Research topic Enhanced automatic speaker recognition in telephony
  University EPFL

- **Ph.D. Candidate** Sacha Krstulović
  Supervisor Dr. Hervé Bourlard
  Research topic Using articulatory features for speech recognition / speaker verification

- **Ph.D. Candidate** Jürgen Lüttin
  Supervisor Steve Renals, Neil Thacker (Univ. Sheffield)
  Research topic Visual Speech and Speaker Recognition
  University University of Sheffield, U.K.

- **Ph.D. Candidate** Perry Moerland
  Supervisor Dr. Emile Fiesler
  Research topic Neural Networks for Hardware
  University EPFL, Lausanne

- **Ph.D. Candidate** Miguel Moreira
  Supervisor Dr. Eddy Mayoraz
  Research topic GLAD – Generalization of LAD
  University EPFL, Lausanne

- **Ph.D. Candidate** Georg Thimm
  Supervisor Dr. Emile Fiesler
  Research topic Ontogenic High Order Neural Networks
  University EPFL, Lausanne

3.2 Lectures

- **Title** Serveurs vocaux interactifs et vérification de l'identité du locuteur
  Speaker Dr. Gérard Chollet
  School École Polytechnique Fédérale de Lausanne
  Date January 13, 1996
  Audience Students of electrical engineering department

- **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** Interactive Voice Servers over the Telephone: Development databases and flexible vocabulary approaches to ASR
  Lecturers Gérard Chollet and Dan Andrei Constantinescu
  Location SRI, Palo Alto, USA
  Dates Jul. 12 and 26, 1996

- **Title** Towards A-LISP: Automatic Language Independent Speech Processing
  Lecturer Gérard Chollet
  Location Bell Labs and CAIP, USA
  Dates Oct. 9 and 16, 1996

- **Title** 1. Towards User-Friendly Neural Networks / 2. Pushing the State-of-the-Art in Neural Computation
  Lecturer Dr. Emile Fiesler
  Location Brigham Young University, Utah
  Date October 31, 1996

- **Title** Vers une reconnaissance de la parole indépendante de la langue à l'aide d'un vocabulaire flexible
  Authors Dan Andrei Constantinescu
  Location ENST, Paris France
  Dates Nov. 28, 1996

- **Title** L'analyse logique de données : un outil pour l'induction de connaissance sous forme intelligible
  Speaker Dr. Eddy Mayoraz
  Location Laboratoire d'Informatique Théorique, École Polytechnique Fédérale de Lausanne
  Date Dec. 10, 1996

3.3 Examinations

- **School** École supérieure d'informatique de gestion (ESIS), Sierre
  Subject Practical projects
  Expert Dr. Jean-Luc Cochard
  Candidates Diploma final examination session after 2 years
  Date Jan. 13, 1996

- **School** Ecole Polytechnique Fédérale de Lausanne
  Subject Diploma project
  Expert Dominique Genoud
  Candidate Andreas Stuber
  Diploma Project Title Adaptative speech recognition techniques
  Date March. 14, 1996

- **School** École supérieure d'informatique de gestion (ESIS), Sierre
  Subject Practical projects
  Expert Dominique Genoud
  Candidates Diploma final examination session after 2 years
  Date June 25, 1996

- **School** Université Henri Poincaré, Nancy I
  Subject Thesis report and examination
  Expert Dr. Jean-Luc Cochard
  Candidate Roselyne Nguyen
  Thesis title Un système multi-agent pour la machine à dicter vocale MAUD : conception et intégration d'une source de connaissances phonologiques
  Date Oct. 28, 1996

- **School** Université d'Avignon et des Pays de Vaucluse
  Subject Thesis examination
  Expert Dr. Jean-Luc Cochard
  Candidate Olivier Oppizzi
  Thesis title Décision incertaine pour la réestimation dans une application de reconnaissance de la parole en mots isolés
  Date Nov. 15, 1996

- **School** Institut national polytechnique de Grenoble
  Subject Thesis report and examination
  Expert Dr. Jean-Luc Cochard
  Candidate Bertrand Caillaud
  Thesis title Apprentissage de connaissances prosodiques pour la reconnaissance automatique de la parole
  Date Dec. 20, 1996

- **School** EPFL
  Subject Artificial and Biological Neural Networks
  Expert Dr. Eddy Mayoraz
  Candidates Diploma final examination session, students in mathematics, computer science and electrical engineering
  Date September 1996

3.4 Student Projects

- **Trainee** Jean-Luc Beuchat
  School EPFL
  Formation Diploma Thesis
  Subject Character Recognition Neural Network
  Duration Oct. 1996 –
  Responsible Georg Thimm, Dr. Gilbert Maître, and Dr. Emile Fiesler

- **Trainee** Jan Erikson
  School EPFL
  Formation Post-diploma internship
  Subject Speaker parametrization using wavelets
  Duration Apr. 1, 1996 – Apr. 30, 1996
  Responsible Dominique Genoud

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

- **Trainee** Marco Grit
  School Technical University, Eindhoven, The Netherlands
  Formation Pre-Diploma Internship
  Subject Superceptron Topology Initialization
  Duration Sept. – Nov. 1996
  Responsible Dr. Emile Fiesler

- **Trainee** Hans Jongebloed
  School Cognitive Science and Engineering, University of Groningen (NL)
  Formation diploma work
  Subject Using a Genetic Algorithm to detect basic patterns in speech signals
  Duration June – Sept. 1996
  Responsible D. A. Constantinescu

- **Trainee** Hans Jongebloed
  School Cognitive Science and Engineering, University of Groningen (NL)
  Formation diploma work
  Subject User-Friendly Vocal Servers
  Duration Sept. – Dec. 1996
  Responsible Dr. Jean-Luc Cochard

- **Trainee** Katrin Keller
  School University of Illmenau, Germany
  Formation Pre-diploma work
  Subject Discrete Ontogenic Neural Networks for Telecommunication
  Duration Aug, 1995 – Feb. 1996
  Responsible Dr. Emile Fiesler

- **Trainee** Sacha Krstulović
  School ESTACA / Univ. Cergy-Pontoise
  Formation DEA
  Subject Inversion acoustico-articulatoire du modèle DRM
  Duration Apr. – Aug. 1996
  Responsible Dr. Gérard Chollet, Dominique Genoud

- **Trainee** Tomas Lundin
  School Uppsala University, Sweden
  Formation Diploma Thesis
  Subject Ontogenic Neural Network Quantization
  Duration July 1996 –
  Responsible Dr. Emile Fiesler and P. Moerland

- **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** Florian Salamin
  School EIV
  Formation Pre-diploma work and diploma work
  Subject Datapump Full-Duplex and Echo Cancellation on Telephone Lines
  Duration March – Dec. 1996
  Responsible F. Corthay (EIV), O. Bornet, Dr. Jean-Luc Cochard

- **Trainee** Jaime Edmundo Soares Reis da Silva
  School University of Porto, Portugal
  Formation Post-diploma internship
  Subject Speech Parametrisation Based on Evolution of Coefficients
  Duration May 1996 – Aug. 1996
  Responsible Dan Andrei Constantinescu

- **Trainee** Robbert Visscher
  School University of Groningen, The Netherlands
  Formation Diploma Thesis
  Subject Superceptron Growing
  Duration May 1996 –
  Responsible Dr. Emile Fiesler

- **Trainee** Peter Weber
  School University of Illmenau, Germany
  Formation Pre-diploma work
  Subject User-Interface Management System for Neural Network Simulator
  Duration Aug. 1995 – Feb. 1996
  Responsible Dr. Emile Fiesler

4 Other Scientific Activities

4.1 Scientific Committees

- **Name** Dr. Jean-Luc Cochard
  Function Member of the Scientific Committee
  Conference XXIème Journées d'Étude sur la Parole, Avignon

- **Name** Dr. Emile Fiesler
  Function Editor-in-Chief
  Book title Handbook of Neural Computation
  Publisher Oxford University Press and Institute of Physics
  ISBN 0–7503–0312–3 and 0–7503–0413–8
  Note Book version published in September 1996; Electronic version expected in early 1997.

- **Name** Georg Thimm
  Function Current Events Editor
  Journal Neurocomputing
  Publisher Elsevier
  ISSN 0925–2312

- **Name** Dr. Eddy Mayoraz
  Function Member of the Editorial Board
  Conference European Symposium of Artificial Neural Networks

4.2 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., Swiss Telecom
  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)

4.3 Collaboration with Local Companies

- **Company's Name** GPIL, Gianni Pante Ingénierie du Logiciel
  Company's Address 6, rue du Grand Verger, 1920 Martigny
  Description Image classification by neural networks for the quality control of watches.
  Internal staff Dr. Emile Fiesler and Miguel Moreira

5 Events and Presentations

- **Event** MicroNeuro'96, Lausanne, Feb. 12-14, 1996
  Speaker Perry Moerland [p-MF96]

- **Event** COST-249 meeting, Kosice, Slovakia, Feb. 28– Mar.1, 1996
  Speaker Dan Andrei Constantinescu [p-CBCC96]

- **Event** Meeting of the "Société académique du Valais", Institut Kurt Boesch, Sion, March 27, 1996
  Speaker Jean-Luc Cochard, *Presentation of IDIAP*

- **Event** M2VTS meeting, Madrid, Spain, April 22–23, 1996
  Speaker Jürgen Lüttin, *Active Shape Models for Visual Feature Extraction*

- **Event** meeting IDSIA–IDIAP, Lugano, April 19, 1996
  Purpose Presentation of IDIAP activities
  Speakers Jean-Luc Cochard, Dan A. Constantinescu, Emile Fiesler, Dominique Genoud, Eddy Mayoraz, Indu Saxena

- **Event** 3rd Slovenian-German and 2nd SDRV Workshop, Speech and Image Understanding, Ljubljana, Slovenia, Apr. 24 – 26, 1996
  Speaker Dan Andrei Constantinescu [p-CC96]

- **Event** ESANN 96, Bruges, April 24-26, 1996
  Speaker Eddy Mayoraz [p-May96]

- **Event** ICASSP'96, Atlanta, GA, May 7-10, 1996
  Speaker Jürgen Lüttin [p-LTB96f]

- **Event** SIWORK 96, Zuerich, Switzerland, May 14 – 15, 1996
  Speaker Jean-Luc Cochard, [p-DCC+96]

- **Event** ICNN'96, Washington D.C., USA, June 2–6, 1996
  Speaker Emile Fiesler [p-CF96]

- **Event** Journées d'Etudes sur la Parole, Avignon, France, June 10-14, 1996
  Speakers Dominique Genoud [p-GGBC96], Murielle Vial [p-CV96], Philippe Langlais [p-LMC96]

- **Event** COST-249 meeting, Stockholm, June 17–19, 1996
  Speaker Jean-Luc Cochard [c-VC96]

- **Event** COST-249/250 meeting, Stockholm, Sweden, Jun. 17–19, 1996
  Speakers Jean Hennebert, Dominique Genoud, Hakan Melin [p-GHH96]

- **Event** M2VTS meeting, Porto Carras, Greece, June 24–26, 1996
  Speaker Gilbert Maître, *Decision in speaker verification and perspectives for multi-modal verification*, Hubert Wassner *Speaker verification*

- **Event** CARNAC seminar, Lausanne, June 27, 1996
  Speaker Perry Moerland, *Non-Negative Neural Networks and their Application to Optical Neural Network Implementations*

- **Event** 4ème Colloque National sur l'Écrit et le Document (CNED'96), Nantes, France, July 3–5, 1996
  Speaker Gilbert Maître [p-MBP96]

- **Event** ICPR'96, Vienna, Austria, August 25-29, 1996
  Speaker Jürgen Lüttin [p-LTB96a, p-LTB96b]
- **Event** International Symposium on Neuro-Fuzzy Systems, Lausanne, Aug. 29–31, 1996
  Speaker Perry Moerland [p-MFS96]
- **Event** AIHENP'96, Lausanne, Switzerland, September 2–6, 1996
  Speaker Michel Maignan [p-FM97]
- **Event** M2VTS demonstrator integration, Sept. 2–10, 1996
  Visitor Franck Rigoulet from MATRA at IDIAP
- **Event** M2VTS level-two demonstrator installation, Sept. 25 – Oct. 1, 1996
  Visitor Gilbert Maître at MATRA
- **Event** EUSIPCO'96, Trieste, Italy, September 10-13, 1996
  Speaker Jürgen Lüttin [p-LTB96e]
- **Event** Invited Talk, IRST, Trento, Italy, September 12, 1996
  Speaker Jürgen Lüttin, *Spatio-Temporal Lip Information for Person Recognition*
- **Event** Foire du Valais (in cooperation with the CREM), Martigny, Sept. 27 – Oct. 6, 1996
  Purpose Presentation of Interactive Voice Servers Technology
- **Event** IVTTA 96, Basking Ridge, NJ, USA, September 30 – October 1, 1996
  Speaker Gérard Chollet, [p-BCC+96]
- **Event** SIPAR'96, Geneva, Switzerland, October 4, 1996
  Speaker Tomas Lundin [p-LFM96], Robbert Visscher [p-VFT96]
- **Event** ICSLP 96, Philadelphia, USA, Oct. 3 – 6, 1996
  Speaker Dan Andrei Constantinescu [p-CBCC96]
- **Event** ICSLP'96, Philadelphia, PA, October 3-6, 1996
  Speaker Jürgen Lüttin, *Speaker Identification by Lipreading, Speechreading Using Shape and Intensity Information*
- **Event** *IEEE* International Conference on Electronics, Circuits, and Systems (ICECS'96), Rhodos, Greece, Oct. 13–16, 1996
  Speaker Perry Moerland [p-SFM96]
- **Event** COST-249 meeting, Zürich, Oct. 17–18, 1996
  Speaker Philippe Langlais
- **Event** ISAI'96, Cancun, Mexico, November 12–15, 1996
  Speaker Emile Fiesler [p-MFP96]
- **Event** ICTAI'96, Toulouse, France, November 16–19, 1996
  Speaker Georg Thimm [p-TF96a]

6 Publications

Books and Book Chapters

[b-FB96] E. Fiesler and R. Beale, editors. *Handbook of Neural Computation.* The Computational Intelligence Library. Institute of Physics and Oxford University Press, New York 1996. The electronic version is expected in early 1997.

[b-FC96] E. Fiesler and K. Cios. Supervised ontogenic networks. In E. Fiesler and R. Beale, editors, *Handbook of Neural Computation*, The Computational Intelligence Library, chapter C1.7. New York, 1996.

[b-Fie96] E. Fiesler. Neural network topologies. In E. Fiesler and R. Beale, editors, *Handbook of Neural Computation*, The Computational Intelligence Library, chapter B2. New York, 1996.

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

[b-LTB96] Juergen Luettin, Neil A. Thacker, and Steve W. Beet. Active shape models for visual speech feature extraction. In D. G. Storck and M. E. Hennecke (editors), editors, *Speechreading by Humans and Machines*, volume 150 of *NATO ASI Series, Series F: Computer and Systems Sciences*, pages 383–390. Springer Verlag, Berlin, 1996.

[b-LVB96] Juergen Luettin, Michael Vogt, and Christoph Bregler. Machine recognition and applications. In D. G. Storck and M. E. Hennecke (editors), editors, *Speechreading by Humans and Machines*, volume 150 of *NATO ASI Series, Series F: Computer and Systems Sciences*, pages 549–555. Springer Verlag, Berlin, 1996.

[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.

[b-NCL+96] F. Néel, G. Chollet, F. Lamel, W. Minker, and A. Constantinescu. *Reconnaissance et comprehension de la parole: évaluation et apllications*, pages 31–367. AUPELF – UREF, 1996.

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

Articles in International Journals

[a-Bre96] Thomas M. Breuel. Finding lines under bounded error. *Pattern Recognition*, 29(1): 167–178, January 1996.

[a-CF96] S. Cuche and E. Fiesler. Generalized cauchy machines. *Neurocomputing*, 1996. submitted.

[a-LT97] J. Luettin and N. A. Thacker. Speechreading using probabilistic models. *Computer Vision and Image Understanding*, 65, 1997. to appear.

[a-MA96] Eddy Mayoraz and Frédéric Aviolat. Constructive training methods for feedforward neural networks with binary weights. *International Journal of Neural Systems*, 7(2): 149–166, 1996.

[a-May96] Eddy Mayoraz. On the power of democratic networks. *SIAM Journal of Discr. Math*, 9(2): 258–268, 1996.

[a-MFS96] P. Moerland, E. Fiesler, and I. Saxena. Incorporation of liquid-crystal light valve nonlinearities in optical multilayer neural networks. *Applied Optics*, 35(26): 5301–5307, 1996.

[a-Moe96] P. D. Moerland. A review of MicroNeuro'96, February 12-14, 1996, Lausanne, Switzerland. *Neurocomputing*, 12(4): 371–373, August 1996.

[a-SF95] I. Saxena and E. Fiesler. Adaptive multilayer optical neural network with optical thresholding. *Optical Engineering*, 34(8): 2435–2440, August 1995. Invited paper.

[a-ST96] I. Saxena and R. B. Torbert. Time resolved polarimetry on an optical fiber ammeter. *Journal of the European Optical Society*, 5: 323–330, 1996.

[a-TF97] G. Thimm and E. Fiesler. High order and multilayer perceptron initialization. *IEEE Transactions on Neural Networks*, 8(2), February 1997.

[a-Thi95] Georg Thimm. Calendar of meetings. *Neurocomputing*, from 1995.

[a-TMF96] G. Thimm, P. Moerland, and E. Fiesler. The interchangeability of learning rate and gain in backpropagation neural networks. *Neural Computation*, 8(2): 451–460, February 1996.

Articles in Conference Proceedings

[p-BCC+96] Olivier Bornet, Gérard Chollet, Jean-Luc Cochard, Andrei Constantinescu, and Dominique Genoud. Secured vocal access to telephone servers. In *Proceedings of IVTTA 96 IEEE Third Workshop Interactive Voice Technology for Telecommunications Applications*, pages 41–44, 1996.

[p-CBCC96] Andrei Constantinescu, Olivier Bornet, Gilles Caloz, and Gérard Chollet. Validating different flexible vocabulary approaches on the swiss french polyphone and polyvar databases. In *Proceedings of ICSLP 96*, pages 2293–2296, 1996.

[p-CC96] Andrei Constantinescu and Gérard Chollet. Swiss polyphone and polyvar: Building databases for speech recognition and speaker verification. In *Proceedings of The 3rd Slovenian-German and 2nd SDRV Workshop, Speech and Image Understanding*, April 24-26 1996.

[p-CF96] S. Cuche and E. Fiesler. Extended cauchy machines. In *Proceedings of the International Conference on Neural Information Processing*, volume 1, pages 275–280, 1996.

[p-CV96] Jean-Luc Cochard and Murielle Vial. Etc_vérif : un environnement multi-agents de reconnaissance automatique de la parole en continu. In *Proceedings of JEP'96: XXIèmes Journées d'étude sur la Parole*, pages 297–300, 1996.

[p-DCC+96] Andrzej Drygajlo, Jean-Luc Cochard, Gérard Chollet, Olivier Bornet, and Philippe Renevey. Sun workstation and swissnet platform for speech recognition and speaker verification over the telephone. In *Proceedings of Workstations und ihre Anwendungen, SIWORK'96*, pages 1–4, May 1996.

[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. Springer Verlag, 1997. 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 (AIHENP)*, Amsterdam, The Netherlands, 1997.

[p-GBGC96] Dominique Genoud, Frédéric Bimbot, Guillaume Gravier, and Gérard Chollet. Combining methods to improve speaker verification decision. In *Proceedings of the Fourth International Conference on Spoken Language Processing (ICSLP'96)*, Philadelphia, October 3-6 1996.

[p-GGBC96] Dominique Genoud, Guillaume Gravier, Frédéric Bimbot, and Gérard Chollet. Amelioration des performances de verification du locuteur par combinaison de methodes. In JEP, editor, *Journees d'etudes sur la parole*, Avignon, June 1996.

[p-GHH96] D. Genoud, J. Hennebert, and H.Melin. Polycost database. Stockholm, June 1996. Joint meeting cost249-250.

[p-JLGW97] 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. Springer Verlag, 1997. to appear.

[p-LFM96] T. Lundin, E. Fiesler, and P. Moerland. Connectionist quantization functions. In *Proceedings of the '96 SIPAR-Workshop on Parallel and Distributed Computing*, pages 33–36. Scientific and Parallel Computing Group, University of Geneva, 1996.

[p-LMC96] Philippe Langlais, Henri Méloni, and Jean-Luc Cochard. Un système prédictif de la structuration syntaxico-rythmique d'un énoncé à l'aide d'informations prosodiques. In *Proceedings of JEP'96: XXIèmes Journées d'étude sur la Parole*, pages 211–214, 1996.

[p-LTB96a] Juergen Luettin, Neil A. Thacker, and Steve W. Beet. Learning to recognise talking faces. In *Proceedings of the International Conference on Pattern Recognition (ICPR'96)*, volume IV, pages 55–59. IAPR, 1996.

[p-LTB96b] Juergen Luettin, Neil A. Thacker, and Steve W. Beet. Locating and tracking facial speech features. In *Proceedings of the International Conference on Pattern Recognition (ICPR'96)*, volume I, pages 652–656. IAPR, 1996.

[p-LTB96c] Juergen Luettin, Neil A. Thacker, and Steve W. Beet. Speachreading using shape and intensity information. In *Proceedings of the 4th International Conference on Spoken Language Processing (ICSLP'96)*, volume 1, pages 58–61, 1996.

[p-LTB96d] Juergen Luettin, Neil A. Thacker, and Steve W. Beet. Speaker identification by lipreading. In *Proceedings of the 4th International Conference on Spoken Language Processing (ICSLP'96)*, volume 1, pages 62–65, 1996.

[p-LTB96e] Juergen Luettin, Neil A. Thacker, and Steve W. Beet. Statistical lip modelling for visual speech recognition. In *Proceedings of the 8th European Signal Processing Conference (Eusipco'96)*, volume I, pages 137–140, 1996.

[p-LTB96f] Juergen Luettin, Neil A. Thacker, and Steve W. Beet. Visual speech recognition using active shape models and hidden markov models. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'96)*, volume 2, pages 817–820, 1996.

[p-May96] Eddy Mayoraz. Bounds on the degree of high order binary perceptrons. In François Blayo and Michel Verleysen, editors, *Proceedings of ESANN'96*. D facto, 1996.

[p-MBP96] Gilbert Maître, Stéphane Brunet, and Gianni Pante. Traitement préliminaire de l'image d'un texte manuscrit en vue de sa reconnaissance: une méthode de sur-segmentation. In *4ème Colloque National sur l'Écrit et le Document (CNED'96)*, pages 85–92, 1996.

[p-MF96] P. D. Moerland and E. Fiesler. Hardware-friendly learning algorithms for neural networks: An overview. In *Proceedings of the Fifth International Conference on Microelectronics for Neural Networks and Fuzzy Systems: MicroNeuro'96*, pages 117–124, 1996.

[p-MFP96] M. Moreira, E. Fiesler, and G. Pante. Image classification by neural networks for the quality control of watches. In R. Soto, J. M. Sanchez, M. Campbell, and F. J. Cantu, editors, *Proceedings ISAI/IFIS'96*, pages 141–149. ITESM, 1996.

[p-MFS96] P. Moerland, E. Fiesler, and I. Saxena. Overcoming inaccuracies in optical multilayer perceptrons. In *Proceedings of the First International Symposium on Neuro-Fuzzy Systems (AT'96)*. AATI, 1996.

[p-PTF96] A. De Pol, G. Thimm, and E. Fiesler. Sparse initial topologies for high order perceptrons. In *Proceedings of the International Conference on Neural Networks*. IEEE, June 1996.

[p-SFM96] I. Saxena, E. Fiesler, and P. Moerland. A method for all-positive optical multilayer perceptrons. In *Proceedings of the Third IEEE International Conference on Electronics, Circuits, and Systems*, volume 1, pages 448–451. IEEE, 1996.

[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. Accepted for publication.

[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*. 1997. Accepted for publication.

[p-TF96a] G. Thimm and E. Fiesler. A boolean approach to construct neural networks for non-boolean problems. In *Proceedings of the 8th IEEE International Conference on Tools with Artificial Intelligence*. November 1996.

[p-TF96b] G. Thimm and E. Fiesler. Neural network pruning and pruning parameters. In Takeshi Furuhashi, editor, *The 1st Workshop on Soft Computing*, Furo-cho, Chikusa-ku, Nagoya 464-01, Japan, August 1996. Dept. of Information Electronics Nagoya University. Is published as http: //www.bioele.nuee.nagoya-u.ac.jp/wsc1/.

[p-VFT96] R. Visscher, E. Fiesler, and G. Thimm. Superceptron construction. In *Proceedings of the '96 SIPAR-Workshop on Parallel and Distributed Computing*, pages 37–40. Scientific and Parallel Computing Group, University of Geneva, 1996.

[p-WC96] Hubert Wassner and Gérard Chollet. New time-frequency derived cepstral coefficients for automatic speech recognition. In *Proceedings of the 8th European Signal Processing Conference (Eusipco'96)*, volume III, pages 1623–1626, 1996.

IDIAP Research Reports

[r-CCC+96] Gérard Chollet, Jean-Luc Cochard, Andrei Constantinescu, Cédric Jaboulet, and Philippe Langlais. Swiss french polyphone and polyvar: telephone speech databases to model inter- and intra-speaker variability. IDIAP-RR 1, IDIAP, 1996.

[r-GGBC96] Dominique Genoud, Guillaume Gravier, Frédéric Bimbot, and Gérard Chollet. Combining methods to improve speaker verification decision. IDIAP-RR 2, IDIAP, 1996.

[r-May96a] Eddy Mayoraz. On the complexity of the class of regions computable by a two-layered perceptron. IDIAP-RR 3, IDIAP, 1996.

[r-BCC+96] Olivier Bornet, Gérard Chollet, Jean-Luc Cochard, Andrei Constantinescu, and Dominique Genoud. Secured vocal access to telephone servers. IDIAP-RR 4, IDIAP / CNRS, 1996.

[r-BHI+96] Endre Boros, Peter L. Hammer, Toshihide Ibaraki, Alexander Kogan, Eddy Mayoraz, and Ilya Muchnik. An implementation of logical analysis of data. IDIAP-RR 5, RUTCOR / IDIAP, 1996.

[r-May96b] Eddy Mayoraz. On variations of the convex hull operator. IDIAP-RR 6, IDIAP, 1996.

[r-BDR96] Hervé Bourlard, Stéphane Dupont, and Christophe Ris. Multi-stream speech recognition. IDIAP-RR 7, IDIAP, 1996.

[r-MM96] Eddy Mayoraz and Miguel Moreira. On the decomposition of polychotomies into dichotomies. IDIAP-RR 8, IDIAP, 1996.

[r-fB96] Vincent fontaine and Hervé Bourlard. Speaker-dependent speech recognition based on phone-like units models — application to voice dialing. IDIAP-RR 9, IDIAP, 1996.

[r-MFP96] M. Moreira, E. Fiesler, and G. Pante. Image classification by neural networks for the quality control of watches. IDIAP-RR 10, IDIAP, 1996.

IDIAP Communications

[c-CFMM96] Jean-Luc Cochard, Emile Fiesler, Gilbert Maitre, and Eddy Mayoraz. Activity report 1995. IDIAP-Com 1, IDIAP, 1996.

[c-SFBC96b] Florian Salamin, cois Corthay Fran Olivier Bornet, and Jean-Luc Cochard. Datapump full-duplex. IDIAP-Com 2, EIV / IDIAP, 1996.

[c-Krs96] Sacha Krstulović. Présentation du modèle drm. IDIAP-Com 3, IDIAP, April 1996.

[c-Jon96] Hans Jongebloed. Voicephone: An interactive vocal server for telephone numbers. IDIAP-Com 4, IDIAP and CSE, University of Groningen, 1996.

[c-VC96] Murielle Vial and Jean-Luc Cochard. Towards a multi-agents approach for understanding speech. IDIAP-Com 5, IDIAP, 1996.

[c-SFBC96a] Florian Salamin, cois Corthay Fran Olivier Bornet, and Jean-Luc Cochard. Annulation d'écho sur une ligne téléphonique. IDIAP-Com 6, EIV / IDIAP, 1996.