



INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Avant-projet PAROLE

Analyse, Perception et Reconnaissance de la parole

Nancy

THÈME 3A

*R*apport
*d'*Activité

1999

Table des matières

1	Composition de l'équipe	3
2	Présentation et objectifs généraux	4
3	Fondements scientifiques	4
3.1	Analyse de la parole	6
3.1.1	Perception	6
3.1.2	Aides auditives	6
3.1.3	Inversion articulatoire	7
3.2	Reconnaissance automatique de la parole	7
3.2.1	Modèles acoustiques	8
3.2.2	Modèles de langage	8
4	Domaines d'applications	9
5	Logiciels	9
5.1	Outils logiciels	9
5.1.1	Snorri	9
5.1.2	Étiquetage de corpus écrits pour la reconnaissance	10
5.1.3	Classifieur automatique de lexique	10
5.1.4	SALT	10
5.1.5	LIPS	10
5.1.6	VINICS	10
5.1.7	ESPERE	11
5.2	Corpus	11
6	Résultats nouveaux	12
6.1	Analyse de la parole	12
6.1.1	Aides auditives	12
6.1.2	Inversion articulatoire	13
6.1.3	Enseignement des sciences de la parole	13
6.2	Reconnaissance automatique de la parole	14
6.2.1	Modèles acoustiques	14
6.2.2	Modèles de langage	15
7	Contrats industriels (nationaux, européens et internationaux)	16
7.1	Parrot	16
7.2	Procoma	16
7.3	Contrat Télécom-PTT Suisses - IDIAP	16

8	Actions régionales, nationales et internationales	16
8.1	Actions régionales	16
8.1.1	Traitement et interprétation de signaux médicaux	16
8.2	Actions nationales	17
8.2.1	Projet du pôle européen de santé sur les aides auditives	17
8.2.2	Action coopérative INRIA sur un système générique de reconnaissance de parole	17
8.3	Actions européennes	17
8.3.1	Projet Européen COST 249 et COST 250	17
8.3.2	AUPELF-UREF	17
8.3.3	VODIS	18
8.3.4	ISAEUS	18
8.4	Visites, et invitations de chercheurs	18
9	Diffusion de résultats	18
9.1	Animation de la Communauté scientifique	18
9.2	Enseignement universitaire	19
9.3	Participation à des colloques, séminaires, invitations	19
10	Bibliographie	19

PAROLE est un avant-projet du LORIA (UMR 7503) commun au CNRS, à l'INRIA, à l'Université Henri POINCARÉ Nancy 1, à l'Université Nancy 2 et à l'Institut National Polytechnique de Lorraine.

1 Composition de l'équipe

Responsable scientifique

Yves Laprie [Chargé de Recherche, CNRS]

Assistants de projet

Martine Kuhlmann [CNRS]

Personnel CNRS

Anne Bonneau [Chargée de Recherche]

Dominique Fohr [Chargé de Recherche]

Personnel INRIA

Khalid Daoudi [Chargé de Recherche, depuis le 1er octobre 1999]

Personnel Université

Jean-Paul Haton [Professeur, U. H. Poincaré, Nancy 1, Institut Universitaire de France]

Marie-Christine Haton [Professeur, U. H. Poincaré, Nancy 1]

Irina Illina [Maître de conférences, I.U.T Charlemagne, U. Nancy 2]

Joseph di Martino [Maître de conférences, U. H. Poincaré, Nancy 1]

Odile Mella [Maître de conférences, U. H. Poincaré, Nancy 1]

Kamel Smaïli [Maître de conférences, U. Nancy 2, délégué au CNRS depuis le 30 septembre]

Chercheurs doctorants

Armelle Brun [MENRT depuis le 1er novembre 1999]

Christophe Cerisara [MENRT jusqu'au 30 septembre 1999]

Vincent Colotte [MENRT]

Virginie Govaere [Bourse UHP]

David Langlois [MENRT]

Slim Ouni [Bourse franco-tunisienne]

Imed Zitouni [Bourse franco-tunisienne puis ATER]

Postdoctorants

Parham Mokhtari [INRIA, jusqu'au 1er mars 1999]

Ingénieurs sur contrat

Christophe Antoine [projet VODIS]

Collaborateurs extérieurs

Noëlle Carbonell [Professeur, U. H. Poincaré, Nancy 1]

Sylvie Coste [Maître de conférences, CRIL, Lens]

Marie-Madeleine Dutel [Service ORL, Hôpital de Nancy]

Jean-Claude Junqua [STL, Santa-Barbara, USA]

Jean-François Mari [Professeur, I.U.T. Charlemagne, U. Nancy 2]

2 Présentation et objectifs généraux

PAROLE est un projet commun à l'INRIA, au CNRS et à l'université Henri Poincaré via le laboratoire LORIA, UMR 7503.

Notre thème de recherche est l'étude de la communication parlée et recouvre un vaste spectre d'activités qui va de l'étude de la perception humaine des indices acoustiques jusqu'à l'étude de l'intégration de la reconnaissance de la parole dans des systèmes embarqués ou multimédia. Ce thème est largement pluridisciplinaire. Il fait appel à des compétences dans des domaines différents tels que la phonétique, les mathématiques, l'acoustique, la psychoacoustique, le traitement du signal, la reconnaissance des formes et la représentation des connaissances, et contribue à l'extension de certains modèles symboliques ou numériques. Ainsi, les modèles de Markov cachés initialement issus de la reconnaissance de la parole, se révèlent être des modèles efficaces dans d'autres domaines de recherche tels que le séquençement du génome, la robotique mobile ou la fouille de données temporelles.

La parole pose des problèmes fondamentaux et appliqués. L'aspect fondamental se rencontre dans nos recherches en phonétique et en informatique puisqu'il faut spécifier des modèles originaux de perception et de nouveaux algorithmes d'apprentissage, de reconnaissance et d'adaptation. L'aspect appliqué correspond à notre volonté de développer de nouveaux systèmes de reconnaissance automatique de la parole, d'apprentissage des langues ou d'aides auditives. Cet aspect nous permet d'évaluer nos résultats théoriques et d'améliorer la fiabilité, la robustesse et les performances de nos algorithmes.

3 Fondements scientifiques

Mots clés : traitement du signal, phonétique, perception, modèles stochastiques, modèles de langage, modèles articulatoires, apprentissage des langues, reconnaissance automatique de la parole, aides auditives, analyse de la parole, indices acoustiques.

Globalement les recherches sur la parole ont donné lieu à deux types d'approches :

- des recherches visant à expliquer comment la parole est produite et perçue, et incluant donc des aspects physiologiques (contrôle du conduit vocal), physiques (acoustique de la

parole), psychoacoustiques (système auditif périphérique), et cognitifs (construction des phrases),

- des recherches visant à modéliser les observations des phénomènes de la parole (analyse spectrale, modèles stochastiques acoustiques et linguistiques).

Les premières recherches sont motivées par la très grande spécificité de la parole parmi tous les signaux acoustiques : l'appareil de production de la parole est facilement accessible (du moins en première approche), les équations acoustiques relativement abordables d'un point de vue mathématique (au prix de simplifications qui ne sont que modérément restrictives), les phrases produites sont régies par le vocabulaire et la grammaire de la langue étudiée. Cela a donc conduit les acousticiens à développer des recherches visant à produire un signal de parole artificiel de bonne qualité, les phonéticiens des recherches visant à trouver l'origine de la variabilité des sons de la parole et à expliquer comment les articulateurs sont utilisés, comment les sons d'une langue s'organisent et comment ils s'influencent dans la parole continue. Enfin, cela a conduit les linguistes à mener des recherches pour savoir comment les phrases sont construites. Il est clair que cette approche donne lieu à de nombreux allers et retours entre la théorie et l'expérimentation et qu'il est difficile de maîtriser simultanément tous ces aspects de la parole.

Les résultats disponibles sur la production et la perception de la parole ne permettent cependant pas d'envisager une approche d'analyse par synthèse. La reconnaissance automatique a donc suscité une seconde approche consistant à modéliser les observations des phénomènes de la parole. Les efforts ont porté sur l'élaboration de modèles numériques (d'abord de simples vecteurs de formes spectrales et maintenant des modèles stochastiques ou neuromimétiques) des réalisations acoustiques des phonèmes ou des mots, et sur le développement de modèles de langages statistiques.

Ces deux approches sont complémentaires, la seconde empruntant à la première les résultats théoriques sur la parole et la première empruntant à la seconde certains outils numériques, les techniques d'analyse spectrale étant sans doute le domaine où les échanges sont les plus marqués. L'existence de ces deux approches est l'une des particularités des recherches en parole menées à Nancy et nous comptons renforcer les échanges entre ces deux types d'approches. Ces échanges sont d'ailleurs conduits à se multiplier depuis que les systèmes de reconnaissance automatique (en particulier destinés à la dictée automatique) sont disponibles pour le grand public : il faut augmenter leur robustesse du point de vue acoustique (robustesse au bruit, adaptation au locuteur) comme du point de vue linguistique.

Les activités de notre équipe se structurent suivant ces deux approches :

Production et perception Nos recherches portent sur l'analyse et la perception des indices acoustiques, l'inversion acoustico-articulatoire et l'analyse de la parole. Elles donnent lieu à un certain nombre d'applications en cours ou à venir : la rééducation vocale, l'amélioration des aides auditives, l'apprentissage des langues.

Modélisation de la parole pour la reconnaissance automatique Nos recherches portent sur les modèles stochastiques, les modèles de langage et les modèles multibandes, ce qui donne lieu à un certain nombre d'applications en cours ou à venir : la reconnaissance automatique de la parole, la dictée automatique, l'alignement texte-parole et la classification de signaux différents de la parole.

3.1 Analyse de la parole

Participants : Anne Bonneau, Jean-Paul Haton, Marie-Christine Haton, Yves Laprie, Joseph di Martino, Christophe Antoine, Vincent Colotte, Virginie Govaere, Parham Mokhtari, Slim Ouni.

3.1.1 Perception

Nous menons des études perceptives afin d'approfondir les connaissances sur les indices essentiels d'identification ainsi que sur les mécanismes de perception des sons de la parole. Les domaines d'application de nos travaux vont de la reconnaissance automatique de la parole aux domaines paramédicaux, comme l'aide aux malentendants, et aux logiciels d'aide à la prononciation.

Nos expériences passées ont concerné la perception du lieu d'articulation des occlusives sourdes du français, le rôle du contexte vocalique dans leur identification ainsi que l'identification de la voyelle à partir du bruit d'explosion de ces consonnes. Nous avons également étudié l'effet des modifications d'amplitude des formants sur la perception des voyelles. Nous savons que les fréquences formantiques ont un rôle déterminant dans la perception des voyelles et nous avons voulu approfondir le rôle de l'amplitude, un paramètre certes moins important, mais qui, pour certains formants et certaines oppositions vocaliques, peut se révéler également déterminant.

3.1.2 Aides auditives

Dans les aides conventionnelles, le signal est capturé à l'aide d'un microphone, reconditionné par l'aide auditive et diffusé dans l'oreille moyenne. Ces aides utilisent des techniques de filtrage et de contrôle automatique du gain. Le filtrage permet de décomposer le signal en bandes de fréquence traitées en parallèle et le contrôle automatique du gain permet de réduire la dynamique du signal afin d'assurer la perception de l'amplitude et de préserver le confort du patient. La qualité globale d'une aide auditive vient de la stratégie d'utilisation de ces outils de base. L'un des objectifs majeurs de la recherche sur les aides auditives est d'exploiter le mieux possible les spécificités de la parole pour guider les techniques de traitement du signal qui deviennent de plus en plus puissantes. Notre contribution intervient à deux niveaux : celui du diagnostic et celui des stratégies de correction du signal de parole.

En ce qui concerne le diagnostic, il apparaît qu'il faut compléter l'audiogramme tonal actuel mais en évitant de verser dans le développement de tests psycho-acoustiques souvent très lourds à mettre en œuvre et demandant une attention prolongée de la part du patient. Nous utilisons donc des stimuli artificiels mais construits à partir de la parole naturelle.

En ce qui concerne les transformations de la parole, il existe un certain nombre de pistes destinées à compléter les techniques actuelles. Les efforts les plus importants correspondent au renforcement des pics spectraux, le but étant de préserver la perception des pics malgré une perte de sélectivité fréquentielle ou temporelle.

3.1.3 Inversion articulatoire

Les travaux sur l'inversion acoustique articulatoire reposent largement sur une approche d'analyse par synthèse articulatoire qui recouvre trois aspects essentiels :

la résolution des équations de l'acoustique Pour résoudre les équations de l'acoustique adaptées au conduit vocal, on fait l'hypothèse que l'onde sonore est une onde plane dans le conduit vocal et que le conduit peut être redressé. Il existe deux grandes familles de résolutions : **(i)** fréquentielles grâce à l'analogie acoustico-électrique, **(ii)** spatio-temporelles, par la résolution directe des équations aux différences finies issues des équations de Webster.

les mesures du conduit vocal Cet aspect représente un obstacle important car il n'existe pas de méthode fiable pour mesurer le conduit vocal avec précision. L'IRM permet de mesurer le conduit vocal en 3D mais n'est pas assez rapide et les rayons X ne permettent que de récupérer une coupe sagittale du conduit vocal.

la modélisation articulatoire L'un des objectifs de la modélisation articulatoire est de décrire avec un petit nombre de paramètres les formes possibles du conduit vocal tout en préservant les déformations observées sur un conduit réel. Les modèles articulatoires actuels sont souvent le résultat d'analyses statistiques de films ciné-radiographiques, comme par exemple le modèle de Maeda.

L'une des difficultés majeures de l'inversion est qu'une infinité de formes de conduits peuvent donner un même spectre de parole. Les méthodes d'inversion acoustico-articulatoire s'organisent en deux familles :

- les méthodes d'optimisation d'une fonction combinant généralement l'effort articulatoire du locuteur et la distance acoustique entre la parole réelle et la parole synthétisée. Ces méthodes font appel à un certain nombre de contraintes permettant de réduire le nombre de formes de conduits possibles.
- les méthodes par tabulation. Ces méthodes reposent sur un dictionnaire de formes articulatoires indexées acoustiquement (généralement par les fréquences des formants). Après avoir récupéré à chaque instant les formes possibles, une procédure d'optimisation permet de trouver une solution d'inversion sous la forme d'un chemin optimal.

Comme notre contribution ne porte que sur l'inversion nous avons repris les méthodes de synthèse articulatoire les plus couramment utilisées. Nous utilisons donc le modèle articulatoire de Maeda, l'analogie acoustique électrique pour calculer le spectre de parole et une méthode spatio-temporelle pour produire le signal de parole.

Pour ce qui concerne l'inversion, nous avons choisi d'utiliser le modèle de Maeda pour contraindre les formes de conduit vocal. Ce choix assure que les phénomènes de synergie et de compensation articulatoire sont toujours possibles ce qui est important pour récupérer des mouvements articulatoires proches de ceux d'un locuteur humain.

3.2 Reconnaissance automatique de la parole

Participants : Dominique Fohr, Jean-Paul Haton, Irina Illina, Odile Mella, Kamel Smaïli, Christophe Antoine, Armelle Brun, Christophe Cerisara, David Langlois, Imed

Zitouni.

La reconnaissance automatique de la parole nécessite l'utilisation imbriquée de modèles acoustiques et de modèles de langage. Les modèles acoustiques permettent de prendre en compte des contraintes acoustiques et phonétiques au niveau d'un son ou d'un groupe de sons alors que les modèles de langage définissent les contraintes syntaxiques et sémantiques au sein d'un groupe de mots ou d'une phrase.

Malgré la forte imbrication entre ces deux types de modèles, nous les présentons dans deux paragraphes successifs pour plus de clarté.

3.2.1 Modèles acoustiques

Les techniques stochastiques sont actuellement les plus utilisées pour la modélisation acoustique de la parole. En effet, ce sont celles qui ont permis d'obtenir les meilleurs résultats en reconnaissance de mots isolés, mots enchaînés et parole continue dans des conditions de laboratoire ou en environnement non bruité. En revanche, dans des conditions réelles de traitement de la parole (milieu bruité, parole spontanée, prononciations diverses et variées, . . .), les performances obtenues par ces techniques sont fortement dégradées ce qui justifie nos recherches actuelles et futures.

Dans cette optique, notre groupe a développé deux classes de modèles stochastiques originaux pour la reconnaissance automatique de la parole : les modèles de Markov cachés du premier et du second ordre (Hidden Markov Models ou HMM), les modèles stochastiques de trajectoires (Stochastic Trajectory Modeling ou STM).

Les HMM nous ont permis de réaliser des systèmes de reconnaissance automatique de lettres épelées, de chiffres connectés ou de la parole continue. Les modèles d'ordre 2 généralisent les modèles d'ordre un. Comparés à ces derniers, les HMM2 permettent une modélisation plus précise des durées des segments sub-phonétiques et atteignent des performances de reconnaissance supérieures au prix d'une complexité supérieure (facteur 2 pour les temps de calcul).

Les modèles stochastiques de trajectoires utilisent une approche novatrice pour reconnaître la parole. Plutôt que d'analyser à intervalle de temps fixé le signal de parole, les STM modélisent la trajectoire du signal dans l'espace de représentation (fréquentiel ou cepstral). L'unité à reconnaître – le mot ou le phonème – est retrouvé grâce à une probabilité d'appartenance à une classe qui intègre les informations de durée et d'évolution des paramètres acoustico-phonétiques.

3.2.2 Modèles de langage

Les systèmes de dictée automatique donnent de bons résultats acoustiques, néanmoins plusieurs problèmes au niveau langagier n'ont toujours pas de solution.

La communauté scientifique travaillant sur la reconnaissance automatique de la parole a pris conscience qu'il devient indispensable de fournir plus d'efforts pour concevoir des modèles de langage plus performants et ayant une meilleure interaction avec les niveaux acoustiques. En effet, les modèles de langage d'aujourd'hui sont, dans la plupart des cas, des modèles stochastiques ayant une portée locale ou à court terme (les modèles avec mémoire cache). Même si ces systèmes donnent de bons résultats, il restent néanmoins limités et ont besoin

d'être constamment améliorés pour s'adapter à la complexité de la langue. Pour ce faire, nous avons accentué nos efforts dans ce domaine et nous avons développé plusieurs modèles, entre autres, ceux fondés sur la prise en compte d'historique de longueur variable. Ces modèles nous ont déjà permis d'améliorer nos résultats en reconnaissance et en prédiction dans le test de Shannon. Afin de maîtriser la complexité du langage, nous avons décidé de prospecter deux nouvelles voies de recherche : l'adaptation et l'introduction de la "sémantique statistique". La première consiste à adapter le modèle de langage dynamiquement au cours de la reconnaissance pour mieux prendre en compte les structures spécifiques de l'application cible. Quant à la seconde, elle consiste à essayer de modéliser quelques phénomènes sémantiques de la langue d'une manière statistique afin de lever certaines ambiguïtés et ainsi améliorer les taux de reconnaissance.

4 Domaines d'applications

Les domaines d'application de nos travaux vont de la reconnaissance automatique de la parole aux domaines para-médicaux. Les méthodes d'analyse de la parole contribueront au développement de nouvelles technologies concernant l'aide à la prononciation (par exemple pour les malentendants ou pour l'apprentissage des langues) et les systèmes d'aides auditives.

Par ailleurs, la parole a et aura un rôle de plus en plus important dans les modalités d'interaction homme-machine. En effet, l'expression orale en langue naturelle est un mode de communication susceptible de séduire le grand public, surtout dans un environnement multimodal où l'association à la parole de gestes de désignation sur un écran tactile permet notamment de simplifier l'interprétation des expressions linguistiques de référence spatiale. D'autre part, le recours à la parole s'impose dans de nombreuses applications nouvelles où l'usage du clavier est malaisé, voire impossible : informatique mobile ou embarquée, bornes interactives, informatique domestique, téléphone.

5 Logiciels

5.1 Outils logiciels

5.1.1 Snorri

Snorri est le logiciel d'étude de la parole que nous avons développé et amélioré depuis 10 ans. Il est destiné à faciliter le travail du chercheur en reconnaissance de la parole, en phonétique, en perception ou encore en traitement du signal. Les fonctions de base de Snorri permettent de calculer plusieurs types de spectrogrammes et d'éditer le signal de parole de manière très fine (couper, coller, filtrages et atténuations diverses) car le spectrogramme permet de connaître la répercussion acoustique de toutes les modifications. À cela s'ajoute un grand nombre de fonctions destinées à étiqueter phonétiquement ou orthographiquement des signaux de parole, des fonctions destinées à extraire la fréquence fondamentale de la parole, des fonctions destinées à piloter le synthétiseur de Klatt et d'autres à utiliser la synthèse PSOLA.

Snorri a servi de base logicielle pour un grand nombre de travaux dans notre équipe (suivi de formants, identification des occlusives, études perceptives, ...). Étant donné l'intérêt qu'il

représente pour l'étude de la parole nous l'avons diffusé auprès d'une quinzaine d'équipes francophones, dont celle du CNET de Lannion. Initialement développé sous Unix et Motif, nous l'avons porté sous Windows et nous le commercialisons depuis cet automne sous le nom de WinSnoori par l'intermédiaire de Babel Technologies (startup située à Mons en Belgique et vendant des logiciels de synthèse et de reconnaissance automatique de la parole).

5.1.2 Étiquetage de corpus écrits pour la reconnaissance

Nous avons développé un outil d'étiquetage permettant de résoudre syntaxiquement un texte. Il permet d'affecter à chaque mot d'une phrase la classe syntaxique en fonction du contexte dans lequel celui-ci apparaît. Cet outil d'étiquetage utilise, pour fonctionner, un dictionnaire de 230 000 formes ainsi qu'un jeu de classes syntaxiques comportant 230 étiquettes. Le taux d'erreur de l'étiqueteur est de 1 %.

5.1.3 Classifieur automatique de lexique

Pour adapter nos modèles de langage aux différentes applications de la dictée automatique, nous avons développé un outil permettant, à partir d'un vocabulaire donné et d'un corpus d'apprentissage, de proposer un jeu de classes permettant d'avoir un modèle de langage de perplexité minimale. Cet outil est fondé sur l'algorithme du recuit simulé et comprend plusieurs variantes : classification initiale aléatoire ou fixée, nombre de classes fixé, perplexité fixée, etc.

5.1.4 SALT

SALT (Semi-Automatic Labelling Tool) est un outil d'étiquetage semi-automatique de grands corpus oraux. À partir du texte de la phrase prononcée, d'un dictionnaire phonétique et de règles phonologiques, il génère un graphe des prononciations possibles pour une phrase. Ensuite, il effectue un alignement forcé de ce graphe sur le signal de parole grâce à des modèles de Markov du second ordre (algorithme de Viterbi). L'étiquetage est affiné itérativement à l'aide d'un logiciel de comparaison d'étiquetage.

5.1.5 LIPS

Dans le cadre de la réalisation de dessins animés, il est nécessaire de synchroniser le mouvement des lèvres des personnages avec la phrase prononcée par l'acteur. Cette phase, réalisée manuellement, peut maintenant être effectuée grâce à notre logiciel LIPS (Logiciel Intégré de Post-Synchronisation) qui permet l'alignement automatique d'un texte anglais ou français avec le signal audio correspondant.

5.1.6 VINICS

L'étude fait partie du projet IMAGIN mené par le CEA dans le domaine des bases de données de centrales nucléaires. Deux aspects de la reconnaissance ont été abordés :

- reconnaissance de la parole continue. Une version de notre système VINICS a été réécrite en C++ par les ingénieurs du CEA.

- Vérification du locuteur. Une interface graphique a été développée pour un de nos systèmes de vérification et l'ensemble a été livré au CEA-Cadarache.

5.1.7 ESPERE

Nous avons développé un moteur de reconnaissance de parole générique fondé sur les modèles de Markov cachés (HMM). Ce moteur ESPERE (Engine for SPEech REcognition) permet de reconnaître aussi bien des mots isolés, connectés, que des mots clefs ou de la parole continue. Entièrement développé en C++, il fonctionne sous UNIX ou sous Windows.

5.2 Corpus

Les recherches menées dans le domaine de la communication parlée ont un point commun, elles nécessitent l'enregistrement, la manipulation et le traitement de corpus de plus en plus importants.

Ainsi, pour mener à bien des études sur les indices phonétiques, il est nécessaire d'enregistrer et d'étiqueter phonétiquement de nombreuses phrases, afin de capturer le maximum d'effets contextuels ; mais ces phrases doivent aussi être prononcées par de nombreux locuteurs, afin cette fois-ci de capturer les variations interlocuteurs. Citons, dans ce cadre, notre participation au projet Européen VODIS pour l'enregistrement, la numérisation et l'étiquetage d'un corpus de plus de 200 automobilistes en conditions réelles.

Depuis plusieurs années déjà, nous avons développé des outils permettant d'éditer, de traiter et d'étiqueter manuellement de telles bases de données de parole comme Snorri présenté dans le paragraphe 5.1.

Un autre exemple concerne la constitution de grands corpus et leur étiquetage automatique en vue d'entraîner les systèmes de reconnaissance de parole faisant appel à des modèles statistiques, stochastiques ou neuromimétiques. En effet, pour évaluer les paramètres de ces modèles, il faut disposer d'une grande quantité de données d'apprentissage. Les modèles étant de plus en plus précis (contextuels, multigaussiennes,...), le nombre de paramètres libres, donc à apprendre, est devenu de plus en plus grand, nécessitant une augmentation considérable de la taille des corpus étiquetés. A l'heure actuelle, les corpus de parole continue contiennent de nombreuses heures de parole tels ceux du LIMSI (BREF 15 Go) de l'ARPA (Wall Street Journal 20 Go) ou des PTT Suisse (Polyphone 10 Go).

De tels corpus, de plusieurs dizaines de milliers de phrases, ne peuvent plus être étiquetés manuellement. Aussi avons nous développé des outils d'étiquetage semi-automatique de grands corpus (cf. paragraphe 6.2.1).

De la même façon, la manipulation de grands corpus de texte est indispensable pour la conception de modèles de langages probabilistes. Ainsi, dans le cadre de la machine à dicter (projet AUPELF-UREF), les modèles bi et trigrammes ont été évalués à partir d'un corpus de 50 millions de mots issus d'articles du journal "Le Monde".

La taille des corpus disponibles ne cesse d'augmenter. Aux Etats Unis, des corpus de plus de 300 millions de mots sont déjà distribués comme le "North American News Text". Il sera donc nécessaire d'améliorer continuellement les outils logiciels pour les traiter.

6 Résultats nouveaux

6.1 Analyse de la parole

6.1.1 Aides auditives

Diagnostic de déficiences auditives Afin d'améliorer le diagnostic des déficiences auditives et ainsi de permettre une correction plus fine de chaque type de déficience par des transformations appropriées du signal, nous construisons actuellement un corpus de sons synthétiques. Pour le moment, nous avons élaboré un corpus de voyelles à deux formants dont les fréquences ont été choisies de manière à mettre en évidence des déficits auditifs dans des régions bien déterminées du spectre. Grâce à un logiciel de simulation des déficiences auditives, élaboré dans notre équipe, nous avons testé la validité de notre diagnostic par une série d'expériences de perception faites sur des normo-entendants [13]. Un logiciel destiné aux audiométristes de l'Hôpital Central de Nancy, avec lesquels nous travaillons, permet de présenter nos stimuli à l'intensité requise par le déficit du patient, en écoute binaurale, monaurale ou dichotique (une partie du signal est présentée dans une oreille, l'autre partie dans l'autre oreille) ainsi que d'injecter du bruit dans le signal. Des tests sur des patients de l'hôpital seront effectués prochainement.

Transformations du signal de parole Par ailleurs, nous avons développé des outils issus de méthodes de synthèse qui nous permettront d'évaluer les transformations de parole susceptibles d'être utilisées par un système d'aide auditive. Nous avons choisi la méthode TD-PSOLA (Time Domain Pitch Synchronous Overlap and Add) pour sa simplicité de mise en oeuvre et en particulier son faible coût calculatoire. Cette méthode nécessite la connaissance préalable d'une décomposition du signal initial en fenêtres recouvrantes. Nous avons donc conçu un algorithme de pitch marquage qui trouve le centre des fenêtres temporelles synchronisées sur les périodes fondamentales. Cet algorithme a l'avantage d'exploiter les résultats de n'importe quel algorithme d'extraction de la fréquence fondamentale (en particulier celui développé cette année [16]). Avec ces outils, nous avons développé une transformation de la parole qui combine un ralentissement sélectif et le renforcement de certains indices acoustiques. Le ralentissement sélectif avait déjà été utilisé pour améliorer l'intelligibilité des bulletins d'information de la NHK, mais les modifications de débit intervenaient au niveau des groupes de mots pour s'adapter à la construction des phrases japonaises. Pour le français, nous avons préféré travailler au niveau phonétique en ralentissant les transitions acoustiques rapides. Les transitions sont détectées en évaluant une fonction de variation spectrale. Le réhaussement porte sur les bursts des occlusives sourdes et les fricatives sourdes. Il consiste à amplifier progressivement le signal temporel de ces indices détectés par examen de critères énergétiques. Le choix de la stratégie globale, en particulier du type d'événements auxquels correspondent les transformations, est destiné à limiter les conséquences perceptives d'erreurs de détection. Cette transformation de la parole peut être utilisée pour compenser la perte de sélectivité temporelle ou pour améliorer la compréhension orale lors de l'apprentissage d'une langue étrangère. Les premières expériences que nous venons de terminer confirment que cette transformation facilite la compréhension orale lors de l'apprentissage de l'anglais.

6.1.2 Inversion articulatoire

Notre approche de l'inversion acoustique articulatoire utilise une méthode variationnelle et le modèle articulatoire de Maeda. Le principal avantage de cette méthode est de combiner au même niveau les critères de proximité acoustique et de régularité des trajectoires articulatoires. Il n'est donc pas nécessaire d'utiliser deux étapes d'optimisation comme cela est souvent le cas.

La mise en oeuvre donne lieu à un processus itératif qui optimise une trajectoire articulatoire initiale obtenue avec une méthode par tabulation. Cette méthode nécessite la construction d'un dictionnaire de formes du conduit vocal indexées par les paramètres acoustiques, en l'occurrence les formants. La construction correspond en pratique à l'échantillonnage de l'espace articulatoire. Cet espace est vaste car le modèle articulatoire sous-jacent fait intervenir sept paramètres et il est donc impossible de l'échantillonner finement. Traditionnellement on se contente donc d'un échantillonnage aléatoire. Ayant constaté l'insuffisance d'une telle approche (les trajectoires initiales ne sont pas toujours satisfaisantes) nous avons donc développé une approche structurant l'espace articulatoire sous la forme d'un arbre d'hypercubes [23]. Un hypercube est une feuille de l'arbre s'il est possible de considérer que la relation articulatoire acoustique y est approximativement linéaire. Outre la construction de l'hypercube lui-même nous avons aussi développé les outils d'interpolation pour calculer les paramètres acoustiques d'un point quelconque de l'espace articulatoire et les outils d'inversion pour récupérer toutes les formes articulatoires à partir des paramètres acoustiques. Au-delà de l'utilisation du dictionnaire de formes lors de l'inversion, ce travail a l'intérêt de permettre l'étude complète du comportement acoustique du modèle articulatoire de Maeda, en particulier des aspects de compensation articulatoire.

6.1.3 Enseignement des sciences de la parole

Nous ne travaillons pas encore directement sur ce thème mais certaines de nos activités s'inscrivent dans la perspective de développer des outils d'apprentissage des langues ou de la phonétique. Outre les travaux sur l'amélioration de l'intelligibilité qui peuvent s'appliquer à l'amélioration de la compréhension orale d'une langue étrangère nous avons développé deux aspects cette année : l'adjonction à WinSnoori d'une interface graphique pour le synthétiseur à formants de Klatt et la construction d'un corpus destiné à l'enseignement de la phonétique du français.

L'interface graphique du synthétiseur de Klatt pour la version Windows de Snorri [22] permet de fixer tous les paramètres concernant les formants et intègre des outils de synthèse par copie pour construire des stimuli à partir de parole naturelle. WinSnoori a fait l'objet d'une présentation lors de la session "Education arena" de la conférence EuroSpeech'99 et été retenu (sous la forme d'une version de démonstration) pour le CD-ROM distribué à cette occasion.

Avec l'équipe de J. Vaissière, de l'ILPGA (Sorbonne Nouvelle), nous avons élaboré un corpus destiné à l'enseignement de la phonétique et du français langue étrangère. Ce corpus, gravé sur un CD, comporte tous les sons de la langue française prononcés isolément, en VCV, et dans des phrases qui présentent des phénomènes articulatoires intéressants. Un petit texte de parole spontanée y figure également [12].

6.2 Reconnaissance automatique de la parole

6.2.1 Modèles acoustiques

Reconnaissance de parole au téléphone L'essor du téléphone et des applications téléphoniques vocales place la reconnaissance de parole téléphonique comme une problématique incontournable pour notre groupe de recherche. Nous étudions l'influence de la paramétrisation acoustique et de la topologie des modèles de Markov sur le corpus de parole téléphonique SPEECHDAT. Dans ce cadre, nous commençons une collaboration avec la société MIC2 pour travailler notamment sur des corpus obtenus dans des application réelles.

Reconnaissance robuste de parole dans une voiture Les voitures sont de plus en plus souvent pourvues d'équipements de haute technologie (systèmes de navigation, téléphones,...) pour lesquels la commande vocale est incontournable. Nous cherchons à augmenter la robustesse de nos modèles acoustiques (HMM) aux bruits habituellement présents dans une voiture. Dans ce cadre, nous collaborons avec la société Parrot basée à Paris.

Décomposition/recombinaison multi-bandes La reconnaissance de parole Multi-Bande est une méthode basée sur des travaux psycho-acoustiques menés par Fletcher dans les années 50, et plus récemment par J. B. Allen. Fletcher et Allen suggèrent qu'un traitement indépendant de différentes bandes fréquentielles est réalisé dans la première partie du cortex auditif, traitement suivi d'une phase pendant laquelle les résultats sont recombines. Ce modèle a été implémenté récemment par notre équipe dans le cadre de la reconnaissance automatique de la parole, et consiste donc à utiliser plusieurs HMMs par phonème, chacun d'entre eux étant entraîné sur une seule plage fréquentielle. Les probabilités retournées par ces HMMs sont ensuite recombines afin d'obtenir un score unique de reconnaissance par phonème.

En dehors des motivations psycho-acoustiques originelles, l'intérêt de cette méthode réside essentiellement dans sa robustesse au bruit. En effet, un bruit n'affecte généralement qu'une zone fréquentielle limitée, et laisse donc les autres bandes intactes. Nous nous sommes attachés au cours de cette année à montrer l'utilité de cette méthode en milieu bruité, qui a donné de meilleurs résultats de reconnaissance qu'une méthode classique utilisant tout le spectre. Par ailleurs, nous avons testé différents modules de recombinaison (linéaire et neuronaux) en comparant leurs points forts et leurs faiblesses selon les conditions expérimentales. Nous avons également proposé deux algorithmes originaux : l'un pour permettre un apprentissage global des HMMs et du module de recombinaison, l'autre pour autoriser un asynchronisme entre les différentes bandes lors de la reconnaissance [15, 14].

Compensation non linéaire Nous avons appliqué à de la parole prononcée en milieu bruité (apparition de l'effet Lombard) une nouvelle méthode de compensation qui permet une transformation non linéaire, aussi bien dans le domaine fréquentiel que cepstral, de la parole ayant servi à l'apprentissage du système.

En se basant sur la méthode *Vector Taylor Series* (VTS), développé à *Carnegie Mellon University* et qui permet de compenser simultanément les bruits convolutif et additif, nous avons proposé une extension de cette méthode pour la compensation, segment par segment,

de la parole bruitée au lieu de trame par trame. Notre but était de prendre en compte les propriétés segmentales des phonèmes.

Étiquetage de grands corpus de parole La plupart des systèmes de reconnaissance automatique de parole continue sont fondés sur des méthodes statistiques qui nécessitent, pour leur apprentissage, de grands corpus de parole étiquetés phonétiquement de façon fiable. Pour de tels corpus, seul l'étiquetage automatique est envisageable mais celui-ci doit être suivi d'une phase de vérification manuelle. Afin de réduire cette phase de vérification manuelle, nous avons élaboré une méthodologie d'étiquetage fondée sur l'utilisation de deux étiqueteurs automatiques et d'un algorithme de décision prenant en compte des règles phonologiques et phonétiques. Cette méthodologie a été mise en œuvre dans le cadre d'une convention avec l'IDIAP et les PTT Suisses pour l'étiquetage du corpus Polyphone (50000 phrases).

Outre l'outil de comparaison, nous avons développé un étiqueteur automatique, le tout constituant l'outil SALT décrit au paragraphe 5.1.4.

Nous avons adapté une version pour la post-synchronisation de dessins animés pour la société PROCOMA (cf. paragraphe 5.1.5)

6.2.2 Modèles de langage

Les modèles de langage sur lesquels nous travaillons sont dans leur globalité d'ordre statistique auxquels on a ajouté des connaissances explicites de la langue. Les modèles statistiques sont une combinaison de modèles fondés sur des séquences de classes et de grammaires de longueur fixe et/ou variable. La nouveauté dans ces modèles est l'utilisation d'horizon variable dans l'estimation de la probabilité du mot à prédire. Nous avons, cette année, amélioré ces modèles qui ont donné de meilleurs résultats en terme de perplexité et de taux d'erreur [26]. Un autre intérêt de cette approche est la possibilité de pouvoir construire automatiquement des séquences de mots ou des concepts évolués du langage qui sont par la suite intégrés au dictionnaire de l'application et qui deviennent des unités lexicales à part entière.

Un des résultats de cette année a été la mise en place d'une approche fondée sur les modèles distants. En effet, ces modèles prennent en compte des co-occurrences de mots non adjacents. Cette notion permet en quelque sorte de pallier les insuffisances des modèles dits locaux, même si les modèles distants appartiennent à cette catégorie de modèles [21]. Ces modèles ont été implantés dans MAUD et ont été testés dans le cadre du jeu de Shannon. Nous essayons à l'heure actuelle de les adapter afin qu'il puissent prendre en compte les phénomènes d'accord.

Nous avons travaillé cette année à l'élaboration d'un modèle de langage sémantique. Pour ce faire, avec l'aide d'un étudiant en linguistique, nous avons développé un jeu de concepts sémantiques. Ces concepts vont servir à l'étiquetage des séquences de mots d'un corpus de textes. Ensuite, ce corpus va être utilisé pour estimer les probabilités du modèle sémantique. Actuellement, le problème est en cours de formalisation mathématique.

L'adaptation des modèles de langage passe, entre autres, par l'adaptation des classes syntaxiques. Cette adaptation se fait sur un corpus de développement qui permet d'ajuster les classes en fonction du domaine d'application en cours de traitement. Plusieurs variantes de cette classification fondée sur le recuit simulé ont été développées. Nous pouvons citer, entre autres, la classification lexicale en prenant comme point de départ une classification fondée sur

des critères linguistiques [24]. L'adaptation des modèles de langage passe également par l'identification thématique. Pour ce faire, nous avons déterminé un certain nombre de thèmes qui vont servir de catégories référentielles pour lesquelles on développe actuellement un modèle de langage par thème. Le corpus du Monde diplomatique nous a servi de base pour l'apprentissage des modèles de langage thématique.

Nous avons travaillé également sur l'amélioration de la coopération entre un modèle à base de classes et un modèle à base de grammaires. Contrairement à ce qui se fait habituellement, nous avons utilisé le modèle classe en prédiction et le modèle grammaire dans la recherche des mots candidats et donc dans le scoring des hypothèses [25]. Cette approche a permis de montrer l'efficacité du modèle classe dans la détection des classes des mots les plus vraisemblables.

7 Contrats industriels (nationaux, européens et internationaux)

7.1 Parrot

Nous avons débuté un contrat avec la société Parrot pour améliorer la robustesse de la reconnaissance de parole dans une voiture (cf. paragraphe 6.2.1)

7.2 Procoma

La société Procoma développe un progiciel permettant de concevoir et réaliser des dessins animés. Nous intervenons dans ce progiciel au niveau de la fourniture d'un outil de post-synchronisation en français et en anglais (cf. paragraphe 5.1.5)

7.3 Contrat Télécom-PTT Suisses - IDIAP

Le but de ce projet a été l'étiquetage semi-automatique du corpus téléphonique Polyphone contenant 50 000 phrases. Nous avons proposé une méthode d'étiquetage, par affinements successifs, fondée sur la comparaison des résultats de deux étiqueteurs automatiques. Nous avons fourni aux Télécom-PTT Suisses, un CDROM contenant l'étiquetage de Polyphone. Par ailleurs, ce corpus nous a servi à mettre au point nos modèles pour un système de reconnaissance de parole téléphonique.

8 Actions régionales, nationales et internationales

8.1 Actions régionales

8.1.1 Traitement et interprétation de signaux médicaux

COLLABORATION AVEC LE SERVICE D'ANESTHÉSIE-RÉANIMATION DU CHRU DE NANCY,
PR. MEISTELMANN

Dans le cadre de notre collaboration avec le service du Professeur Meistelman commencée en 97, nous avons réalisé en coopération avec le groupe MAIA une nouvelle interface permettant d'acquérir, de visualiser et de traiter des signaux électro-encéphalographiques lors d'une anesthésie.

8.2 Actions nationales

8.2.1 Projet du pôle européen de santé sur les aides auditives

180000 aides auditives sont prescrites chaque année en France mais un grand nombre de patients (environ 40%) ne les utilisent pas parce qu'elles ne leur apportent pas les améliorations attendues. Ces échecs sont dûs à deux raisons principales :

- les audiogrammes tonal et vocal sont insuffisants pour établir un diagnostic précis des déficiences auditives,
- les aides actuelles n'utilisent que de simples filtrages.

Nous avons donc lancé un projet, en collaboration avec l'équipe d'audiologie du Professeur B. Frachet de l'Hôpital Avicenne à Bobigny et celle du Professeur C. Simon de l'Hôpital central de Nancy, sur les transformations de la parole adaptées aux aides auditives. Ce projet porte sur deux aspects : d'une part, l'élaboration de stimuli artificiels pour établir un meilleur diagnostic, et d'autre part, l'étude des transformations de la parole qui renforcent la perception des indices acoustiques des sons de la parole.

8.2.2 Action coopérative INRIA sur un système générique de reconnaissance de parole

Le but de cette action est mettre en commun les compétences de plusieurs équipes françaises de recherche en parole (LIA, IRISA, ENST, LORIA, IRIT) pour définir et réaliser un système de reconnaissance de parole continue, grand vocabulaire.

8.3 Actions européennes

8.3.1 Projet Européen COST 249 et COST 250

Les projets COST 249 et COST 250 sont deux projets européens qui ont débuté en 1990 et qui s'intéressent au traitement et à la reconnaissance de la parole téléphonique. Ces projets sont des groupes de réflexion et d'échange concernant l'implémentation de systèmes de reconnaissance de la parole téléphonique.

8.3.2 AUPELF-UREF

Dans le cadre de l'appel d'offre AUPELF-UREF (Association des Universités Partiellement ou Entièrement de Langue Française, Universités des Réseaux d'Expression Française), débuté en 1995, nous avons participé au projet *B1* destiné à évaluer notre système de dictée vocale intégrant nos HMM2. Lors de la première campagne, notre machine à dicter MAUD a été classée seconde dans la catégorie des systèmes de 20000 mots. Actuellement, nous travaillons à la définition de la seconde campagne de tests.

Un autre aspect de ce projet porte sur l'évaluation des modèles de langage, et plus précisément, sur la capacité de prédiction de celui-ci. Chaque modèle concurrent doit reconstituer les phrases pour lesquelles certains mots ont été supprimés. Le nombre normalisé de phrases reconstituées correctement donne une valeur approchée de la perplexité. Un autre volet de cette évaluation consiste à connaître le degré de sélection du modèle de langage dans un graphe

d'homophones. Trois laboratoires participent à cette évaluation : LIMSI, LIA et LORIA. Cette campagne nous a permis de travailler ensemble et d'échanger des ressources, des idées et de proposer des publications communes.

8.3.3 VODIS

Projet Telematics Language Engineering VODIS n°1-2277 avec Renault, Peugeot, Volkswagen, BMW, Lernout et Hauspie, Bosch

Le VODIS (*Voice Operated Driver Information Systems*) a pour objectif le pilotage par la voix d'un autoradio intégrant un téléphone portable et un système de navigation. En 1998, nous avons terminé l'enregistrement, l'étiquetage et la diffusion d'une base de données de 200 locuteurs. De plus, cet étiquetage a été certifié à l'aide d'un programme de reconnaissance.

8.3.4 ISAEUS

Projet Telematics/TIDE ISAEUS n° 3004 avec les universités de Bonn et Madrid, Thomson et Teima – est le nom du dernier projet européen dans lequel nous sommes impliqués depuis janvier 1997. Il se termine en juillet 2000. Ce projet porte sur la rééducation vocale des personnes déficientes auditives, enfants ou adultes [17]. Notre contribution porte sur l'autocontrôle et la correction de la voix et prolonge donc les travaux entrepris dans le cadre de SIRENE II. Outre l'extension de l'existant, notre objectif est de renforcer les aspects liés à la pédagogie pour les sujets à entraîner [19, 18] et les aspects liés à l'usage professionnel par les thérapeutes [20]. Un aspect important du projet concerne l'amélioration du système d'entraînement par une mise en commun des expériences sur une base multilingue faisant intervenir l'allemand, l'espagnol et le français.

8.4 Visites, et invitations de chercheurs

Nous avons accueilli les chercheurs dont les noms suivent :
 Parham Mokhtari, université de Canberra, à partir de juin 98.
 Mohamed Debyeche, université d'Alger du 15 octobre au 15 novembre 1999.

9 Diffusion de résultats

9.1 Animation de la Communauté scientifique

Relectures pour les journaux IEEE Transactions on speech and audio processing, Speech communication, Langue, Journal of Phonetics.

Journées de formation sur la reconnaissance de parole à l'intention d'un groupe d'industriels, Nancy, septembre 99.

Présentation d'une démonstration à la semaine de la science, Nancy, octobre 99.

9.2 Enseignement universitaire

- Forte participation à divers enseignements dans les établissements lorrains (Université de Nancy 1 et II, INPL) : Maîtrise et DEA d’Informatique, IUT, MIAGE, DESS Informatique, DESS Information Scientifique et Technique, DEA de Chimie Informatique et Théorique ;
- Responsabilité du DEA d’Informatique de Nancy (J.-P. Haton) ;
- Responsabilité de l’antenne d’Épinal de la faculté des Sciences (M.-C. Haton) ;
- Co-responsabilité de l’option «Ingénierie des Réseaux et des Systèmes du DESS Informatique» (O. Mella) ;
- Responsable pédagogique 2ème année d’IUP-MIAGE (Kamel Smaïli) ;

9.3 Participation à des colloques, séminaires, invitations

- Séjour post-doctoral de Christophe Cerisara à Speech Technology Laboratory (bourse postdoctorale INRIA)
- Participation à des jurys de thèses de doctorat D. Fohr, J.-P. Haton, M.-C. Haton, Y. Laprie, K. Smaïli ;
- Participation à des jurys de mémoire d’orthophonie A. Bonneau, M.-C. Haton, Y. Laprie
- On se reportera à la bibliographie pour la liste des conférences et *workshops* auxquels les membres de l’action ont participé.

10 Bibliographie

Ouvrages et articles de référence de l’équipe

- [1] A. BONNEAU, F. CHARPILLET, S. COSTE-MARQUIS, J.-P. HATON, Y. LAPRIE, P. MARQUIS, « Towards a Multilevel Model for Hypothetical Reasoning in Continuous Speech Recognition », *in : Levels in Speech Communication : Relations and Interactions*, C. Sorin, J. Mariani, et H. Méloni (éditeurs), Elsevier, 1994.
- [2] A. BONNEAU, L. DJEZZAR, Y. LAPRIE, « Perception of the Place of Articulation of French Stop Bursts », *Journal of the Acoustical Society of America* 100, 1, 1996, p. 555–564.
- [3] J. DI MARTINO, « Dynamic time warping algorithms for isolated and connected word recognition », *in : New Systems and Architectures for Automatic Speech Recognition and Synthesis, Nato Asi Series, vol F16*, R. de Mori et C. Y. Suen (éditeurs), Springer-Verlag, Berlin, 1984.
- [4] D. FOHR, J.-P. HATON, Y. LAPRIE, « Knowledge-based Techniques in Acoustic-Phonetic Decoding of Speech: Interest and Limitations », *International Journal of Pattern Recognition and Artificial Intelligence* 8, 1, 1994, p. 133–153.
- [5] M.-C. HATON, « Issues in Using Models for Self Evaluation and Correction of Speech », *in : Computational Models of Speech Pattern Processing*, M. Ponting, K. (éditeur), *Computer and Systems Sciences*, Springer-Verlag, Berlin, 1998.
- [6] I. ILLINA, M. AFIFY, Y. GONG, « Environment Normalization Training and Environment Adaptation Using Mixture Stochastic Trajectory Model », *Speech Communication* 24, 1998.
- [7] J.-C. JUNQUA, J.-P. HATON, *Robustness in Automatic Speech Recognition*, Kluwer Academic, 1996.

- [8] Y. LAPRIE, M.-O. BERGER, « Cooperation of Regularization and Speech Heuristics to Control Automatic Formant Tracking », *Speech Communication* 19, 4, octobre 1996, p. 23, <http://www.loria.fr/publications/1996/96-R-167/96-R-167.ps>.
- [9] O. MELLA, D. FOHR, « TwoTools for Semi-automatic Phonetic Labelling of Large Corpora », in : *First International Conference on Language Resources and Evaluation, Grenade, Espagne*, mai 1998, <http://www.loria.fr/publications/1998/98-R-028/98-R-028.ps>.
- [10] K. SMAÏLI, I. ZITOUNI, F. CHARPILLET, J. P. HATON, « An Hybrid Language Model For a Continuous Dictation Prototype », in : *5th European Conference On Speech Communication And Technology - Eurospeech'97, Rhodes, Greece*, p. 2723–2726, septembre 1997.

Articles et chapitres de livre

- [11] M. AFIFY, Y. GONG, J.-P. HATON, « A Minimum Cross-Entropy Approach to Hidden Markov Model Adaptation », *IEEE Signal Processing Letters* 6, 6, juin 1999, p. 132–134.

Communications à des congrès, colloques, etc.

- [12] A. BONNEAU, Y. LAPRIE, J. VAISSIERE, « Hypertext atlas of speech sounds », in : *Method and Tool Innovations for Speech Science Education, Workshop of the European Speech Communication Association, Londres*, p. 65–68, avril 1999.
- [13] A. BONNEAU, P. MOKHTARI, « A phonetically-guided diagnosis of auditory deficiency based on synthetic speech stimuli », in : *European Conference on Speech Communication and Technology, Budapest, 1*, Technical University of Budapest and The Scientific Society for Telecommunications, p. 559–562, septembre 1999, <http://www.loria.fr/publications/1999/99-R-132/99-R-132.ps>.
- [14] C. CERISARA, D. FOHR, J.-P. HATON, « Robust behavior of multi-band paradigm », in : *Workshop on Robust Methods for Speech Recognition in Adverse Conditions, Tampere, Finland*, mai 1999.
- [15] C. CERISARA, J.-P. HATON, D. FOHR, « Towards a global optimization scheme for multi-band speech recognition », in : *EUROSPEECH 99, Budapest, Hongrie, 2*, p. 587–590, septembre 1999.
- [16] J. DI MARTINO, Y. LAPRIE, « An Efficient F0 Determination Algorithm Based on the Implicit Calculation of the Autocorrelation of the Temporal Excitation Signal », in : *Eurospeech'99, Budapest, Hongrie*, septembre 1999, <http://www.loria.fr/publications/1999/99-R-104/99-R-104.ps>.
- [17] R. GARCIA GOMEZ, M.-C. HATON, J.-P. HATON, C. ANTOINE, P. ALINAT, « Speech training for deaf and hearing-impaired people », in : *European Speech Communication Association, Budapest, Hongrie*, septembre 1999.
- [18] V. GOVAERE, « A Combination of Representation Styles for the Acquirement of Speech Abilities », in : *Artificial Intelligence in Education*, I. Press (éditeur), S.P. Lajoie and M.Vivet, p. 371–378, 1999.
- [19] V. GOVAERE, « Représentation des connaissances à l'aide d'une qualification floue et de frames dans un environnement manipulant des normes », in : *IIIeme Colloque Jeunes Chercheurs en Sciences Cognitives, Soulac*, p. 110–115, avril 1999.
- [20] M.-C. HATON, J.-P. HATON, « Use of articulatory and spectral information for speech training », in : *The XIVth International Congress of Phonetic Sciences, San Francisco*, août 1999.
- [21] D. LANGLOIS, K. SMAÏLI, « A New Based Distance Language Model for a Dictation Machine: application to MAUD », in : *EUROSPEECH 99, Budapest, Hongrie, 4*, p. 1779–1782, septembre 1999.

-
- [22] Y. LAPRIE, « Snorri, a software for speech sciences », *in: Esca workshop, Method and tool innovations for speech science education*, Matisse, Londres, p. 89–92, avril 1999, <http://www.loria.fr/publications/1999/99-R-157/99-R-157.ps>.
- [23] S. OUNI, Y. LAPRIE, « Design of hypercube codebooks for the acoustic-to-articulatory inversion respecting the non-linearities of the articulatory-to-acoustic mapping », *in: 6TH EUROPEAN CONFERENCE ON SPEECH COMMUNICATION AND TECHNOLOGY - EUROSPEECH'99, Budapest, Hongrie, 1*, p. 141–144, septembre 1999.
- [24] K. SMAÏLI, A. BRUN, I. ZITOUNI, J.-P. HATON, « Automatic and manual clustering for large vocabulary speech recognition : a comparative study », *in: Eurospeech'99, Budapest, Hongrie*, septembre 1999.
- [25] K. SMAÏLI, I. ZITOUNI, J.-P. HATON, « Towards a Better Collaboration Between a n-class and a n-gram Language Model », *in: Proceedings of the International Workshop on speech and communication, Moscou*, octobre 1999.
- [26] I. ZITOUNI, J. F. MARI, K. SMAÏLI, J.-P. HATON, « Variable-Length Sequence Language Model for Large Vocabulary Continuous Dictation Machine », *in: EUROSPEECH99, Budapest, Hongrie*, septembre 1999, <http://www.loria.fr/publications/1999/99-R-102/99-R-102.ps>.