

Projet PAROLE

Analyse, Perception et Reconnaissance de la parole

Nancy

THÈME 3A



*R*apport
d'Activité

2000

Table des matières

1	Composition de l'équipe	3
2	Présentation et objectifs généraux	5
3	Fondements scientifiques	5
3.1	Analyse de la parole	7
3.1.1	Perception	7
3.1.2	Indices acoustiques	7
3.1.3	Aides auditives	8
3.1.4	Inversion articulatoire	8
3.2	Reconnaissance automatique de la parole	9
3.2.1	Modèles acoustiques	10
3.2.2	Modèles de langage	10
4	Domaines d'applications	11
5	Logiciels	11
5.1	Outils logiciels	11
5.1.1	Snorri	11
5.1.2	Étiquetage de corpus écrits pour la reconnaissance	12
5.1.3	Classifieur automatique de lexique	12
5.1.4	SALT	12
5.1.5	LIPS	12
5.1.6	VINICS	12
5.1.7	ESPERE	13
5.2	Corpus	13
6	Résultats nouveaux	14
6.1	Analyse de la parole	14
6.1.1	Indices acoustiques	14
6.1.2	Compréhension orale	14
6.1.3	Inversion articulatoire	15
6.1.4	Enseignement des sciences de la parole	16
6.2	Reconnaissance automatique de la parole	16
6.2.1	Modèles de Markov Cachés	16
6.2.2	Réseaux Bayésiens (RB)	18
6.2.3	Modèles de langage	19
6.3	PROCOMA	20
6.4	MIC2	20

7	Actions régionales, nationales et internationales	21
7.1	Actions régionales	21
7.1.1	Action « Assistance à l'apprentissage des langues » (thème Téléopérations et assistants intelligents du Pôle Intelligence Logicielle du Plan État Région)	21
7.2	Actions nationales	21
7.2.1	Projet du pôle européen de santé sur les aides auditives	21
7.2.2	Action coopérative INRIA sur un système générique de reconnaissance de parole	22
7.2.3	Projet RNRT IVOMOB	22
7.2.4	Projet PRIAMM SAALSA	22
7.3	Actions européennes	23
7.3.1	Projet Européen COST 249 et COST 250	23
7.3.2	AUPELF-UREF	23
7.3.3	VODIS	24
7.3.4	ISAEUS	24
7.4	Visites, et invitations de chercheurs	25
8	Diffusion de résultats	25
8.1	Animation de la Communauté scientifique	25
8.2	Enseignement universitaire	25
8.3	Participation à des colloques, séminaires, invitations	25
9	Bibliographie	26

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

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]

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, délégué au CNRS depuis le 1er septembre]

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

Nathalie Parlangau-Vallès [Maître de conférences, I.U.T Charlemagne, U. Nancy 2, depuis le 1er septembre]

Kamel Smaïli [Maître de conférences, U. Nancy 2, délégué au CNRS]

Post-doctorants

Christophe Cerisara [A.T.E.R., U. H. Poincaré, Nancy 1, depuis le 1er septembre]

Imed Zitouni [A.T.E.R., U. H. Poincaré, Nancy 1, jusqu'au 31 aout]

Brigitte Bigi [A.T.E.R., U. H. Poincaré, Nancy 1, depuis le 1er octobre]

Angel de la Torre Vega [université de Grenade, jusqu'au 1er juillet]

Chercheurs doctorants

Yassine Benayed [Bourse tunisienne]

Armelle Brun [MENRT]

Vincent Colotte [MENRT]

Murat Deviren [Bourse INRIA, depuis le 1er octobre]

Salma Jamoussi [MENRT depuis le 1er octobre]

David Langlois [MENRT]

Slim Ouni [Bourse franco-tunisienne]

Ingénieurs sur contrat

Christophe Antoine [projet VODIS jusqu'au 29 février, puis collaborateur extérieur, ingénieur MIC2]

Michel Pitermann [projet PRIAMM, depuis le 15 septembre]

Koray Balci [poste d'accueil INRIA, depuis le 1er octobre]

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]

Virginie Govaere [Poste de jeune docteur à l'IUFM depuis mai]

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

Mariapaola D'imperio-Pitermann [Université de Columbus, Ohio]

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.

L'objectif de notre projet est de traiter automatiquement des signaux de parole pour en comprendre la signification, ou pour analyser et renforcer la structure acoustique. Il s'inscrit dans la perspective de construire des interfaces vocales efficaces et nécessite des travaux en analyse, en perception et en reconnaissance automatique de la parole.

Nos activités se structurent suivant deux thèmes :

Analyse de la parole Nos travaux portent sur l'analyse et la perception des indices acoustiques, l'inversion acoustico-articulatoire et l'analyse de la parole. Ils 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 travaux portent sur les modèles stochastiques (HMM¹, modèles graphiques et trajectoires acoustiques), l'approche multi-bandes, l'adaptation d'un système de reconnaissance à un nouveau locuteur ou au canal de communication et sur les modèles de langage, 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.

Notre culture est pluri-disciplinaire et allie des travaux en phonétique et en reconnaissance des formes. Cette pluri-disciplinarité se révèle être un atout décisif pour aborder de nouveaux thèmes de recherche, l'apprentissage des langues ou les approches multi-bandes notamment, pour lesquels il faut à la fois disposer de compétences en reconnaissance automatique de la parole et en phonétique.

Notre politique de relations industrielles consiste à favoriser les contrats s'insérant assez précisément dans nos objectifs scientifiques. Nous sommes impliqués dans plusieurs coopérations avec des industriels utilisant la reconnaissance automatique de la parole, notamment MIC2 avec qui nous avons une coopération en cours sous la forme d'un projet RNRT, Syncmagic Procoma avec qui nous avons un contrat PRIAMM sur le lipsync et Babel Technologies qui commercialise notre logiciel d'analyse de la parole WinSnoori. Par ailleurs, jusqu'en juillet, nous étions impliqués dans le projet européen ISAEUS sur la rééducation vocale. Nous travaillons également avec des enseignants de langue de Nancy dans le cadre d'un projet du Plan État Région.

3 Fondements scientifiques

Mots clés : traitement du signal, phonétique, télécommunications, santé, perception, modèles stochastiques, modèles de langage, modèles articulatoires, apprentissage des langues,

1. Hidden Markov Models

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, donc incluant 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 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 emprunte à la première les résultats théoriques sur la parole et la première emprunte 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 elles. 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 au plan acoustique (robustesse au bruit, adaptation au locuteur) comme au plan 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 et

donneront lieu à un certain nombre d'applications : 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 multi-bandes. Elles donnent et donneront lieu à un certain nombre d'applications : 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, 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 [15].

Nos expériences 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 [14]. 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 Indices acoustiques

Nous reprenons un travail entrepris sur les indices forts il y a quelques années. Au moment où nous avons introduit le concept d'indice fort, nous désirions pallier une lacune des systèmes de reconnaissance de la parole : l'absence de certitude. En effet, du fait des nombreuses sources de variation qui influencent le signal de parole, les valeurs prises par un indice donné pour deux ou plusieurs unités différentes se recouvrent partiellement. Dans les systèmes de reconnaissance, un coefficient de confiance est attribué en fonction de la valeur de chaque indice considéré et de chaque unité candidate à l'identification. Ainsi l'identification d'un son ou d'un trait repose sur une combinaison de poids complexe. Un tel procédé n'aboutit jamais à une identification certaine. Or certaines configurations acoustiques signalent sans aucune ambiguïté la présence ou l'absence d'un trait; les jugements définitifs parfois émis par les lecteurs de spectrogrammes nous le confirment. Nous avons donc entrepris de décrire ces formes et nous avons défini deux types d'indices acoustico-phonétiques : des indices « forts », de préférence ou d'exclusion, et des indices « faibles ». Les indices forts de préférence autorisent l'identification immédiate d'un trait, les indices forts d'exclusion éliminent directement un candidat à l'identification. Les

indices forts ont donc pour fonction de faire reposer la reconnaissance d'un trait phonétique sur un certain nombre d'informations présentées comme certaines. L'intérêt de tels indices pour l'analyse lexicale est évident : ils permettent d'élaguer le nombre d'hypothèses de mots, toujours très important dans un système de reconnaissance de la parole à moyen ou grand vocabulaire.

Avec l'apparition de nouvelles technologies qui permettent de renforcer les indices importants, l'intérêt des indices forts ne se limite plus à la reconnaissance de la parole mais trouve des applications dans l'apprentissage des langues et les aides auditives. En effet, un indice fort est un indice très discriminant d'un point de vue phonétique et bien marqué d'un point de vue acoustique. Le renforcement de ce type d'indices doit permettre aux apprenants de mieux assimiler les caractéristiques des sons de la langue qu'ils étudient et aux handicapés de mieux percevoir les sons de parole.

3.1.3 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.

3.1.4 Inversion articuloire

Les travaux sur l'inversion acoustique articuloire reposent largement sur une approche d'analyse par synthèse articuloire 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, Khalid Daoudi, Michel Pitermann, Yassine Benayed, Murat Deviren, Angel de la Torre Vega.

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.

Aussi notre groupe travaille-t-il sur l'amélioration de la modélisation de parole par des modèles de Markov cachés (Hidden Markov Models ou HMM) et a-t-il développé deux classes de modèles stochastiques originaux pour la reconnaissance automatique de la parole : les réseaux bayésiens et les modèles stochastiques de trajectoires (Stochastic Trajectory Modeling ou STM).

Les **modèles de Markov cachés** nous ont permis de réaliser des systèmes de reconnaissance automatique de lettres épelées, de chiffres connectés ou de parole continue et de tester différents algorithmes de paramétrisation dans le cas de la parole bruitée ou téléphonique.

Les **réseaux bayésiens** consistent à associer un graphe orienté non-cyclique à la distribution jointe d'un ensemble de variables aléatoires donné. Les nœuds de ce graphe représentent les variables, alors que les liens entre les nœuds codent les indépendances conditionnelles qui existent (ou qui sont supposées exister) dans la distribution jointe. Les HMM sont un cas particulier des réseaux bayésiens. Ces derniers nous offrent donc un cadre théorique général qui nous permet de proposer de nouveaux modèles capables de représenter la parole plus fidèlement que les HMM.

Les **modèles stochastiques de trajectoires (STM)** 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 retrouvée 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. 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 dans deux nouvelles voies de recherche : l'adaptation, l'introduction de la "sémantique statistique" et la compréhension. 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. Cela est rendu possible grâce à des techniques d'identification thématique. La seconde 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'applications 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, serveurs vocaux, 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 sa 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, jusqu'alors 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. Deux versions du logiciel ont été implantées : l'une sous PC-Linux, l'autre sous PC-Windows.

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 que connectés, ou 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, ce qui nécessite 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 suisses (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 5.1.4).

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 [13]), 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

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

6.1.1 Indices acoustiques

Nous travaillons actuellement à l'amélioration des détecteurs d'événements phonétiques et à la détection d'indices forts (censés ne pas déclencher de fausses alarmes) pour les voyelles d'avant. En ce qui concerne les détecteurs d'événements acoustico-phonétiques, nous essayons d'améliorer la détection du bruit des consonnes occlusives ainsi que sa segmentation en deux parties : l'attaque et la friction. Une bonne détection de l'attaque, et en particulier des attaques bien marquées, conditionne la détection des indices forts.

6.1.2 Compréhension orale

Aides 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. 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 [16].

Aides à la compréhension orale Les outils de transformation de la parole que nous avons développés précédemment peuvent être utilisés, soit pour créer de nouveaux systèmes d'aide auditive soit pour améliorer la compréhension orale lors de l'apprentissage des langues. Dans les deux cas le succès de l'approche repose à la fois sur la pertinence des indices acoustiques qui ont été détectés et sur la stratégie exploitant ces indices. Notre travail a donc porté sur le choix d'indices facilement repérables dans le signal et sur les algorithmes permettant de les localiser. Les indices ou régions qui sont ralentis correspondent à des régions où la vitesse spectrale, mesurée par l'algorithme « Spectral Variation Function », est élevée. Il faut noter que les erreurs de détection sont presque exclusivement des insertions à proximité de régions déjà détectées. Elles ne dégradent donc pas la compréhension orale mais allongent seulement l'étendue du signal sur laquelle porte le ralentissement.

À la différence du ralentissement, il est important de ne pas amplifier trop de sons. Par conséquent, nous avons retenu les occlusives et les fricatives sourdes qu'il est possible de détecter en commettant très peu d'erreurs d'insertions. La détection exploite des critères énergétiques et spectraux [25, 24, 22]. Ces transformations ont été évaluées avec succès [23] dans le cadre de l'apprentissage des langues.

Transformations du signal de parole La prosodie couvre les aspects de l'intonation (mesurée par la fréquence fondamentale), de l'énergie et du rythme. Nous avons ajouté cette année la possibilité de manipuler complètement le rythme de la parole. Pour cela nous utilisons les

outils de reconnaissance automatique de la parole qui donnent la segmentation du signal en sons. Il suffit alors de jouer avec PSOLA (Pitch Synchronous Overlap and Add) sur la durée de chacun des sons pour modifier le rythme.

Par ailleurs, nous poursuivons nos travaux sur le vocodeur de phase dont l'avantage est de ne nécessiter aucun calcul du fondamental. L'idée est de séparer la contribution de l'amplitude de celle de la phase et de reconstruire un signal qui correspond à la transformation envisagée. L'une des difficultés est l'apparition d'un phénomène acoustique appelé «phasiness» qui donne l'impression d'une voix enregistrée avec un microphone trop éloigné du locuteur. Ce phénomène s'explique par la destruction de la structure des phases du signal initial, et par conséquent de la forme du signal temporel. Pour éliminer cet effet, nous avons conçu une méthode d'optimisation destinée à assurer que la forme du signal temporel, et par conséquent la structure des phases, est bien conservée. Cette étape d'optimisation peut être vue comme une procédure de synchronisation des phases et elle est déclenchée à chaque début de région voisée.

6.1.3 Inversion articuloire

Notre objectif est d'étudier l'influence de contraintes destinées à approcher la nature physique du conduit vocal sur le mécanisme d'inversion acoustique articuloire. Un objectif préliminaire mais essentiel consiste à développer une procédure d'inversion aussi indépendante que possible des contraintes artificielles, souvent ajoutées pour réduire la complexité du problème mais peu justifiées d'un point de vue phonétique ou articuloire.

Notre approche repose sur l'approximation de la relation articuloire acoustique par une table de couples de paramètres articuloires et paramètres acoustiques. Cette table, appelée «codebook articuloire», est structurée sous la forme d'une hiérarchie d'hypercubes, la relation étant considérée comme linéaire dans les hypercubes les plus fins. Nous avons abordé deux problèmes : l'interpolation des paramètres acoustiques à partir des points du codebook et l'inversion à partir des points du codebook.

L'interpolation est effectuée en calculant le jacobien par rapport au sommet de l'hypercube le plus proche du point à interpoler (un hypercube est en effet défini par 128 sommets puisque le modèle articuloire est un modèle à sept paramètres). En comparant les paramètres acoustiques calculés directement avec le synthétiseur articuloire et ceux calculés par interpolation, nous avons vérifié [32] que le codebook et la procédure d'interpolation permettait d'approcher à moins de 10Hz les valeurs des trois premiers formants.

Le second problème est celui de l'inversion. Pour retrouver la ou les trajectoires articuloires à l'origine du signal de parole nous opérons en deux étapes : d'abord retrouver à chaque instant tous les paramètres articuloires susceptibles d'être à l'origine des paramètres acoustiques observés, ensuite reconstruire une trajectoire articuloire régulière à partir de ces points. La première étape exploite le codebook articuloire et recherche tous les hypercubes donnant des paramètres acoustiques proches de ceux qui ont été mesurés dans la parole. Pour chacun des sommets d'un hypercube possible nous calculons à l'aide de la décomposition en valeurs singulières la solution articuloire la plus proche du sommet en nous assurant que ce point appartient bien à l'hypercube. L'étape de reconstruction d'une trajectoire régulière repose sur un algorithme [31] de lissage non linéaire à base de programmation dynamique.

6.1.4 Enseignement des sciences de la parole

Les outils de traitement du signal qui nous ont servi pour la compréhension orale (l'algorithme de détection de la fréquence fondamentale, l'algorithme de marquage des périodes du fondamental et la mise en œuvre de la méthode PSOLA) peuvent être utilisés dans le cadre de l'apprentissage de la prosodie.

L'opportunité de travailler avec des enseignants de langue (français en tant que langue étrangère et anglais) nous a été offerte dans le cadre du Plan État Région. Depuis le mois d'octobre nous organisons un séminaire commun qui nous a permis de nous fixer comme objectif l'enseignement de la prosodie, sans doute l'un des aspects les moins bien traités par les logiciels éducatifs actuels du domaine.

Nous avons complété nos outils d'analyse de la parole qui offrent maintenant la possibilité de manipuler tous les paramètres de la prosodie. Notre objectif est de corriger et de guider l'apprenant en modifiant sa prosodie, soit en corrigeant ses phrases de manière à les lui faire entendre telles qu'il aurait dû les prononcer, ou au contraire en amplifiant ses erreurs. Pour cela, nous sommes en train de mettre en place une étude de perception destinée à évaluer l'ampleur des modifications pour qu'elles soient perceptibles.

6.2 Reconnaissance automatique de la parole

Mots clés : télécommunications, modèles stochastiques, modèles acoustiques, modèles de langage, reconnaissance automatique de la parole, apprentissage.

Nos travaux sur la reconnaissance automatique de la parole sont classés suivant le type de modélisation stochastique choisi.

6.2.1 Modèles de Markov Cachés

Moteur de reconnaissance de parole générique Afin de concevoir et de tester de nouveaux algorithmes pour la reconnaissance automatique de la parole, nous avons développé le moteur générique ESPERE basé sur les modèles de Markov cachés (HMM) et permettant de définir des modèles, de réaliser leur apprentissage et d'évaluer leurs performances sur de grands corpus de parole continue [28].

Après la conception et la mise au point de ce moteur, il s'est avéré nécessaire d'optimiser le calcul de la vraisemblance lors de la phase de reconnaissance. En effet, les nouveaux modèles acoustiques, comme les modèles triphones que nous avons implantés, se complexifient et nécessitent l'utilisation de plusieurs dizaines de milliers de gaussiennes. Le calcul exhaustif de la vraisemblance n'est plus envisageable et nous avons implanté une technique qui permet d'approcher la valeur optimale à l'aide d'un partitionnement de l'ensemble de gaussiennes.

Modèles et algorithmes Dans des applications réelles, le système de reconnaissance automatique de la parole doit être capable de s'adapter à un nouveau locuteur et de prendre en compte des mots hors vocabulaire.

L'adaptation au locuteur consiste à modifier les moyennes et les écarts types des gaussiennes des modèles en fonction des premières phrases prononcées par le locuteur. Cette année,

nous avons implanté dans le cadre d'un travail de DEA une méthode MLLR (Maximum Likelihood Linear Regression) consistant à calculer une ou plusieurs transformations linéaires pour modifier ces moyennes.

Pour certaines applications vocales, la requête d'un locuteur contient des mots indispensables (mots clés) à la compréhension de la requête et des mots outils ou de politesse non nécessaires (mots hors vocabulaires). Nous avons débuté une thèse sur ce sujet.

Paramétrisation L'essor du téléphone et des applications téléphoniques vocales fait de la reconnaissance de parole téléphonique un problème incontournable pour notre groupe de recherche. Nous avons étudié l'influence de la paramétrisation acoustique et de la topologie des modèles de Markov sur le corpus de parole téléphonique SPEECHDAT et sur une base de données réelles fournie par la société MIC2 [28].

Cette étude nous a montré l'importance de la détection du début et de la fin de la phrase prononcée par le locuteur et nous a conduits à concevoir une nouvelle méthode fondée sur l'énergie du signal, la durée et une modélisation acoustique.

La nouvelle paramétrisation et cette détection parole-non parole ont été intégrées dans notre système ESPERE pour la reconnaissance au téléphone.

Robustesse au bruit 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 avons mené une étude avec Angel de la Torre Vega de l'Université de Granada sur la compensation de bruit. Il s'agit d'évaluer l'influence du bruit dans chacune des bandes de fréquences utilisées pour l'analyse du son (24 bandes critiques) et de compenser cette influence de façon non linéaire [27].

Dans le cadre d'un séjour post-doctoral au Panasonic Speech Technology Laboratory, Christophe Cerisara a étudié et conçu une nouvelle méthode d'adaptation des modèles de Markov à l'environnement. Deux objectifs principaux étaient visés :

- améliorer la robustesse aux bruits additifs et convolutifs d'un système de reconnaissance basé sur les modèles de Markov,
- concevoir un algorithme dont le coût en temps de calcul et en mémoire est suffisamment faible pour permettre une adaptation en temps réel sur un système embarqué.

Cette méthode est basée sur un algorithme d'adaptation proposé par Sagayama^[SYTJ97], qui utilise l'approximation jacobienne de la fonction caractérisant les effets du bruit additif sur le signal de parole. L'algorithme conçu dans le cadre de notre étude améliore l'adaptation Jacobienne selon deux critères : la prise en compte simultanée des bruits additifs et convolutifs et la diminution de la complexité en temps de calcul et en place mémoire. Un article a été publié suite à ce travail [21].

[SYTJ97] S. SAGAYAMA, Y. YAMAGUCHI, S. TAKAHASHI, T. J.-I., « Jacobian Approach to Fast Acoustic Model Adaptation », *in* : *Proc. ICASSP '97, Munich*, p. 835-838, 1997.

Alignement et Post-synchronisation Nous avons adapté une version de notre programme d'étiquetage de corpus à la post-synchronisation de dessins animés dans le cadre d'un contrat avec la société PROCOMA (cf. paragraphe 5.1.5). Nous avons publié les résultats d'une première évaluation de la première maquette opérationnelle [30].

Décomposition/recombinaison multi-bandes La reconnaissance de parole Multi-Bandes 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 recombinaison. Ce modèle a été mis en œuvre récemment par notre équipe dans le cadre de la reconnaissance automatique de la parole, et consiste 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 recombinaison 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 [20].

6.2.2 Réseaux Bayésiens (RB)

Notre stratégie de recherche dans ce domaine consiste à concevoir de nouveaux systèmes de reconnaissance dont la robustesse est liée à la fidélité de la modélisation de la parole plutôt qu'à des améliorations de notre système à base de HMM. Ceci est motivé par le fait que les HMM ne modélisent que la dynamique temporelle du signal de parole alors que sa dynamique fréquentielle, très informative d'un point de vue phonétique, n'est pas prise en compte. Pour avoir une modélisation plus fidèle, il est donc naturel de penser à des modèles qui capturent les caractéristiques à la fois temporelles et fréquentielles de la parole.

Pour ce faire, nous nous inspirons de l'approche multi-bandes, mais au lieu d'utiliser des HMM indépendants pour chacune des bandes spectrales, nous lions les états correspondants aux différentes bandes. De cette façon, nous obtenons un RB plus complexe que les HMM mais qui permet de mieux représenter les aspects temporels et fréquentsiels de la parole.

Nous avons en effet développé un RB multi-bandes qui permet de prendre en compte l'asynchronisme et la dépendance entre les bandes de fréquence. Nous avons ensuite adapté un algorithme d'inférence (junction tree algorithm) à notre RB. Puis, nous avons développé un algorithme d'apprentissage pour ce modèle. Nous avons implémenté ce nouveau système et l'avons testé sur la base de données américaine TIDIGITS. Les tests ont été effectués sur de la parole propre et bruitée alors que l'apprentissage était fait exclusivement sur de la pa-

role propre. En mode propre, notre RB donne de meilleurs résultats qu'un HMM classique: ce dernier obtient 93.4% alors que les RB 2, 3 et 4-bandes donnent respectivement 97.4%, 97.3% et 95.4%. Ceci montre qu'effectivement, notre modèle capture mieux les caractéristiques dynamiques de la parole. En mode bruité, nous avons effectué plusieurs expérimentations en ajoutant, aux données de test, différents bruits colorés à bandes limitées. Nous avons comparé notre RB à un HMM, à un modèle multi-bandes classique et à un RB multi-bandes synchrone. Dans tous les cas de figure, les performances de notre modèle surpassent largement celles des autres modèles: avec un bruit situé en hautes fréquences et avec un rapport signal sur bruit de 26dB, notre modèle donne 84.9% alors que le meilleur des autres modèles n'obtient sur la même tâche que 59.5%. Tous ces résultats montrent d'une part que notre modèle capture mieux les caractéristiques dynamiques du signal et d'autre part qu'il est bien adapté à la reconnaissance de la parole corrompue par un bruit à bande limitée [26].

6.2.3 Modèles de langage

Globalement, les modèles de langage sur lesquels nous travaillons sont des modèles statistiques 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 grammes² de longueur fixe et/ou variable hiérarchique ou non [12][35]. La nouveauté dans ces modèles est l'utilisation d'horizons variables 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 (22%), de précision (19%) et de taux d'erreur (23%). 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 [33].

Les premiers résultats de l'approche fondée sur les modèles distants sont très satisfaisants en terme de perplexité mais en terme de reconnaissance, les résultats ne se sont pas améliorés sensiblement [29]. 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. Un des résultats majeurs de cette année est le développement d'un formalisme permettant un choix dynamique du modèle de langage en fonction de l'historique traité. Ce formalisme a été appliqué pour l'instant aux bigrammes et bigrammes distants. La sélection du modèle de langage adéquat en fonction de l'historique a permis une amélioration de 7% la perplexité. Cette méthode va être généralisée à d'autres types de modèles de langage.

Le développement d'une composante sémantique statistique a débuté par l'élaboration manuelle d'un jeu de classes sémantiques qui servira à étiqueter des corpora. En fait, ces classes correspondent à des concepts inspirés du contenu des résultats du système de dictée VIA VOICE³. L'étude de ces messages a révélé l'importance de l'introduction de la sémantique pour lever certaines ambiguïtés. Nous concentrons nos efforts actuellement sur la formalisation mathématique du modèle correspondant.

2. suites de mots

3. Le système de dictée automatique d'IBM

L'adaptation des modèles de langage constitue l'une de nos préoccupations majeures. Nous abordons le problème par le biais de l'identification thématique et, une fois le thème identifié, le modèle de langage approprié est chargé de l'identification des suites de mots en reconnaissance. Nous avons développé quelques techniques d'identification fondées sur des méthodes statistiques [18, 19]. Une des difficultés a été de construire le dictionnaire caractéristique pour chaque thème. Plusieurs types de vocabulaire ont été testés, certains sont fondés sur la fréquence d'utilisation des mots, d'autres sur les mots caractéristiques positifs et négatifs, etc. Les techniques actuelles d'identification des thèmes nous ont permis d'atteindre le score de 54% d'identification correcte en première position. En sachant que durant l'adaptation, on peut combiner linéairement n modèles de langages, lorsqu'on s'intéresse aux trois meilleurs résultats d'identification, nos algorithmes nous ont permis d'atteindre le taux de 93%. Une application de ces techniques est le routage automatique de mails sur lequel nous travaillons avec l'entreprise MIC2.

Un autre point sur lequel nous travaillons actuellement est l'élimination des événements impossibles des modèles de langage statistiques. En effet, les modèles statistiques estiment V^2 paramètres où V désigne la taille du vocabulaire, qui, dans notre cas, dépasse les 20000 mots. Ce que nous proposons est de supprimer les événements impossibles de l'espace total des événements. La suppression de ces événements est une opération délicate : il faut en fait distinguer les événements impossibles des non observés. Pour ce faire, nous nous sommes fondés sur des classes syntaxiques construites manuellement pour engendrer des règles de génération de biclasses impossibles, desquelles on extrait des bigrammes impossibles. Ces règles sont fondées sur l'information mutuelle, la syntaxe, la statistique et la phonologie. Pour l'instant nous n'avons trouvé que 10% de bigrammes impossibles, mais cela ne constitue qu'un début. La formule de Katz^[Kat87] a été modifiée en conséquence pour tenir compte non seulement des événements peu fréquents mais également de ces événements impossibles qui n'ont jamais été mis en évidence auparavant par notre communauté [17].

6.3 PROCOMA

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

6.4 MIC2

Avec l'entreprise parisienne MIC2, nous débutons un contrat portant sur la définition, la réalisation et l'amélioration d'un moteur de reconnaissance vocale.

Pour sa mise en œuvre le système de reconnaissance sera complété par :

- une procédure d'adaptation au locuteur et/ou à l'environnement par les méthodes MLLR (Maximum Likelihood Linear Regression) ou MAP (Maximum a posteriori Probability),

[Kat87] S. KATZ, « Estimation of Probabilities from Sparse Data for the Language Model Component of a Speech Recognizer », *IEEE Transactions on Acoustics Speech and Signal Processing* 35, 3, 1987, p. 400–401.

- la détection de début et de fin de parole réalisée sous la forme d'un automate avec des transitions qui sont dues soit à une modification du niveau d'énergie, soit à une modification spectrale par rapport à des gabarits de référence du bruit et de la parole,
- un module de paramétrisation de l'analyse MFCC (Mel Frequency Cepstral Coefficients) permettant de fixer le facteur d'échantillonnage, les filtres, la largeur de bande des filtres, le déplacement entre deux fenêtres d'analyse consécutives et le nombre des coefficients,
- une fonction de compensation de la réponse fréquentielle des microphones par soustraction cepstrale,
- un module de réduction du bruit par soustraction spectrale.

Une maquette portant sur une application d'interrogation de cours d'actions cotées en bourses par le téléphone sera réalisée. Par ailleurs, le système sera testé sur des corpus de parole permettant de valider les solutions algorithmiques retenues.

7 Actions régionales, nationales et internationales

7.1 Actions régionales

7.1.1 Action « Assistance à l'apprentissage des langues » (thème Téléopérations et assistants intelligents du Pôle Intelligence Logicielle du Plan État Région)

Depuis le mois d'octobre nous avons commencé à travailler avec des enseignants de français langue étrangère et d'anglais dans l'objectif de mieux exploiter les outils d'analyse et de reconnaissance automatique en apprentissage des langues. Outre un séminaire mensuel (les trois premiers ont porté sur l'utilisation de la phonétique en apprentissage, sur un tour de vue des possibilités des logiciels commercialisés et sur les outils d'analyse de l'équipe) nous définissons actuellement nos objectifs en termes d'apprentissage de la prosodie.

7.2 Actions nationales

7.2.1 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.

7.2.2 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, à grand vocabulaire.

7.2.3 Projet RNRT IVOMOB

Mise au point d'un moteur de reconnaissance automatique de la parole pour l'interface vocale des télé services accessibles depuis un véhicule automobile en mouvement.

Projet RNRT MIC2, Mémodata, Technium et LORIA

Les interfaces des serveurs vocaux interactifs exploitent actuellement la technique DTMF robuste mais qui est d'une trop grande rusticité pour constituer un véritable navigateur. La mise au point d'une véritable interface vocale en langue française pourra permettre de développer de nombreux services accessibles par le public : commerce électronique associant internet et téléphone, recherche vocale de site web, exécution par commande vocale de calcul de performances d'actions et demande de transmission du résultat par e-mail ou télécopie, etc.. Pour maîtriser un tel « navigateur » vocal, il est indispensable d'associer plusieurs disciplines : spécialistes du traitement du signal et de la reconnaissance automatique de la parole, linguistes spécialisés dans le traitement automatique de la langue française, ergonomes spécialisés dans les interfaces vocales, spécialistes du télémarketing, etc...

Ce projet est constitué des 6 sous-projets suivants :

- réalisation de module de pré-traitement des signaux vocaux,
- constitution de corpus acoustico-phonétique et textuel,
- conception d'un système hybride permettant de faire coopérer plusieurs techniques de RAP,
- conception d'une méthodologie de développement des applications vocales,
- réalisation d'une librairie d'automates vocaux,
- réalisation d'une maquette validant l'interopérabilité de l'ensemble des composants.

7.2.4 Projet PRIAMM SAALSA

Dans le cadre du programme PRIAM (Programme pour la Recherche et l'Innovation dans l'Audiovisuel et le Multimédia), l'objectif du projet SAALSA (Système Auto Adaptatif de Lip Sync Automatique) est de développer de nouveaux outils de production pour l'automatisation de la phase de synchronisation de la voix et des mouvements de bouche en animation 2D/3D

en appliquant les résultats des travaux de recherche en traitement automatique de la parole effectués dans les laboratoires LORIA et ENST. Traditionnellement, cette phase de synchronisation est réalisée manuellement par un opérateur expérimenté qui annote le signal sonore en l'écoutant. Cette tâche est longue, rébarbative et coûteuse. Sur la base d'un premier logiciel développé par le LORIA en collaboration avec la société PROCOMA qui automatise une partie de cette phase de synchronisation en utilisant des algorithmes de reconnaissance automatique de la parole, nous proposons de développer un prototype logiciel mieux adapté à la réalité du marché en constante évolution. Ce prototype, nommé SAALSA, achèvera l'automatisation complète du procédé, en s'adaptant systématiquement à chaque nouvel interlocuteur.

Aussi dans ce projet, l'équipe Parole du LORIA se propose-t-elle d'appliquer et de valider les résultats de ses travaux de recherche en parole dans le domaine de l'animation 2D/3D. Plus précisément, il est proposé d'appliquer à la synchronisation voix-mouvement de bouche les algorithmes de reconnaissance automatique de la parole et d'alignement fondés sur les modèles de Markov cachés (HMM, Hidden Markov Models), développé par l'équipe.

L'un des points cruciaux dans les systèmes à base de Markov est l'adéquation entre les modèles appris lors de la phase d'apprentissage et la parole rencontrée lors de la phase d'alignement. Les voix utilisées dans le monde de l'animation sont souvent caricaturales ou extrêmes (voix enfantine, chuchotée, criée, très aiguë ou très grave, imitations d'animaux ou d'accents?) alors que les modèles acoustiques de LIPS ont été appris sur des corpus de parole standard (locuteur d'âge moyen présentant peu d'accent). Ces différences entre les conditions d'apprentissage et les conditions d'alignement génèrent des erreurs et de mauvaises synchronisations. Pour pallier ces difficultés, deux solutions sont envisageables : d'une part, exploiter les résultats d'alignement du logiciel LIPS et, d'autre part, adapter les modèles acoustiques en fonction de la voix du locuteur, par exemple en utilisant des techniques MAP (Maximum A Posteriori) ou MLLR (Maximum Likelihood linear Regression).

7.3 Actions européennes

7.3.1 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.

7.3.2 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.

7.3.3 VODIS

Projet Telematics Language Engineering VODIS n°1–2277 avec Renault, Peugeot, Volkswagen, BMW, Lernout & Hauspie.
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. Ce projet s'est terminé le 29 février.

7.3.4 ISAEUS

Projet Telematics/TIDE ISAEUS n°3004 avec les universités de Bonnet Madrid, Thomson, Teimaet URAPEDA

Le projet ISAEUS s'est terminé le 31 juillet 2000. Les travaux des années antérieures ont permis d'établir un cahier des charges de bonne qualité, couvrant tous les aspects de la rééducation vocale, ce qui est tout à fait novateur pour le français. L'apprentissage de la voix et de la parole s'effectue grâce à un ensemble de leçons organisées en un graphe pertinent dans le cas général selon le point de vue du thérapeute. Des efforts ont été faits dans le sens de l'adaptation à un sujet donné, en tenant compte du bilan orthophonique initial, des aptitudes du sujet révélées au cours des sessions d'apprentissage et de leur évolution. Ils permettent de considérer comme possible l'apprentissage avec guidage automatique ou semi-automatique de la part du système, notamment dans le cas de personnes adultes. Une partie de ces idées sont exposées dans la thèse de Virginie Govaere, soutenue en mai 2000 [11]. Les travaux de l'équipe en matière d'analyse et de reconnaissance de la parole, notamment notre expertise en matière de modèles de Markov cachés, ont permis la mise en place d'un prototype intégrant les leçons d'ISAEUS (depuis les niveaux suprasegmentaux jusqu'au niveau des sons et des mots) et un certain nombre d'utilitaires : un tutoriel sur la phonétique acoustique, des réservoirs de sons et de mots organisés pour des usages multiples, des recommandations et conseils circonstanciés sous forme textuelle ou de séquences video en langage parlé complété (langage dans lequel la production usuelle de parole est accompagnée de signes permettant de lever les ambiguïtés dues aux sosies labiaux). La validation du système a été effectuée en collaboration avec l'URAPEDA et des orthophonistes indépendants, dans un cadre privé ou hospitalier, sur les sept premiers mois de cette année, avec des utilisateurs finaux, adultes et enfants, sourds ou handicapés de la parole. Chaque composante a été étudiée, des propositions d'amélioration ont été faites, des perspectives ont été ouvertes. Ce travail important s'est conclu par la rédaction de documents relatifs à la validation et au plan d'exploitation d'ISAEUS en anglais, et d'un manuel utilisateur en français (deliverables du contrat européen) [38, 37, 36].

7.4 Visites, et invitations de chercheurs

Jan Cernocky, Institut de radio-électronique de Brno, séminaire sur l'inversion acoustico-articulatoire, le 27 mars; Gérard Chollet, ENST Paris, séminaire sur le codage à bas débit par HMM, le 4 décembre; Valérie Hazan, University College of London, séminaire sur le renforcement du signal de parole afin d'améliorer son intelligibilité, les 11 et 12 décembre; Ming Ouhyoung et Michael Li, National Taiwan University, du 14 au 19 décembre, dans le cadre d'un projet sur la réalisation d'une tête parlante.

8 Diffusion de résultats

8.1 Animation de la Communauté scientifique

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

Co-responsabilité du thème Télé-opérations et assistants intelligents dans le cadre du pôle Intelligence logicielle du Plan État Région (Yves Laprie).

Co-responsabilité de l'action « Assistance à l'apprentissage des langues » dans le cadre du thème Télé-opérations et assistants intelligents du Plan État Région (Anne Bonneau).

Membre élu du bureau du G.F.C.P, groupe francophone de la communication parlée, (Yves Laprie).

8.2 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 DESS IST de l'UHP (M. C. Haton) ;
- Responsabilité du DEA d'Informatique de Nancy (J.-P. Haton) ;
- Responsabilité de l'antenne d'Épinal de la faculté des Sciences (M. -C Haton) ;
- Responsabilité du DESS Informatique de l'UHP (O. Mella) ;

8.3 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 ;
- On se reportera à la bibliographie pour la liste des conférences et *workshops* auxquels les membres de l'action ont participé.

9 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.
- [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.

Thèses et habilitations à diriger des recherches

- [11] V. GOVAERE, *Evaluation et guidage d'un utilisateur dans un environnement d'apprentissage. Application au domaine de la rééducation de la parole*, Thèse d'université, Loria, mai 2000.
- [12] I. ZITOUNI, *Modélisation du langage pour les systèmes de reconnaissance de la parole destinés aux grands vocabulaires : application à MAUD*, Thèse d'université, Nancy, mars 2000, <http://www.loria.fr/publications/2000/A00-T-174/A00-T-174.ps>.

Articles et chapitres de livre

- [13] F. BIMBOT, M. EL BEZE, S. IGOUNET, M. JARDINO, K. SMAÏLI, I. ZITOUNI, «An alternative scheme for perplexity estimation and its assessment for the evaluation of language models», *Computer Speech and Language*, 2000, à paraître.
- [14] A. BONNEAU, «Identification of vocalic features from French stop bursts», *Journal of Phonetics*, 2001, à paraître.

Communications à des congrès, colloques, etc.

- [15] A. BONNEAU, Y. LAPRIE, V. COLOTTE, «Towards phonetic tools for speech training», *in: Integrating speech technology in (language) learning (InSTIL2000)*, Dundee, Scotland, août 2000.
- [16] A. BONNEAU, P. MOKHTARI, «Un diagnostic phonétique pour les déficiences auditives», *in: Journées d'étude sur la parole, Aussois - Savoie - France*, Institut de la Communication Parlée (I.C.P.), juin 2000, <http://www.loria.fr/publications/2000/A00-R-225/A00-R-225.ps>.
- [17] A. BRUN, D. LANGLOIS, K. SMAÏLI, J. P. HATON, «Discarding Impossible Events from Statistical Language Models», *in: International Conference on Spoken Language Processing, Pékin*, octobre 2000, <http://www.loria.fr/publications/2000/A00-R-221/A00-R-221.ps>.
- [18] A. BRUN, K. SMAÏLI, J.-P. HATON, «Experiment Analysis in Newspaper Topic Detection», *in: String Processing and Information Retrieval - SPIRE'2000, A Coruna, Espagne*, IEEE Computer Society, p. 55 – 64, septembre 2000.
- [19] A. BRUN, K. SMAÏLI, J.-P. HATON, «Topic Identification Challenge Based on Short Word History», *in: Traitement Automatique du Langage Naturel - TALN'00, Lausanne - Suisse*, p. 383–392, octobre 2000.
- [20] C. CERISARA, D. FOHR, J.-P. HATON, «Asynchrony in Multi-Band Speech Recognition», *in: ICASSP, Istanbul*, juin 2000.
- [21] C. CERISARA, L. RIGAZIO, R. BOMAN, J.-C. JUNQUA, «Transformation of Jacobian matrices for noisy speech recognition», *in: ICSLP'2000, Beijing, Chine*, octobre 2000.
- [22] V. COLOTTE, Y. LAPRIE, «Amélioration automatique de l'intelligibilité de la parole», *in: Journées d'Études de la Parole, Aussois*, Institut de la Communication Parlée, p. 105–108, juin 2000, <http://www.loria.fr/publications/2000/A00-R-142/A00-R-142.ps>.
- [23] V. COLOTTE, Y. LAPRIE, «Automatic enhancement of speech intelligibility», *in: IEEE International Conference on Acoustics, Speech, and Signal Processing - ICASSP'2000, Istanbul*, juin 2000.
- [24] V. COLOTTE, Y. LAPRIE, «Detecting relevant acoustic events for piloting improvement of intelligibility», *in: European Signal Processing Conference, Tampere*, septembre 2000, <http://www.loria.fr/publications/2000/A00-R-143/A00-R-143.ps>.
- [25] V. COLOTTE, Y. LAPRIE, «Modification sélective du débit de parole», *in: Reconnaissance des Formes et Intelligence Artificielle - RFIA '2000, Paris, 2*, p. 141–148, janvier 2000.
- [26] K. DAOUDI, D. FOHR, C. ANTOINE, «A new approach for multi-band speech recognition based on probabilistic graphical models», *in: ICSLP, Beijing, China*, octobre 2000.

- [27] A. DE LA TORRE, D. FOHR, J.-P. HATON, « Compensation of Noise Effects for Robust Speech Recognition in Car Environments », *in: ICSLP, Beijing Chine*, octobre 2000.
- [28] D. FOHR, O. MELLA, C. ANTOINE, « The automatic speech recognition engine ESPERE : experiments on telephone speech », *in: ICSLP, Pékin, Chine*, octobre 2000.
- [29] D. LANGLOIS, K. SMAÏLI, J.-P. HATON, « Dealing with distant relationships in natural language modelling for automatic speech recognition », *in: World Multiconference on Systemics, Cybernetics and Informatics, Orlando (Florida), 6*, International Institute of Informatics and Systemics, p. 400–405, juillet 2000, <http://www.loria.fr/publications/2000/A00-R-144/A00-R-144.ps>.
- [30] O. MELLA, D. FOHR, L. MARTIN, A. CARLEN, « A tool for the synchronization of speech and mouth shapes: LIPS », *in: International Conference on Spoken Language Processing, Pékin, Chine*, octobre 2000.
- [31] S. OUNI, Y. LAPRIE, « Improving acoustic-to-articulatory inversion by using hypercube codebooks », *in: International Conf. on Spoken Language Processing (ICSLP2000), Beijing, Chine, II*, p. 178–181, octobre 2000, <http://www.loria.fr/publications/2000/A00-R-239/A00-R-239.ps>.
- [32] S. OUNI, Y. LAPRIE, « Utilisation d'un dictionnaire hypercubique pour l'inversion acoustico-articulaire », *in: 23èmes Journées d'Etudes sur la Parole, Aussois, France*, p. 409 – 412, juin 2000, <http://www.loria.fr/publications/2000/A00-R-132/A00-R-132.ps>.
- [33] I. ZITOUNI, K. SMAÏLI, J.-P. HATON, « Variable-Length Class Sequences Based on a Hierarchical Approach: MCNV », *in: 4th Word Multiconference on Systemics, Cybernetics and Informatics, Orlando, USA*, juillet 2000, <http://www.loria.fr/publications/2000/A00-R-193/A00-R-193.ps>.
- [34] I. ZITOUNI, K. SMAÏLI, « Vers une meilleure modélisation du langage : la prise en compte des séquences dans les modèles statistiques », *in: XXIIIèmes Journées d'Etude sur la Parole JEP'2000, Aussois, France*, juin 2000, <http://www.loria.fr/publications/2000/A00-R-169/A00-R-169.ps>.
- [35] I. ZITOUNI, K. SMAÏLI, J. P. HATON, « Beyond the Conventional Statistical Language Models: The Variable-Length Sequences Approach », *in: International Conference on Speech Language Processing, Pekin*, octobre 2000.

Rapports de recherche et publications internes

- [36] M.-C. HATON, J.-P. HATON, « Le système ISAEUS, manuel utilisateur », *Rapport intermédiaire*, Commission européenne, juillet 2000.
- [37] M.-C. HATON, « Exploitation plan of the French ISAEUS system », *Rapport intermédiaire*, Commission européenne, juillet 2000.
- [38] M.-C. HATON, « Report about the user experience with ISAEUS », *Rapport intermédiaire*, Communauté européenne, juillet 2000.