INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Projet parole

Analyse, Perception et Reconnaissance de la parole

Lorraine

THÈME 3A

Table des matières

1. Composition de l'équipe

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

Responsable scientifique
Yves Laprie [Chargé de Recherche, CNRS]

Assistante de projet
Martine Kuhlmann [CNRS]

Personnel CNRS
Anne Bonneau [Chargée de Recherche]
Christophe Cerisara [Chargé de Recherche]
Dominique Fohr [Chargé de Recherche]

Personnel INRIA
Khalid Daoudi [Chargé de Recherche]

Personnel Université
Jean-Paul Haton [Professeur, U. H. Poincaré, Institut Universitaire de France]
Marie-Christine Haton [Professeur, U. H. Poincaré]
Irina Illina [Maître de conférences, I.U.T Charlemagne, U. Nancy 2]
Joseph di Martino [Maître de conférences, U. H. Poincaré]
Odile Mella [Maître de conférences, U. H. Poincaré, déléguée au CNRS depuis le 1er septembre 2002]
Nathalie Parlangeau-Vallès [Maître de conférences, I.U.T Charlemagne, U. Nancy 2]
Kamel Smaïli [Professeur, U. Nancy 2]

A.T.E.R
Vincent Colotte [A.T.E.R, U. H. Poincaré, Nancy 1, jusqu'au 30 septembre 2002]
David Langlois [A.T.E.R, U. H. Poincaré, Nancy 1, depuis le 1er octobre 2001]

Chercheurs doctorants
Vincent Barreaud [ATER]
Yassine Benayed [Bourse tunisienne]
Armelle Brun [MENRT]
Murat Deviren [Bourse INRIA]
Salma Jamoussi [MENRT]
Fabrice Lauri [Bourse CIFRE]
Vincent Robert [Professeur du secondaire]

Ingénieurs sur contrat
Christophe Antoine [collaborateur extérieur, ingénieur DIALOCA]
Sen Zhang [ingénieur expert INRIA]
Koray Balci [poste d'accueil INRIA, jusqu'au 30 septembre 2002]
Michel Pitermann [postdoctorant INRIA région, jusqu'au 30 septembre 2002]

Spécialiste INRIA
Filipp Korkmazsky [depuis le 1er novembre 2002]

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 (HMM[1], réseaux bayésiens 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 traduction automatique, l'alignement texte-parole, l'indexation de documents sonores.

Notre culture est pluridisciplinaire et allie des travaux en phonétique et en reconnaissance des formes. Cette pluridisciplinarité 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 DIALOCA 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, Sextant avec qui nous menons une étude sur la reconnaissance de parole avec des locuteurs non natifs dans un environnement bruité et Babel Technologies qui commercialise notre logiciel d'analyse de la parole WinSnoori. Par ailleurs, nous somme impliqués dans les projets européens OZONE et MIAMM. Nous travaillons également avec des enseignants de langue de Nancy dans le cadre d'un projet du Plan État Région.

3. Fondements scientifiques

3.1. Introduction

Mots clés : *traitement du signal*, *phonétique*, *télécommunications*, *santé*, *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, 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

[1] Hidden Markov Models

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, la traduction automatique l'alignement texte-parole et l'indextation de documents sonores.

3.2. 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.

3.2.1. 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 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 [1]. 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.

3.2.2. 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.

3.2.3. 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 [13].

3.2.4. 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 articulatoire 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.

3.3. 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, Khalid Daoudi, Yassine Benayed, Murat Deviren, Angel de la Torre Vega, Fabrice Lauri, Vincent Barreaud, Salma Jamoussi, Nathalie Parlangeau-Vallès, Sen Zhang, Filipp Korkmazsky, Michel Pitermann, Joseph di Martino.

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 langages 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.

3.3.1. 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 parole, d'alignement texte-parole et d'indexation sonore. Nous les utilisons également pour mettre en oeuvre et valider nos travaux de recherche sur les différents algorithmes de paramétrisation, de robustesse et d'adapatation à l'environnement dans le cas de la parole émise dans des conditions très variées (bruitée, spontanée, téléphonique, locuteurs non natifs, etc.).

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.

3.3.2. 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 (modèles avec mémoire cache ou triggers). Même si ces systèmes donnent de bons résultats, ils restent néanmoins limités et ont besoin d'être constamment améliorés pour s'adapter à la complexité de la langue. Afin de maîtriser cette complexité, nous avons continué à travailler selon 3 axes : Adaptation dynamique des modèles de langage, Choix du meilleur modèle en fonction de l'historique, Afin de maîtriser cette complexité du langage, nous avons continué nos travaux de recherche portant sur l'adaptation dynamique des modèles de langage, par le biais d'études concernant l'identification thématique. Le second axe 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. Enfin, nous avons décidé de prospecter dans une nouvelle voie de recherche : la compréhension.

4. Domaines d'application

Les domaines d'application de nos travaux vont de la reconnaissance automatique de la parole aux domaines paramédicaux. 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. Enfin, la multiplication des documents sonores disponibles sur le Web

et non répertoriés par les moteurs de recherche classiques comme Google ou Yahoo, ouvre une nouvelle voie d'application. En effet l'indexation automatique de documents sonores ou audiovisuels permettra qu'ils soient référencés et donc exploitables.

Notre intérêt pour l'apprentissage de la voix et de la parole a permis dans le passé le développement d'un ensemble d'outils éducatifs utilisant l'entrée vocale et les techniques d'analyse et de reconnaissance élaborées dans l'équipe. Nous poursuivons dans cette optique des travaux sur le suivi de l'apprenant en situation d'apprentissage d'un cours diffusé sur le web : constitution automatique de « sous-sites », donc de portions de cours, suivi de la navigation pour conseiller l'élève.

5. Logiciels

5.1. Outils logiciels

5.1.1. PhonoLor

PhonoLor est un phonétiseur permettant de transformer la transcription orthographique d'un mot ou d'une suite de mots en une transcription phonétique. Ce logiciel utilise des règles de phonétisation apprises sur un corpus d'exemples.

5.1.2. Snorri et WinSnoori

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 1999 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).

Cette année nous avons étendu sensiblement les possibilités de WinSnoori dans le domaine de l'édition de signaux de parole. Outre la possibilité de n'afficher qu'une partie du spectrogramme et d'ajouter du bruit à des signaux, on peut maintenant appliquer plusieurs filtres dessinés directement sur le spectrogramme.

Cette nouvelle fonctionnalité permet de rehausser ou d'abaisser le niveau d'énergie des formants et des harmoniques. Elle est destinée à construire des stimuli de perception pour évaluer la contribution des indices acoustiques à l'identification des sons de la parole.

5.1.3. 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 %.

5.1.4. 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.

5.1.5. 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.

5.1.6. 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.

5.1.7. 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.

5.1.8. SALSA

Dans le cadre du projet PRIAMM SAALSA, nous avons un logiciel de post synchronisation de dessins animés en collaboration avec la société Syngmagic et l'ENST.

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.

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. Le logiciel Snorri présenté dans le paragraphe 5.1 en est un exemple.

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. 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 5.1.5).

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, 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

Notre objectif est la définition d'indices forts (très fiables) pour la détection de sons bien prononcés. Nous avons commencé ce travail par les occlusives sourdes du français en considérant à la fois le bruit d'explosion et les transitions formantiques.Après avoir élaboré une procédure de segmentation automatique de l'attaque du bruit, et testé nos indices pour ce segment[33], nous étudions les transitions formantiques. Un premier ensemble d'indices a été défini et nous vérifions actuellement les suivis de transitions avant de le tester. Trouver des indices forts à partir des transitions est une tâche difficile tant sur le plan acoustique que phonétique. En effet, le suivi de formants est une tâche particulièrement ardue sur ces segments à la frontière entre occlusive et voyelle dont la fréquence évolue très vite et qui sont très sensibles au contexte phonétique. C'est pourquoi nous avons choisi de considérer les oppositions de trajectoires formantiques (au lieu des fréquences d'attaque habituellement étudiées).

6.1.2. Compréhension orale

Afin d'améliorer la compréhension orale en renforçant l'intelligibilité de la parole, nous avons développé au cours des deux dernières années des outils de transformation de la parole qui ralentissent sélectivement le débit et amplifient certains indices acoustiques. Pour ne pas introduire d'artefacts acoustiques qui risqueraient de détériorer l'identification des sons, nous avons adopté une stratégie qui consiste à renforcer seulement les consonnes sourdes et les transitions spectrales rapides. Une première expérience, effectuée l'an passé, a montré que nos transformations amélioraient significativement la compréhension des phrases françaises par des étudiants étrangers. Nous préparons cette année une nouvelle expérience destinée à valider notre approche, avec un double objectif. Le premier objectif est de tester l'effet de nos modifications sur des sons isolés, afin d'ajuster nos transformations et éventuellement d'écarter celles qui auraient un effet trop néfaste sur l'identification du son. Pour cela, nous élaborons actuellement un corpus de logatomes VCV. Notre deuxième objectif est de montrer que notre stratégie améliore l'intelligibilité de la parole continue. Pour cela nous avons enregistré une présentation radiophonique du journal météo. Les tests de perception commenceront dès que les corpora seront achevés et les transformations effectuées [18].

6.1.2.1. Transformations du signal de parole

La méthode de transformation du signal de parole que nous avons utilisée pour améliorer la compréhension orale est PSOLA (Pitch Synchronous Overlap and Add) pour sa facilité de mise en œuvre et la qualité obtenue pour le ralentissement temporel. Le manque de précision du placement des marques de synchronisation entraîne cependant l'appratition d'un bruit de synthèse entre les harmoniques de la parole. Nous avons donc apporté deux types d'amélioration.

Tout d'abord nous avons introduit un algorithme d'élagage dans la recherche des marques de synchronisation avec la fréquence fondamentale qui augmente la robustesse du marquage pour les segments de parole présentant de forte variation des formants.

Ensuite, nous avons amélioré la localisation des marques d'analyse et de synthèse. Pour l'analyse il suffit de suréchantillonner le signal ou d'utiliser un algorithme de détection de FO qui donne une précision inférieure à un échantillon. Pour l'étape de synthèse les améliorations ont consisté à rééchantillonner dynamiquement le signal de parole pour pouvoir le placer exactement à l'endroit des marques de synthèse. Ces deux améliorations réduisent fortement la présence de bruit entre les harmoniques ce qui permet d'obtenir des signaux de parole de très bonne qualité [20][19].

6.1.3. Vocodeur de Phase

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. Au cours de cette année, nous avons travaillé sur la possibilité de déclencher la procédure d'optimisation à tout moment au cours de la zone voisée. Nous y sommes parvenu au moyen d'une technique qui consiste à interpoler linéairement les phases dans une zone de jonction autour de l'instant de synchronisation considéré. Nous avons pu constater que grâce à cette technique et pourvu que le zone de jonction ne soit pas trop étroite temporellement les fichiers sons obtenus étaient d'excellentes qualité et presque exempts de « phasiness »[27].

6.1.4. Inversion articulatoire

L'inversion est effectuée en deux étapes : d'abord retrouver à chaque instant tous les paramètres articulatoires susceptibles d'être à l'origine des paramètres acoustiques observés, ensuite reconstruire une trajectoire articulatoire régulière à partir de ces points. Le modèle articulatoire de Maeda décrit la forme du conduit vocal avec sept paramètres et nous utilisons trois paramètres acoustiques qui sont les trois premiers formants.

L'inversion que nous avons développée utilise une table de formes de conduit vocal indexées par les formants. À un triplet de formants correspond théoriquement une infinité de formes du conduit vocal qui, pratiquement, est un espace de dimension quatre. Il serait éventuellement possible de réduire l'indétermination mais il faudrait utiliser des paramètres acoustiques moins robustes et donc plus difficiles à extraire du signal de parole.

Malgré l'introduction de contraintes de continuité pour retrouver des trajectoires articulatoires suffisamment régulières il reste un grand nombre de solutions possibles dont certaines ne sont pas en accord avec les observations phonétiques. La séquence /iy/ doit par exemple s'accompagner de la protrusion des lèvres ce qui n'est pas le cas des meilleures solutions trouvées lors de l'inversion. Il faut noter que malgré cette incohérence phonétique les solutions trouvées donnent, avec une très bonne précision, les formants observés dans le signal d'origine. Nous avons donc introduit des contraintes très simples pour orienter l'inversion vers les trajectoires articulatoires que produirait un locuteur humain. Nous avons ainsi seulement imposé l'observation d'une forte protrusion accompagnée d'une position de la machoire relativement fermée pour /y/. Ces deux contraintes réduisent très fortement le nombre de solutions possibles, et les solutions trouvées sont cette fois tout à fait cohérentes avec les observations articulatoires. Cependant, il apparaît aussi que ces contraintes pourtant assez faibles risquent dans certains cas d'éliminer toutes les solutions inverses. Nous allons donc poursuivre ce travail en étudiant conjointement l'influence de la précision sur les formants et de contraintes articulatoires sur les solutions récupérées lors de l'inversion.

6.1.5. 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 2000 nous organisons un séminaire commun qui nous a permis de 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.

La construction de tutoriaux sur la parole nécessite de pouvoir insérer des outils d'analyse de la parole directement dans des pages Web ou des documents Powerpoint. Nous avons donc développé un ensemble de contrôles ActiveX (propres au système Windows) à partir de notre logiciel d'analyse de la parole WinSnoori. WinSnoori est programmé en C++ qui est le langage le plus simple à utiliser pour construire des contrôles ActiveX. Par ailleurs, la plupart des phonéticiens et enseignants des sciences de la parole préfère le système

Windows. Ces deux raisons expliquent le choix des contrôles ActiveX. Pour l'instant, les objets et contrôles signal, spectrogramme, annotation, et coupe spectrale sont disponibles avec notamment la possibilité de modifier tous les paramètres prosodiques.

Nous avons commencé l'écriture d'un cours interactif sur la prosodie, à l'aide de Snorri ActiveX. Grâce à des transformations adéquates du signal de parole, nous mettons en évidence les caractéristiques acoustiques de l'accent français et de l'accent anglais.

6.2. Reconnaissance automatique de la parole

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

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

6.2.1.1. 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). Il permet de définir des modèles, de réaliser leur apprentissage et d'évaluer leurs performances sur de grands corpus de parole continue. Cette année, nous avons ajouté au moteur la possibilité de calculer les N meilleures solutions pour permettre de tester des méthodes de post-traitement.

La robustesse des systèmes de reconnaissance de la parole vis-à-vis du locuteur et de l'environnement étant primordiale pour l'avenir de la reconnaissance automatique de la parole, nous menons de nombreux travaux de recherche dans ce domaine qui sont décrits dans les deux paragraphes suivants.

6.2.1.2. Adaptation des modèles au locuteur ou à l'environnement

L'adaptation au locuteur consiste à modifier les moyennes et les écarts type des gaussiennes des modèles en fonction des premières phrases prononcées par le locuteur. Nous poursuivons une thèse sur le sujet. Pour cela, nous avons d'abord amélioré la méthode MLLR (Maximum Likelihood Linear Regression) qui consiste à calculer une ou plusieurs transformations linéaires pour modifier les moyennes des modèles HMMs. Puis, afin de mieux prendre en compte la quantité disponible de données d'adaptation, nous avons développé une approche hiérarchique de MLLR (Sructural-MLLR) [34].

Cependant lorsque la quantité de données est vraiment trop faible, la méthode MLLR dégrade les performances de la reconnaissanc. Aussi avons-nous développé la méthode S-MAP (Structural Maximum A Posteriori adaptation)[34].

Dans un contexte plus général d'adaptation à l'environnement, nous avons étudié la méthode S-MAP dans le cadre des modèles segmentaux [29] et poursuivi notre étude sur une méthode d'adaptation MLLR dans laquelle la dépendance entre les classes de régression est modélisée par un processus autorégressif multi-échelles

Nous avons également étudié le problème de la reconnaissance de parole prononcée par des locuteurs non natifs en présence de bruit. Nous avons montré que l'utilisation de modèles phonétiques mixtes, c'est-à-dire l'association de modèles phonétiques de la langue cible avec ceux de la langue maternelle du locuteur, améliorait de façon notable les performances. De plus, l'adaptation MLLR de ces modèles mixtes au locuteur et au bruit améliore encore les performances [28].

6.2.1.3. Robustesse aux variabilités du signal de parole

L'un des facteurs importants dans l'environnement d'un système de reconnaissance de parole est le bruit. Ainsi 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 donc à augmenter la robustesse de nos modèles acoustiques (HMM) aux bruits [8].

Dans ce cadre nous avons débuté une thèse sur la compensation de la parole bruitée. Nous avons proposé une approche originale pour compenser en temps réel les bruits éventuellement non stationnaires sans utiliser de données d'adaptation.

Par ailleurs, nous avons également amélioré l'algorithme d'adaptation Jacobienne développé l'année précédente, en incorporant un module permettant d'estimer dynamiquement certains paramètres de l'algorithme d'adaptation. Le but est double : éliminer la nécessité d'utiliser un corpus de développement, et par conséquent, rendre la méthode plus « stable » car moins dépendante des conditions d'apprentissage et de développement [17].

Nous avons aussi réalisé une étude comparative de différentes méthodes pour adapter rapidement des modèles acoustiques à des bruits convolutifs ou additifs. les méthodes suivantes ont été testées sur la base de données de Vodis (corpus enregistré dans une voiture) : PMC (Parallel Model Combination), CMS (Cepstral Mean Subtraction) et un algorithme qui combine ces deux approches dans le domaine spectral [16].

Nous avons également mené une étude en coopération avec l'Université de Granada sur la comparaison de deux approches pour améliorer la robustesse au bruit ambiant dans une voiture : la compensation du bruit et l'adaptation des modèles acoustiques [22].

Enfin nous avons commencé un travail de fond sur l'adaptation des modèles acoustiques aux bruits non-staionnaires, comme les bruits musicaux souvent présents en arrière-plan dans les documentaires télévisés par exemple. Les méthodes étudiées tentent de décomposer le signal perçu en ses composants issus des différentes sources sonores, comme par exemple le locuteur principal et la musique. En ce sens, notre travail présente des liens avec le domaine de l'analyse de scènes auditives. Nous avons ainsi étudié différentes approches pour traiter ce problème : (i) la reconnaissance en données manquantes qui « masque » les parties de l'espace temps-fréquence qui ne sont pas issues du locuteur principal et (ii) une modification de l'algorithme PMC pour traiter ces problèmes particuliers. Nous avons également proposé un formalisme commun expliquant ces deux approches. Les problèmes restant à résoudre concernent essentiellement le développement d'algorithmes d'estimation dynamique du bruit non-stationnaire et non-prédictible ainsi que la validition de ces approches sur des systèmes de reconnaissance grand-vocabulaire.

6.2.1.4. Détection de mots clés

La détection de mots clés consiste à repérer des mots ou des expressions dans le signal de parole sans s'appuyer sur une transcription complète de ce signal. L'approche choisie repose sur une reconnaissance flexible à base de modèles de Markov cachés et sur le concept de Modèle du Monde. Dans ce cadre, nous avons poursuivi l'étude sur la détection de mots clés en nous focalisant sur le rejet des fausses alarmes fondé sur les SVM (Support Vector Machine)[11][9][10][12]. Mais nous avons aussi initié une nouvelle étude sur la détection de mots clés. Dans celle-ci, nous avons choisi de représenter le modèle du monde par une boucle de modèles HMM phonétiques. Le modèle de chacun des mots clés est construit à partir de l'ensemble des modèles phonétiques qui composent le modèle du monde. Lors de la phase de reconnaissance des mots-clés, le modèle du monde est reconnu plus facilement de par sa structure. Il faut donc mettre en oeuvre un mécanisme de pondération pour favoriser la détection des mots clés. Cette pondération est une fonction du nombre de phonèmes constituant le mot. Nous avons commencé à tester cette approche sur un bulletin d'information radiophonique dans le cadre du projet Raives (cf. paragraphe 8.2.3). Les modèles phonétiques indépendants du contexte ont été appris sur Bref80 (parole lue). Une double adpatation de ces modèles aux données radiophoniques a été faite par la méthode S-MLLR [34] en deux étapes : une première, supervisée, sur une émission radiophonique différente et une seconde, non supervisée, sur le bulletin lui-même lors de la détection. L'évaluation du système se fait en fonction du nombre d'omissions et d'insertions de mots clés (courbes ROC). Les premiers résultats obtenus sont encourageants.

6.2.1.5. Segmentation parole/musique

Lorsqu'on souhaite transcrire ou indexer des données radiophoniques ou audiovisuelles, il est nécessaire dans un premier temps de découper le signal en segments de type Parole et/ou Musique. L'approche choisie est basée sur des modèles de mélanges de lois gaussiennes (GMM). Chaque classe (parole, musique) est caractérisée par un GMM distinct. D'autres approches sont en cours d'expérimentation.

6.2.1.6. Paramétrisation Acoustique

Nous avons conçu un système de reconnaissance de lettres isolées prononcées en américain, indépendant du locuteur. Cette tâche est particulièrement intéressante du fait de la difficulté de reconnaissance du vocabulaire

sur lequel nous avons travaillé. Nous avons utilisé les modèles de Markov cachés pour concevoir notre moteur de reconnaissance. L'originalité de notre étude réside dans le fait que nous avons introduit des dérivées d'ordre élevé (de 1 à 5) dans les vecteurs acoustiques. Nous nous sommes rendu compte qu'à chaque fois que nous ajoutions une dérivée, le taux de reconnaissance augmentait. Avec 5 dérivées, nous avons obtenu, sur la base de données internationale ISOLET notre meilleur taux de reconnaissance, à savoir 97.54%. Ce résultat est comparable à ce qui a été d'ores et déjà publié, avec l'avantage pour notre méthodologie qu'elle est de beaucoup plus simple [26].

6.2.2. Réseaux Bayésiens Dynamiques

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 sommes inspirés de l'approche multi-bandes, mais au lieu d'utiliser un HMM indépendant pour chacune des bandes spectrales, nous couplons ces HMMs en ajoutant des dépendances (ou liens en terme de graphes) entre les processus cachés gouvernant les dynamiques des différentes bandes. De cette façon, nous obtenons un réseau bayésien dynamique (RBD), certes plus complexe qu'un HMM, mais qui permet de prendre en compte l'asynchronisme et la dépendance entre les bandes de fréquence. Nous avons effectué plusieurs expérimentations en ajoutant, aux données de test, différents bruits colorés à bandes limitées. Nous avons comparé notre RBD à un HMM, à un modèle multi-bandes classique et à un RBD multi-bandes synchrone. Dans tous les cas de figure, les performances de notre modèle surpassent largement celles des autres modèles. 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 pour la reconnaissance de la parole corrompue par un bruit à bande limitée [7].

Un des problèmes à résoudre lors de la conception d'un RBD est le choix de la structure graphique appropriée. Nous avons ainsi développé une technique de conception de RBDs pour la modélisation de la parole qui ne fait pas d'hypothèses de dépendance *a priori* entre les variables observées et cachées, mais plutôt qui apprend les dépendances à partir des données d'apprentissage (*l'apprentissage structurel*). Cette technique garantit de meilleures performances de reconnaissance que les HMM tout en permettant un contrôle à l'utilisateur sur la complexité souhaitée du moteur de reconnaissance. En outre, elle ajoute un degré de discrimination supplémentaire dans la modélisation. Cette année nous avons étendu cette technique afin de pouvoir traiter de la parole continue [25][23]. Nous avons ensuite développé un algorithme rapide de décodage qui permet d'inférer un séquence de RBDs avec des structures graphiques différentes [24].

Un article de synthèse de nos travaux sur cette nouvelle stratégie pour la conception de systèmes de reconnaissance de la parole [21] a été primé par un *"Best Paper Award"* de l'*International Society of Applied Intelligence* lors de la "Fifteenth International Conference on Industrial and Engineering Application of Artificial Intelligence and Expert Systems (IEA/AIE-2002)".

6.2.3. Modèles de langage

L'adaptation des modèles de langage constitue l'une de nos préoccupations majeures. Nous abordons le problème par le biais de l'identification thématique qui a pour objet d'assigner un label thématique à un segment de texte parmi un ensemble prédéfini de labels possibles. Dans le cadre de l'application à la reconnaissance automatique de la parole, le thème est identifié avec les premiers mots dictés puis le modèle de langage approprié intervient dynamiquement dans le processus de reconnaissance. Cette année, nous avons développé différentes méthodes d'identification thématique et avons effectué une étude sur leur complémentarité. Les méthodes de base utilisées sont l'unigramme, le cache, la TFIDF, la perplexité et une méthode originale fondée sur la similarité du mot et du thème [15][14]. Un des résultats important de ce travail sur l'identification de thèmes a consisté à mettre en avant l'importance des vocabulaires thématiques.

Les méthodes de combinaison que nous avons développées sont fondées sur la SVM, les réseaux de neuronne et le vote majoritaire. La combinaison des méthodes de base nous a permis d'accroître les performances de l'identification (93%). Nous avons voulu savoir s'il était possible d'aller au delà. Pour ce faire, nous avons demandé à des expérimentateurs humains d'identifier des thèmes à partir de nos échantillons de texte. L'expérience à montrer que dans plusieurs cas les étiqueteurs humains n'étaient pas d'accord sur la classe à donner à un même document. Ce résultat nous a permis de conclure qu'il est difficile, étant donné la qualité des documents, d'obtenir de meilleurs résultats.

Un autre volet de notre travail concerne les relations syntaxiques ou sémantiques entre les composantes d'une phrase ou d'un texte qui sont en grande partie distantes. Cette étude a été systématisée et finalisée cette année. Elle a abouti au développement de méthodes de combinaisons dédiées à ces modèles et fondées sur les parties de l'historique utilisées par chacun des modèles à combiner. L'application du principe de sélection décrit dans le rapport précédent à la détection de séquences de mots a permis de trouver des séquences d'une validité linguistique intéressante. Ces séquences ont permis d'améliorer la perplexité de nos modèles de plus de 20% et de réduire le taux d'erreur d'un système de reconnaissance de plus de 12% [31][32].

Depuis deux ans maintenant, nous introduisons dans les modèdes statistiques de langage une connaissance supplémentaire : la notion d'événements impossibles de la langue, non considérée dans les modèles classiques. Leur prise en compte pourrait permettre de modéliser de manière plus précise les contraintes langagières. La première phase de ce travail consiste à recenser un ensemble d'événements impossibles. Nous avons complété ce recensement cette année en utilisant diverses sources de connaissance supplémentaires. De plus, les travaux de l'année passée ont montré la nécessité d'une méthode de report efficace de la masse de probabilités initialement allouée à ces événements impossibles, vers les événements possibles. Nous avons déterminé que cette méthode doit être adaptée aux données de test afin que la modification globale des distributions de probabilités ait une incidence sur les performances du modèle final.

Enfin, nous continuons à travailler sur la formalisation du problème de la compréhension de la parole. Nous considérons ce processus comme un processus de traduction d'un signal vers un autre signal. La premère partie de ce processus consiste à traduire le texte intial en concepts. Nous avons testé plusieurs méthodes : réseaux de neuronnes, méthodes fondées sur l'information mutuelle [30] et réseaux Bayésiens. Les résultats obtenus avec cette dernière méthode sont très satisfaisants. Les concepts obtenus sont d'une très grande qualité. Le travail consiste maintenant à franchir la deuxième étape permettant de passer des concepts à la compréhension effective.

6.3. SEXTANT

Nous avons mené une étude sur la reconnaissance de parole avec des locuteurs non natifs dans un environnement bruité.

6.4. DIALOCA

Nous continuons notre collaboration avec la société DIALOCA et cette année nous avons développé un nouveau moteur de reconnaissance nommé MICLOR qui nous permet de tester de nouveaux algorithmes pour l'adaptation au locuteur ou la robustesse de bruit.

6.5. 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 8.2.5)

8. Actions régionales, nationales et internationales

8.1. Actions régionales

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

Nous réalisons un logiciel d'apprentissage de la prosodie de l'anglais par des élèves français, qui utilise des techniques de visualisation et de transformation du signal, et qui est destiné à être utilisé par des enseignants de langue lors de leur cours. Outre les outils de traitement du signal et de reconnaissance automatique, le logiciel comprend un cours de prosodie, destiné aux enseignants, ainsi qu'une base de données intégrant des phrases modèles. Notre objectif est double : former les enseignants d'abord, puis par leur intermédiaire, sensibiliser les élèves à la prosodie d'une langue par l'écoute, la visualisation, et l'exagération des fautes et des cibles à atteindre à partir de leur propre production.

Le cours de prosodie proposé aux enseignants est divisé en trois parties. (1) Une présentation des indices acoustiques correspondant aux unités prosodiques et des outils de traitement du signal. (2) Le cours sur la prosodie proprement dit, accompagné par des exemples de phrases acoustiquement modifiées afin de mettre en relief les unités prosodiques de chaque langue et leurs indices acoustiques. (3) Une illustration des erreurs typiques et des corrections possibles à partir de phrases prononcées par des locuteurs natifs et des élèves français.

Lors de leur cours, les enseignants peuvent utiliser une base de phrases prononcées par des locuteurs anglais. Un système d'alignement automatique (en cours de réalisation) permettra de comparer les réalisations de l'élève avec celles du modèle. L'enseignant peut également utiliser les outils de traitement du signal afin de modifier la réalisation de l'élève en exagérant ses fautes ou en effectuant des corrections afin que la production de l'élève se rapproche de celle de la cible. Le logiciel doit être simple et pratique pour pouvoir être réellement utilisé par les enseignants et les élèves. Pour cela nous avons inclus les principales fonctions d'édition et de transformation du signal du logiciel WinSnorri au sein d'outils Windows comme Word, et Powerpoint, par le biais de fonctions Active X. L'utilisateur peut, à partir du document Word, appeler un signal sonore, visualiser son spectrogramme, la courbe intonative, l'intensité et l'étiquetage du signal en phonèmes ou en mots, si celui-ci est disponible. Il peut également modifier les indices prosodiques (durée, fréquence fondamentale et/ou intensité, simultanément ou non) en chaque point du signal. La réalisation de ce logiciel réunit des spécialistes de parole, de traitement du signal, d'ergonomie, des enseignants de langue et des ergonomes.

8.2. Actions nationales

8.2.1. Action de recherche coopérative INRIA

Retour d'efforts articulatoires

Partenaires : ENST Paris, LPL Aix-en-Provence, Projets PAROLE et ISA. L'objectif à long terme de ce projet est de pouvoir guider articulatoirement un sujet apprenant une langue étrangère. Pour cela nous nous proposons d'utiliser une méthode d'inversion acoustique articulatoire pour retrouver les paramètres qu'a utilisés le locuteur, des techniques de suivi de mouvement pour connaître le mouvement de la mâchoire et des lèvres, des outils de synthèse articulatoire et de tête parlante pour construire le retour audiovisuel. Durant cette année nous avons principalement mis en place la plate-forme de capture d'images et de stéréovision pour reconstruire en trois dimensions le visage des locuteurs. Par ailleurs, nous avons développé un modèle biomécanique permettant d'animer un visage parlant.

8.2.2. Projet MathSTIC

Modèles probabilistes graphiques pour la reconnaissance automatique de la parole

Partenaires : ENST Paris, ENS Cachan, Institut Elie-Cartan et le LORIA.

Ce projet met en collaboration mathématiciens et spécialistes du traitement du signal et de la parole pour étudier de façon approfondie d'une part le formalisme des réseaux Bayésiens dynamiques, d'autre part les

problèmes de robustesse en reconnaissance de la parole. L'objectif étant de développer de nouveaux modèles stochastiques de la parole susceptibles de conduire à la conception de système(s) de reconnaissance vocale robuste(s).

8.2.3. Projet STIC-SHS RAIVES

Les documents sonores font à l'heure actuelle partie de ce que l'on appelle le "Web invisible". Nous avons initié un projet avec les laboratoires DDL (Lyon2) et IRIT (Toulouse III) qui a pour objectif de structurer ces documents sonores, en particulier radiophoniques, à partir d'une indexation par leur contenu, de manière à leur donner un sens du point de vue d'un utilisateur du Web, et, de produire à partir de ces documents des connaissances exploitables [35]. Ce contenu pourra alors être accessible aux moteurs de recherche et devenir disponible aux internautes au même titre que le contenu textuel de pages HTML. Ce projet contribuera donc au développement d'une nouvelle génération de moteurs de recherche, capables d'accéder par leur contenu à des documents sonores, et pourquoi pas visuels. Les méthodes seront mises au point sur des données radiophoniques classiques avant d'être adaptées à des données radiophoniques provenant d'Internet.

L'indexation par le contenu de documents sonores s'appuie sur des techniques utilisées en traitement automatique de la parole, mais doit être distinguée de l'alignement automatique d'un texte sur un flux sonore ou encore de la reconnaissance automatique de la parole. Ce serait alors réduire le contenu d'un document sonore à sa seule composante verbale. Or, la composante non-verbale d'un document sonore est importante et correspond souvent à une structuration particulière du document. Par exemple, dans le cas de documents radiophoniques, on voit l'alternance de parole et de musique, plus particulièrement de jingles, pour annoncer les informations. Ainsi, nous pouvons considérer un ensemble de descripteurs du contenu d'un document radiophonique : segments de Parole/Musique, "sons clés", la langue, locuteur associé à une éventuelle identification de ce locuteur, mots clés et thèmes. Cet ensemble peut être bien entendu enrichi. Extraire l'ensemble des descripteurs est sans doute suffisant pour référencer un document sur Internet. Mais il est intéressant d'aller plus loin et de donner accès à des parties précises du document. Chaque descripteur doit être associé à un marqueur temporel qui donne accès directement à l'information. Cependant, l'ensemble des descripteurs appartenant à des niveaux de description différents, leur organisation n'est pas linéaire dans le temps : un même locuteur peut parler en deux langues sur un même segment de parole ou bien étant donné un segment de parole dans une langue donnée, plusieurs locuteurs peuvent intervenir. Il faut donc aussi être capable de fournir une structuration de l'information sur différents niveaux de représentation.

Notre contribution pour la première année de ce projet a été la participation à la définition et à la constitution du corpus, la définition d'un guide d'étiquetage, l'annotation manuelle de 2 heures de données radiophoniques et la mise au point d'un premier système de détection de mots clés.

8.2.4. 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 DIALOCA, Mémodata, Technium et LORIA

Les interfaces des serveurs vocaux interactifs exploitent actuellement la technique DTMF robuste mais d'une trop grande rusticité pour constituer un véritable navigateur. La mise au point d'une véritable interface vocale en langue française est de nature à 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.

Actuellement, nous nous focalisons sur les problèmes de débruitage.

8.2.5. 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 INRIA-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 INRIA-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 avons proposé de développer un prototype logiciel mieux adapté à la réalité du marché en constante évolution. Cela a consisté en l'amélioration de la phonétisation (transformation du texte en suite de phonèmes), l'apprentissage de nouveaux modèles génériques et l'implantation de méthodes d'adaptation au locuteur (MLLR) pour mieux tenir compte des voix utilisées dans le monde de l'animation qui sont souvent caricaturales ou extrêmes [36] (voix enfantine, chuchotée, criée, très aiguë ou très grave, imitations d'animaux ou d'accents).

8.3. Actions européennes

8.3.1. OZONE

Projet IST, OZONE (New technologies and services for emerging nomadic societies) n°IST-2000-30026 avec Philips, Interuniversitair Micro-Electronica Centrum, Laboratoires d'Electronique Philips, EPICTOID, Eindhoven University of Technology, THOMSON multimedia R&D France

Avec plusieurs autres projets de l'INRIA (dont le projet Langue et Dialogue et l'action MAIA à Nancy) nous participons à ce projet où notre rôle concerne le développement d'une interface multimodale généraliste destinée à utiliser des services nomades. L'objectif général du projet OZONE est de développer un cadre logiciel qui puisse s'adapter très souplement aux différents types de matériels et de situations rencontrés dans le cadre de l'informatique nomade.

L'interface multimodale reposera sur la parole et le geste et devra être capable de modéliser la situation dans laquelle se trouve l'utilisateur et les services nomades auxquels l'utilisateur a accès. Les équipes de l'INRIA seront impliqués dans un démonstrateur embarqué dans un cybercar. Ce démonstrateur sera implanté à Rocquencourt.

Le travail de cette année a essentiellement porté sur l'élaboration de l'architecture du système OZONE, et plus particulièrement sur la couche "Service Enabling" destinée au développement des applications. Cette couche logicielle intègre à la fois des aspects de sensibilité au contexte ("context awareness") et de multimodalité.

8.3.2. MIAMM

MIAMM est un projet transversal regroupant les équipes Langue et Dialogue et Parole. Plusieurs partenaires participent à ce projet : CANON, DFKI, SONY, TNO et le LORIA. On se donne comme objectif de fournir une plateforme de conception d'un système de dialogue multimodal. Cette plateforme est fondée sur une série de scénarios d'interaction utilisant plusieurs modalités. La contribution de notre équipe se situe au niveau de la prise en charge de la modalité « parole » pour le Français. Le système ESPERE développé dans l'équipe

constituera le moteur de reconnaissance du français dans MIAMM. Aussi, le module de détection Parole/non Parole sera fourni aux partenaires dans le but d'être utilisé dans le système de reconnaissance de l'allemand.

Pour le système de reconnaissance du français, nous avons calculé un modèle de langage de type bigramme à partir d'un corpus développé au Loria. Ce corpus de 71000 requêtes concerne un bureau d'ordinateur et est constitué d'un vocabulaire d'une centaine de mots. Un corpus de test a été créé également afin de valider le système de reconaissance de cette application. Plusieurs locuteurs (Hommes et femmes) ont prononcé chacun une vingtaine de phrases. Nous avons obtenu les premiers résulats et nous nous concentrons actuellement sur la mise au point des paramètres acoustico-langagiers afin d'améliorer les performances.

8.4. Visites, et invitations de chercheurs

- Jacqueline Vaissière, professeur à la Sorbonne Nouvelle, séminaire intitulé « La prosodie de l'anglais »,
- Elizabeth Guimbretière, professeur à l'université Jussieu, intitulé « L'accent en français »,
- G. Konopczynski, professeur à l'université de Franche-Comté séminaire intitulé « L'apprentisage du langage »,
- A. Germain, directrice du centre de pédagogie universitaire d'Ottawa (Canada), séminaire intitulé « Les transformations de la parole et l'enseignement de la prosodie »,
- Kumiko Tanaka-Ishii, université de Tokyo, séminaire intitulé « Some language technologies for small sized corpora »,
- Laurent Younes, ENS Cachan, séminaire intitulé « Apprentissage statistique »,
- Sylvain MARCHAND, LABRI, Université de Bordeaux, séminaire intitulé « Analyse, transformation et synthèse du son dans les modèles sinusoïdaux »,
- Driss Aboutajdine, professeur à la faculté des sciences de Rabat, séminaire intitulé « Présentation du pôle de compétence STIC du Maroc »

9. Diffusion des résultats

9.1. Animation de la Communauté scientifique

Relectures pour les journaux IEEE Transactions on speech and audio processing, IEEE Transaction in Information Theory, Speech communication, Journal of Phonetics, JASA.

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).

Membre de IASTED Technical Committee on Pattern Recognition (K. Daoudi).

Membre du comité de programmes du European Symposium of Young Researchers in Artificial Intelligence (K. Daoudi).

Présidence de l'Association Française des Sciences et Technologies de l'Information ASTI (Jean-Paul Haton).

Membre du comité de programmes de l'International Conference on Speech and Language Processing ICSLP (Jean-Paul Haton).

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 DESS IST de l'UHP (M. C. Haton) ;
- Responsabilité du DESS Informatique de l'UHP (O. Mella) ;

9.3. Participation à des colloques, séminaires, invitations

- Poster de Khalid Daoudi sur les réseaux bayésiens en reconnaissance de la parole lors de l'"International Meeting on Bayesian Statistics (Valencia 7)" à Tenerife, Espagne.
- Murat Deviren a été sélectionné par AAAI (American Association or Artificial Intelligence) comme présentateur au "Seventh SIGART/AAAI Doctoral Consortium at AAAI-02" à Edmonton, Canada. Une bourse de la NSF lui a ainsi été attribuée.
- 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.

10. Bibliographie

Bibliographie de référence

[1] A. BONNEAU. *Identification of vocalic features from French stop bursts.* in « Journal of Phonetics », 2001.

[2] C. CERISARA, D. FOHR. *Multi-band automatic speech recognition.* in « Computer Speech and Language », numéro 2, volume 15, avril, 2001, pages 151-174.

[3] M.-C. HATON. *Issues in Using Models for Self Evaluation and Correction of Speech.* éditeurs M. PONTING., in « Computational Models of Speech Pattern Processing », série Computer and Systems Sciences, Springer-Verlag, Berlin, 1998.

[4] I. ILLINA, M. AFIFY, Y. GONG. *Environment Normalization Training and Environment Adaptation Using Mixture Stochastic Trajectory Model.* in « Speech Communication », volume 24, 1998.

[5] J.-C. JUNQUA, J.-P. HATON. *Robustness in Automatic Speech Recognition.* Kluwer Academic, 1996.

[6] Y. LAPRIE, M.-O. BERGER. *Cooperation of Regularization and Speech Heuristics to Control Automatic Formant Tracking.* in « Speech Communication », numéro 4, volume 19, octobre, 1996, pages 23.

Articles et chapitres de livre

[7] K. DAOUDI, D. FOHR, C. ANTOINE. *Dynamic Bayesian Networks for Multi-Band Automatic Speech Recognition.* in « Computer Speech and Language », 2002, à paraître.

[8] J.-P. HATON. *Méthodes robustes pour la reconnaissance automatique de la parole.* in « La parole, des modèles cognitifs aux machines communicantes », J. Mariani, mars, 2002.

Communications à des congrès, colloques, etc.

[9] Y. BENAYED, D. FOHR, J.-P. HATON, G. CHOLLET. *Keyword Spotting using Support Vector Machines.* in « Fifth International Conference on Text, Speech and Dialogue - TSD'2002 , Brno, Czech Republic », septembre, 2002.

[10] Y. BENAYED, D. FOHR, J.-P. HATON, G. CHOLLET. *Recognition and Rejection Performance in Wordspotting Systems Using Hidden Markov modeling techniques.* in « International Workshop speech and computer - SPECOM'2002, St-Petersburg, Russia », septembre, 2002.

[11] Y. BENAYED, D. FOHR, J.-P. HATON, G. CHOLLET. *Recognition and Rejection Performance in Wordspotting Systems Using Support Vector Machines.* in « 2nd WSEAS International Conference on Signal, Speech and Image Processing - WSEAS ICOSSIP'2002, Koukounaries, Skiathos Island, Greece », septembre, 2002.

[12] Y. BENAYED, D. FOHR, J.-P. HATON, G. CHOLLET. *Support Vector Machines for Keyword Spotting.* in « International Workshop speech and computer - SPECOM'2002, St-Petersburg, Russia », septembre, 2002.

[13] A. BONNEAU, P. MOKHTARI. *A platform for the diagnosis of auditory deficiency.* in « 4th International Workshop on Enterprise Networking and Computing in Health Care Industry - Healthcom 2002, Nancy, France », juin, 2002.

[14] A. BRUN, K. SMAÏLI, J.-P. HATON. *Contribution to Topic Identification by Using Word Similarity.* in « 7th International Conference on Spoken Language Processing - ICSLP'2002, Denver, Colorado, USA », septembre, 2002.

[15] A. BRUN, K. SMAÏLI, J.-P. HATON. *WSIM : une méthode de détection de thème fondée sur la similarité entre mots.* in « Traitement Automatique des Langues Naturelles - TALN'2002, Nancy, France », juin, 2002.

[16] C. CERISARA, D. FOHR. *Fast Channel and Noise Compensation in the Spectral Domain.* in « XI European Signal Processing Conference - EUSIPCO 2002, Toulouse, France », septembre, 2002, http://www.loria.fr/publications/2002/A02-R-180/A02-R-180.ps.

[17] C. CERISARA, J.-C. JUNQUA, L. RIGAZIO. *Dynamic estimation of a noise over estimation factor for Jacobian-based adaptation.* in « IEEE International Conference on Acoustics, Speech, and Signal Processing - ICASSP 2002, Orlando, Florida », IEEE, mai, 2002, http://www.loria.fr/publications/2002/A02-R-179/A02-R-179.ps.

[18] V. COLOTTE, Y. LAPRIE, A. BONNEAU. *Modifying speech to improve the perception of L2.* in « Integrating speech technology in learning - INSTIL 2002, Davis, Ca, USA », mars, 2002.

[19] V. COLOTTE, Y. LAPRIE. *Amélioration de la précision de la resynthèse avec TD-PSOLA.* in « XXIVème Journées d'Etude sur la Parole - JEP 2002 , Nancy, France », juin, 2002.

[20] V. COLOTTE, Y. LAPRIE. *Higher precision pitch marking for TD-PSOLA.* in « XI European Signal Processing Conference EUSIPCO, Toulouse, France », septembre, 2002.

[21] K. DAOUDI. *Automatic Speech Recognition : the New Millennium.* in « International Conference on Industrial and Engineering Application of Artificial Intelligence and Expert Systems - IEA/AIE'2002, Cairns, Australia », série Lecture Notes in Artificial Intelligence, volume 1358, Springer-Verlag, éditeurs M. A. T. HENDTLASS., pages 253-263, juin, 2002, http://www.loria.fr/publications/2002/A02-R-255/A02-R-255.ps.

[22] A. DE LA TORRE, D. FOHR, J.-P. HATON. *Statistical Adaptation of Acoustic Models to Noise Conditions for Robust Speech Recognition.* in « International Conference on Spoken Language Processing - ICSLP 2002,

Denver, USA », pages 1437-1440, septembre, 2002.

[23] M. DEVIREN, K. DAOUDI. *Apprentissage de structures de réseaux bayésiens dynamiques pour la reconnaissance de la parole.* in « XXIVèmes Journées d'Études sur la Parole - JEP'2002, Nancy, France », pages 293-296, juin, 2002.

[24] M. DEVIREN, K. DAOUDI. *Continuous Speech Recognition Using Dynamic Bayesian Networks : A Fast Decoding Algorithm.* in « First European Workshop on Probabilistic Graphical Models - PGM'02, Cuenca, Spain », novembre, 2002.

[25] M. DEVIREN, K. DAOUDI. *Continuous Speech Recognition using Structural Learning of Dynamic Bayesian Networks.* in « XI European Signal Processing Conference - EUSIPCO'2002, Toulouse, France », septembre, 2002.

[26] J. DI MARTINO. *On The Use of High Order Derivatives for High Performance Alphabet Recognition.* in « International Conference on Acoustics Speech and Signal Processing - ICASSP 2002, Orlando, Florida, USA », mai, 2002, http://www.loria.fr/publications/2002/A02-R-053/A02-R-053.ps.

[27] J. DI MARTINO, Y. LAPRIE. *Un Algorithme de Réduction de la Réverbération de Signaux Issus du Vocoder de Phase.* in « XXIVe Journées d'Etude sur la Parole - JEP 2002, Nancy, France », juin, 2002, http://www.loria.fr/publications/2002/A02-R-054/A02-R-054.ps.

[28] D. FOHR, O. MELLA, I. ILLINA, F. LAURI, C. CERISARA, C. ANTOINE. *Reconnaissance de la parole pour les locuteurs non natifs en présence de bruit.* in « XXIVèmes Journées d'Etude sur la Parole - JEP'02, Nancy, France », pages 297-301, juin, 2002.

[29] I. ILLINA. *Tree-Structured Maximum a Posteriori Adaptation for a Segment-Based Speech Recognition System.* in « 7th International Conference on Spoken Language Processing - ICSLP'02, Denver, Colorado, USA », septembre, 2002.

[30] S. JAMOUSSI, K. SMAILI, J.-P. HATON. *Neural Network and Information Theory In Automatic Speech Understanding.* in « International Workshop Speech and Computer 2002 - SPECOM'2002, St-Petersburg, Russia », septembre, 2002.

[31] D. LANGLOIS, K. SMAÏLI, J.-P. HATON. *Détection de séquences par sélection de l'historique : application à la reconnaissance automatique de la parole.* in « XXIVe Journées d'Etudes sur la Parole - JEP'2002, Nancy, France », pages 301, juin, 2002, http://www.loria.fr/publications/2002/A02-R-155/A02-R-155.ps.

[32] D. LANGLOIS, K. SMAÏLI, J.-P. HATON. *Retrieving phrases by selecting the history : application to Automatic Speech Recognition.* in « 7th International Conference on Spoken Language Processing - ICSLP'2002, Denver, USA », volume 1, pages 721, septembre, 2002.

[33] Y. LAPRIE, A. BONNEAU. *Segmentation du bruit d'explosion des occlusives.* in « XXIVe Journées d'Etude sur la Parole - JEP'2002, Nancy, France », juin, 2002, http://www.loria.fr/publications/2002/A02-R-260/A02-R-260.ps.

[34] F. LAURI, I. ILLINA, D. FOHR. *Comparaison de SMLLR et de SMAP pour une adaptation au locuteur en*

utilisant des modèles acoustiques markoviens. in « XXIVe Journées d'Etude sur la Parole - JEP'02, Nancy, France », pages 289-292, juin, 2002.

[35] I. MAGRIN-CHAGNOLLEAU, N. VALLÈS-PARLANGEAU. *Audio-Indexing : what has been accomplished and the road ahead.* in « Sixth International Joint Conference on Information Sciences - JCIS'02, Durham, North Carolina, United States », pages 911-914, mars, 2002.

Rapports de recherche et publications internes

[36] D. FOHR, O. MELLA. *Le prototype SAALSA : Automatisation de la post-synchronisation.* Rapport de fin de contrat, décembre, 2002.