

Nouvelles Brèves du GFCP

Manifestations scientifiques

Acoustique et Aéroacoustique de la production des sons de parole.

Le mercredi 7 janvier 1998, 13h30 – 18h00.

A l'Ecole Nationale Supérieure des Télécommunications. 46 rue Barrault, 75634 PARIS. Metro Corvisart. Amphi Jade.

Organisé par le GFCP.

Objectifs :

Les études sur la production de la parole ont beaucoup évolué ces dernières années, en particulier au niveau des aspects de contrôle du système physiologique complexe qu'est l'appareil phonatoire humain. Par contre, les modèles de génération des sons de parole ont progressé lentement et, même si des résultats fondamentaux ont déjà été obtenus, un certain nombre de sons ne sont pas encore modélisables de façon satisfaisante, en particulier parmi les consonnes. Ceci peut s'expliquer par des difficultés de résolution formelle mais aussi, et surtout, parce que la plus grande partie des phénomènes opérants dépasse le cadre des approximations usuelles.

Le but de l'après-midi est de présenter certaines améliorations de ces théories au premier ordre et leurs conséquences sur la compréhension et la modélisation de la production des sons de parole. Bien qu'abordant des domaines de pointe, les différents exposés se veulent suffisamment introductifs pour intéresser les non-spécialistes.

Programme :

Introduction (Xavier Pelorson, Institut de la Communication Parlée)

Principes de base de la production des sons de parole: génération de son par un écoulement, propagation dans le conduit vocal, rayonnement. Hypothèses de base. Limites de validité.

I. Modélisation des sources auto-oscillantes (cordes vocales, lèvres). (Joël Gilbert, Laboratoire d'Acoustique de l'Université du Maine)

Principe de fonctionnement, Aspects théoriques, Couplage acoustique

Application à la parole. Illustration sonore sur la base d'un modèle de cordes vocales. Importance de certains phénomènes d'écoulement.

II. Propagation acoustique (Vincent Pagneux, Laboratoire d'Acoustique de l'Université du Maine)

Théorie modale de la propagation, Couplage fluide-parois, Méthodes de résolution.

Application à la parole. Illustration de l'existence et de l'importance des modes d'ordre supérieurs. Mesures sur maquettes, illustration sonore (voyelles, fricatives ...).

III. Introduction à l'aéroacoustique (Marc-Pierre Verge, Laboratoire d'Acoustique Musicale)

Description systématique des interactions écoulement-acoustique, Théorie de Lighthill, formulation de Howe, Powell ...

Participation aux frais :

Permanents : 150 Francs,

Etudiants : 50 Francs.

Un aide mémoire sera distribué à l'issue de l'après midi.

Important : Afin de faciliter l'organisation de cette journée, merci de bien vouloir vous pré-inscrire à l'avance.

Renseignements, pré-inscriptions : Xavier Pelorson, ICP, 46, av. F. Viallet, F-38031 Grenoble Cedex 01; tel : 04-76-57-45-36, Email : pelorson@icp.inpg.fr

Bourses du GFCP

Le GFCP a décidé d'attribuer quelques bourses pour des étudiants souhaitant assister aux séminaires, sur la base de 50 % du prix normal du billet de train AR (2e classe) entre le laboratoire et le lieu du séminaire (ce qui finance à peu près l'intégralité du prix du transport au tarif découverte 30 ou au tarif -50% pour les moins de 25 ans titulaires d'une carte 12-25). Le montant total que le GFCP peut allouer à ces bourses étant limité, les candidats à de telles bourses sont priés de se faire connaître le plus rapidement possible. Les demandes doivent être adressées par e-mail à Frédéric BIMBOT (bimbot@irisa.fr).

LES XXII^{èmes} JOURNEES D'ETUDE SUR LA PAROLE

à MARTIGNY (Suisse), du 15 au 19 juin 1998

Organisées par

l'Institut Dalle Molle d'Intelligence Artificielle Perceptive (IDIAP)
et le Groupe Francophone de la Communication Parlée (GFCP),

sous le patronage

- de la Société Française d'Acoustique (SFA)
- de l'Européen Speech Communication Association (ESCA)
- de SWISSCOM
- du FNSRS (Fonds National de la Recherche Scientifique)
- et de la ville de Martigny.

Objectifs

Les XXII^{èmes} JEP seront consacrées aux travaux couvrant l'ensemble des aspects fondamentaux de la communication parlée et les différents problèmes de son traitement automatique. Une attention particulière sera apportée aux recherches qui visent à la compréhension et/ou à la simulation des processus de décision permettant la fusion de données et de connaissances hétérogènes (numériques et symboliques, sensorielles et cognitives, multimodales, etc.).

Principaux thèmes abordés

Les articles proposés aux XXII^{èmes} JEP s'inscriront dans l'un des thèmes généraux suivants :

1. Production et perception de la parole
2. Phonétique et phonologie
3. Prosodie
4. Reconnaissance et compréhension de la langue
5. Reconnaissance/identification de la langue et du locuteur
6. Synthèse vocale
7. Modèles de langage pour le traitement automatique de la parole
8. Applications intégrant une composante orale
9. Dialogue à composante orale, etc.

Les problèmes peuvent être abordés de différents points de vue :
acoustique, médical, sociologique, psychologique, neuroscientifique,
linguistique, technique, etc.

Comité scientifique

M. ADDA-DECKER, LIMSI, Paris (F)
R. ANDRE-OBRECHT, IRIT, Toulouse (F)
F. BIMBOT, ENST, Paris - IRISA, Rennes (F)
J.-F. BONASTRE, LIA, Avignon (F)
J.-L. COCHARD, IDIAP, Martigny (CH)
P. DUPONT, CMU, Pittsburgh (USA)
P. DELEGLISE, LIUM, Le Mans (F)
J.-M. HOMBERT, DDL, Lyon2 (F)
Y. LAPRIE, CRIN, Nancy (F)
R.K. MOORE, DERA, Malvern (UK)
C. MONTACIE, LAFORIA, Paris (F)
P. PERRIER, ICP, Grenoble (F)
J. SCHOENTGEN, ULB, Bruxelles (B)
R. SOCK, IP, Strasbourg (F)
B. TESTON, LPL, Aix-en-Provence (F)
J. VAISSIERE, IPP, Paris (F)

Comité d'organisation

H. BOURLARD (Président), C. PILLET (Secrétaire),
O. BORNET, G. CALOZ, D. GENOUD, C. JABOULET, S. KRSTULOVIC.

Organisation des sessions

Les travaux seront exposés aux XXIIèmes JEP sous la forme de conférences à thèmes, de communications orales ou affichées et de démonstrations.

Le mode de présentation sera adapté aux contraintes scientifiques et techniques concernant les thèmes et le nombre de contributions.

Soumission des articles

Le texte complet des propositions de communications sera évalué par deux rapporteurs désignés par les membres du comité scientifique. La langue d'expression est le français. Cependant le comité scientifique examinera les propositions de communications écrites en anglais, pour autant que le(s) auteur(s) s'engage(nt) à traduire son (leur) texte en français pour son édition dans les actes.

Les articles ne dépasseront pas 4 pages A4, accompagnées d'un court résumé.

La description exacte des format de mise en page (MS-WORD et LaTeX) sera bientôt disponible sur notre serveur WEB. Les articles qui ne respecteraient pas ces règles de mise en page pourront être renvoyés pour correction.

Les textes sont à envoyer pour le 19 janvier 1998 au plus tard au secrétariat des JEP 98.

Dates à retenir

- 03 février 1998 (date limite de soumission des articles complets)
- 15 mars 1998 (décision concernant les articles retenus et notification aux auteurs)
- 14 avril 1998 (date limite de réception des articles prêts à imprimer)

Questions Scientifiques

Les questions scientifiques sont à adresser à :

M. Frédéric Bimbot
Président du GFCP
Téléphone : +33-2-99.84.75.06
E-mail : Frederic.Bimbot@irisa.fr

RENSEIGNEMENTS

Secrétariat XXIIèmes JEP
IDIAP
CP 592
CH-1920 Martigny (SUISSE)
Téléphone : +41 27 721 77 11
Télécopie : +41 27 721 77 12
E-mail : JEP98@idiap.ch

WEB: <http://www.idiap.ch/JEP98/>

Colloque RLA2C

Le GT-1 organise à Avignon un colloque centré sur "la reconnaissance du locuteur et ses applications commerciales et criminalistiques" du 20 au 23 avril 1998.

Adresse Web : <http://lia.univ-avignon.fr/RLA2C>

Informations : RLA2C@univ-avignon.fr

Séminaire sur la reconnaissance automatique de la langue parlée

Le GT-1 prépare l'organisation d'un séminaire sur la reconnaissance de la langue parlée vers l'automne 1998.

Informations : Régine Andre-Obrecht (obrecht@irit.fr)

Nos Thèses et DEAs : passés et futurs...

Christian Abry

Thèse pour le Doctorat d'Etat ès Lettres et Sciences Humaines

Vendredi 28 novembre à l'Université-Stendhal ou Grenoble III

De la parole articulatoire à la parole narrative

ou

Du Soi vocal à l'Autre narratif dans une "Théorie de l'esprit"

Résumé

Les travaux et les positions de recherche exposés sous ce titre nous feront passer des plus petites unités de la parole, les coordinations articulatoires mises en jeu dans le contrôle de la production des sons, aux objets qui sont traditionnellement considérés comme les plus complexes dans l'exercice de la parole, les récits.

Dans l'état des connaissances actuelles, une vision nouvelle sur les fondements de la parole pourra être proposée sur la base de plusieurs homologues en neurocognition : comparée (primatologie), développementale (comportement et corticogénèse) et pathologique (aphasiologie et neuropsychiatrie cognitive). Cette proposition d'une histoire naturelle de la parole sera évaluée à la lumière d'expériences menées in vivo et en robotique sur le contrôle de la production vocale oro-faciale.

Il n'en ira pas de même pour les fondements de la parole narrative. En conséquence, la démarche suivie dans ce domaine ne sera ni expérimentale, ni théorique, mais volontairement descriptive. C'est ainsi à l'exploration d'une parole narrative ethnographiquement située, en l'occurrence, dans la région alpine, que nous serons conviés, et ceci grâce à un fonds narratif de tradition orale, unique en France par son extension et sa densité. Il nous permettra tout à la fois d'établir une véritable géographie sociale contrastée des croyances et de mettre en évidence ce qu'une approche expérience-centered pourrait apporter, dans ce champ, pour explorer une ontologie des croyances qui nous semble nécessaire pour les fonder en amont de l'ethnohistoire et de l'étymologie.

Mais au préalable, le nécessaire recours à la parole dialectale, ici francoprovençale, nous aura donné l'occasion de passer du laboratoire des sciences de la parole au terrain, pour les sons comme pour les noms.

Au bout du compte l'ethnographie des origines de la narration, jointe à l'étymologie des nominations, se révéleront plus payantes que les reconstructions structurales ou comparatives, qui resteront épistémologiquement disqualifiées aussi longtemps qu'on ne pourra distinguer dans ce domaine les analogies des homologues. Une ethnographie payante pour la parole narrative, puisque nous verrons se donner explicitement comme telle la matrice lexico-narrative à l'origine de tous les récits : une matrice fonctionnellement apotropaïque, narrativement auto-étiologique et mnémotechniquement énumérative, que l'on aurait été bien en peine d'inventer par la seule logique de la narration, quelle que soit la méthode d'analyse des exemplaires et des prototypes utilisée, dans l'espace comme dans le temps.

Enfin deux très récentes percées convergentes en expérimentation neurofonctionnelle sur le cerveau, l'une dans le domaine du monitoring de la parole intersensorielle, l'autre dans l'exercice de la narration, nous permettront de poser l'intérêt de la cognition en parole pour les recherches sur la "Théorie de l'esprit". Aurions-nous dès maintenant les linéaments de la genèse qui va du soi vocal à l'autre narratif... la capacité mentale d'une telle théorie nous étant absolument nécessaire aussi bien pour le suivi de notre voix que pour l'écoute du récit d'origine de tous les récits.

Driss MATROUF

"Adaptation des modèles acoustiques pour la reconnaissance de la parole bruitée".

driss@limsi.fr

17/12/97

Résumé

Cette thèse se situe dans le cadre de la reconnaissance de la parole bruitée par des modèles statistiques (Modèles de Markov Cachés à Distributions Continues : MMCDC).

Le meilleur système de reconnaissance pour des conditions particulières d'utilisation est celui qui est appris dans ces mêmes conditions. L'objectif principal de ce travail est d'approcher au mieux cette configuration. Pour cela nous proposons une nouvelle approche d'adaptation qui consiste en une compensation explicite des bruits (additif et convolutif). Elle est fondée sur le modèle du canal de transmission : $y = (s+n)*h$, où y , s , n et h désignent respectivement le signal observé, le signal propre correspondant, le bruit additif ambiant et le bruit convolutif. Cette technique permet avec peu d'approximations d'adapter les paramètres statistiques statiques et dynamiques du système.

L'hypothèse que les données d'apprentissage sont propres est rarement vérifiée dans la pratique. Nous proposons une technique plus générale que la première pour compenser conjointement les bruits additifs et convolutifs de test et d'apprentissage.

Dans certaines applications on s'intéresse à la qualité du signal acoustique et non pas au message linguistique qu'il porte. Nous proposons une technique de débruitage fondée sur la modélisation du signal propre par des MMC autorégressifs (MMCAR) et du bruit par une gaussienne AR. Le processus d'estimation des filtres est itératif et nécessite une bonne initialisation pour converger vers la bonne solution. Pour cela nous proposons une nouvelle stratégie combinant l'utilisation des MMC fondés sur le cepstre, l'utilisation des MMCAR et la technique d'adaptation développée dans cette thèse. Cette technique de débruitage donne des résultats très satisfaisants du point de vue perceptif (augmentation significative du rapport signal sur bruit avec très peu de distorsions).

Girin Laurent

Institut National Polytechnique de Grenoble

Soutenance de thèse

le 19 décembre 1997

National Polytechnique de Grenoble, 46 avenue Félix Viallet,

Spécialité Signal-Image-Parole, intitulée:

Débruitage de parole par un filtrage utilisant l'image du locuteur

Thèse préparée à l'Institut de la Communication Parlée, INPG/Université

RESUME:

Un problème majeur pour les systèmes de télécommunications est celui du débruitage de parole, c'est-à-dire l'atténuation des effets d'un bruit parasite en vue d'améliorer l'intelligibilité et la qualité du message. Or, l'homme possède en ce domaine une compétence particulière, celle de pouvoir extraire l'information grâce aux signaux captés visuellement sur le visage de l'interlocuteur. L'utilisation "automatisée" des informations visuelles a déjà permis d'améliorer la robustesse des systèmes de reconnaissance de parole, principalement en milieu bruité. Cette thèse traite du problème inédit d'un débruitage audiovisuel qui n'implique pas une tâche de classification, mais de génération de parole acoustique intelligible à partir de parole acoustique dégradée et d'une information optique supplémentaire (forme du contour labial). Abordé sous l'angle du filtrage, l'objectif est l'élaboration d'un système de débruitage fiable, simple et efficace, utilisant des filtres rehausseurs estimés pour une grande part à partir d'informations visuelles.

girin@icp.inpg.fr

Étude des marqueurs discursifs dans le dialogue finalisé

Nathalie Colineau

Spécialité Sciences Cognitives

Date : 12 décembre 1997

Lieu : Amphi maison J. Kuntzman

Heure : 10h30

Résumé :

L'étude que nous avons réalisée se place dans le cadre des analyses de discours, avec pour objectif de donner une interprétation des énoncés produits, en s'appuyant sur des marques linguistiques, prosodiques, actionnelles et contextuelles.

L'idée sur laquelle repose l'étude de ces marques est qu'un locuteur donne dans son énonciation suffisamment d'indices pour être compris, indices qui permettront à l'interlocuteur d'interpréter ce que le locuteur a voulu dire. En effet, il apparaît que ces indices linguistiques reflètent l'organisation voulue par le locuteur et que l'allocutaire les exploite comme guides lors de l'interprétation des énoncés.

Dans un premier temps, une analyse linguistique a été faite sur un corpus de dialogues oraux finalisés. Ce travail a permis de déterminer un ensemble de marques pertinentes, et de répertorier leurs différents emplois dans chaque contexte d'énonciation. Trois types de marques ont été retenus: les marques intervenant dans la détermination des actes de dialogue, les marques de connexion dont le rôle est de marquer la structure du discours et les marques de cohésion rendant compte des chaînes thématiques.

Cependant les marques linguistiques à elles seules ne permettent pas dans tous les cas de déterminer l'acte de dialogue accompli, d'autres marques de nature extra-linguistique doivent être intégrés dans l'analyse. En effet, la situation dans laquelle sont produits les énoncés et le rôle de chacun des partenaires du dialogue contribuent aussi à la détermination des actes de dialogue accomplis. Afin d'intégrer ces informations dans l'analyse, un modèle des enchaînements d'énoncés a été élaboré, tenant compte à la fois du type d'actes de dialogue accomplis à chaque tour ainsi que du rôle occupé par le locuteur dans l'interaction.

Dans un second temps, un modèle informatique a été élaboré comme outil pour tester la pertinence de ces marques et ainsi la validité de notre approche. La tâche du système consiste à partir des marques discursives relevées, à déterminer si l'énoncé produit est une requête, une question, une information donnée, etc.

Enfin, pour valider notre approche et montrer la pertinence des marques discursives collectées, nous avons analysés trois corpus de dialogues oraux. Nous montrons ainsi le comportement règle de ces marques et l'importance de leur prise en compte dans l'interprétation des énoncés.

Laboratoire de Communication Langagière et d'Interaction Personne Système (CLIPS-IMAG)

e-mail : nathalie.colineau@imag.fr

<http://hermes.imag.fr/~colineau/>

Génération multiparamétrique de la prosodie du français par apprentissage automatique

Yann MORLEC

vendredi 19 Décembre, Grenoble
Institut de la Communication Parlée / INPG
morlec@icp.inpg.fr
http://www.icp.inpg.fr/~morlec

RESUME :

Cette thèse a pour objet la conception, l'apprentissage automatique et l'évaluation d'un modèle de génération de la prosodie du français.

Ce modèle, d'inspiration cognitive, suppose un encodage direct des informations linguistiques et paralinguistiques véhiculées par un énoncé via des formes prosodiques prototypiques. Ces formes - ou mouvements - sont enchaînées au sein d'un même niveau linguistique (phrase, groupe, mot...) et se superposent à des enchaînements encodant les niveaux supérieurs. Cette morphologie superpositionnelle opère de manière multiparamétrique.

Dans le cadre de cette thèse, nous avons étudié deux niveaux linguistiques - la phrase et le groupe - et deux dimensions prosodiques - la mélodie et le rythme. Pour ces deux niveaux, nous avons élaboré un système de génération conjoint des contours mélodiques et rythmiques, paramétré par un apprentissage automatique hiérarchique sur des corpus de parole naturelle. L'architecture du modèle est un réseau de neurones récurrent à connexions partielles.

Les corpus de parole ont été conçus de façon à prendre en compte le phénomène linguistique étudié avec une représentativité statistique suffisante, et à faire varier systématiquement le contenu des niveaux linguistiques inférieurs.

À partir d'un corpus de phrases isolées prononcées par un locuteur selon six consignes actitudinales, nous avons entrepris la génération des prototypes mélodiques et rythmiques du niveau de la phrase véhiculant l'attitude du locuteur vis-à-vis de son propos.

Nous avons également étudié, sur la base d'un corpus de phrases déclaratives, les modulations du niveau porte de groupe prenant en charge la fonction de structuration de l'énoncé.

Chaque étape d'apprentissage de notre modèle a été sanctionnée par un test de perception destiné à évaluer sa capacité à capturer les informations nécessaires et suffisantes pour communiquer les fonctions linguistiques données comme consignes a priori dans chaque corpus.

A. BIEM

EXTRACTION DE PARAMETRES DISCRIMINATIFS;
APPLICATION A LA RECONNAISSANCE DE LA PAROLE.

le 28 - 11 - 1997

À l'Université Paris 6

Ce travail de thèse a été réalisé au laboratoire ATR au Japon

Résumé

La reconnaissance des formes suppose l'application de deux processus fondamentaux: la paramétrisation du signal et le décodage ou classification.

La paramétrisation, ou extraction de traits, a pour but premier de trouver des paramètres pertinents du signal. L'extracteur de traits se fonde généralement sur la statistique des données (PCA, LDA,...) ou la connaissance a priori (expertise dans le domaine concerné). Ainsi, en reconnaissance de la parole, les modèles auditifs ou perceptifs sont abondamment utilisés.

Le décodage, réalisé à travers plusieurs types de décodeurs tels que les modèles de Markov cachés ou les réseaux de neurones, s'occupe principalement de la classification des traits extraits et trouve sa fondation dans la théorie bayésienne de décision qui garantit théoriquement une probabilité d'erreur minimale.

Les méthodes classiques en reconnaissance des formes, et en particulier en reconnaissance de la parole, souffrent par conséquent d'un problème de cohérence: les processus de paramétrisation et de décodage sont fondés sur des critères différents, bien qu'ayant une cible commune. Ceci entraîne une sous-optimalité des méthodes conventionnelles.

Dans cette thèse, nous proposons la méthode DFE (Discriminative Feature Extraction) comme une solution à ce problème. La méthode DFE (Biem et al, 1993) suppose un extracteur de traits paramétrisé et l'optimise en fonction des erreurs de classification. Ceci réalise un système global, intègre et optimise en vue d'une minimisation beaucoup plus directe des erreurs du système. La méthode DFE se fonde sur le cadre de travail MCE (Minimum Classification Error) (Juang et Katagiri, 1992) et peut être considérée comme une extension de ce cadre de travail à l'extraction des traits. Le principal avantage réside dans le

fait que cette méthode peut être appliquée à n'importe quels types de paramètres utilisant un décodeur quelconque et garantit l'optimalité locale de l'extracteur de paramètres et du décodeur ainsi conçus. À travers quelques applications en reconnaissance de la parole, en l'occurrence, l'optimisation d'une pondération cepstrale et le réajustement des fréquences centrales, des bandes passantes et des gains d'un banc de filtres, nous montrons que les traits DFE permettent de concevoir un système plus performant que l'utilisation classique du critère MCE ou du Maximum de Vraisemblance (MLE). L'analyse du banc de filtres et de la pondération cepstrale semble suggérer que les traits DFE, bien que basés sur le critère de la minimisation des erreurs, ont tout de même une bonne interprétation acoustique et par conséquent ouvre la voie à un nouveau type d'analyse du signal.