

# Nouvelles Brèves du GFCP

## Manifestations scientifiques

PERCEPTION DU MONDE ET PERCEPTION DU LANGAGE  
15, 16 et 17 Octobre 1999

Université Marc Bloch STRASBOURG

Laboratoire de Psychobiologie du Comportement Moteur et des Sports - Strasbourg

Sciences Cognitives, Linguistique et Intelligence Artificielle (SCOLIA)

Groupe Francophone de la Communication Parlée - Société Française d'Acoustique (GFCP-SFA)

Institut de la Communication Parlée de Grenoble (ICP)

Réseau Cognisciences Est (CNRS)

Centre Interdisciplinaire de Recherche en Linguistique Et Psychologie cognitive (CIRLEP)  
Université de Reims.

Qui observe depuis une vingtaine d'années la vaste scène de ce qu'on nomme les sciences cognitives – notamment, la biologie, la neurophysiologie, les sciences du sport, la psychologie, la linguistique – y constate un mouvement par lequel les chercheurs, sous la pression du “réel”, tendent à prendre leur distance par rapport à l'appareil conceptuel hérité de ce que l'on pourrait appeler la “science classique”.

C'est la saisie de ce mouvement inédit – de son enracinement philosophique, de son orientation, des contenus épistémologiques, théoriques et méthodologiques qu'il met peu à peu au jour – que les organisateurs du congrès “Perception du monde et perception du langage” ont en vue.

L'être humain en tant que sujet agissant et doué de parole, se construit, évolue et *communique* au sein d'un univers, objectif et intersubjectif, *complexe*. Les avancées considérables liées aux progrès technologiques ont conduit à une révision souvent radicale des principes théoriques à partir desquels on pensait rendre compte des mécanismes qui sous-tendent l'appréhension de l'univers par la “conscience incorporée”. En effet, une meilleure compréhension des mécanismes dynamiques et complexes qui, du mouvement de l'action corporelle au mouvement de l'énonciation ‘langagière’, sont responsables du fonctionnement du système nerveux et de la constitution de ses ‘objets’ n'a été rendue possible qu'en vertu d'une nouvelle façon de penser le vivant; au vrai, qu'en vertu d'une nouvelle façon de penser le connaître scientifique, d'une nouvelle *épistémè*.

Aussi, l'objet de ce colloque est de redessiner un projet épistémologique, théorique et pratique dont les thématiques abordent les perceptions sur un plan philosophique, dans le cadre du mouvement et de la corporéité ainsi que dans le champ de la parole.

Les intervenants s'attacheront à mettre en évidence et à établir des relations entre les différents aspects de la perception. Les exposés d'une trentaine de minutes seront regroupés par centres d'intérêts; ils se voudront accessibles à un assez large public: universitaires de diverses disciplines, doctorants, jeunes chercheurs, etc.

### Vendredi 15 Octobre 1999

8h.30 : Accueil des participants

9h. : Le mot du président de l'Université Marc Bloch de Strasbourg. D. Payot

10h. à 10h.30 : J.P. Durafour : Mise en perspective thématique du colloque.

10h.30 à 11h. : J.L. Petit : De l'intentionnalité de l'acte.

11h. à 11h.30 : E. Coseriu : Le langage : diacriticon tes ousias.

11h.30 à 12h. : A. Murgia : Le tournant kantien dans la philosophie de l'esprit.

12h à 12h.30 : *Questions aux intervenants.*

**12h.30 à 14h.30 : Déjeune**

14h.30 à 15h. : J.F. Demonet : Les bases neurophysiologiques des perceptions.

15h. à 15h.30 : Ph. Fuchs : Problématique du comportement humain dans un environnement virtuel.

15h.30 à 16h. : A. Rauch : Perceptions du corps; considérations historiques.

16h. à 16h.30 : *Questions aux intervenants.*

**16h.30 à 17h. : Pause.**

17h. à 17h.30 : D. Keller : Corps et perceptions dans l'exercice physique et sportif.

17h.30 à 18h. : J. Stewart : Le sens biologique.

18h. à 18h.30 : *Questions aux intervenants.*

**Samedi 16 Octobre 1999**

10h. à 10h.30 : E. Tyvaert : Le rôle de la langue dans la tentative de maîtrise du monde par la parole.

10h.30 à 11h. : F. Cordier : Ontologies et catégories.

**11h. à 11h.30 : Pause.**

11h.30 à 12h. C. Chevrier-Muller : La perception et les pathologies du langage.

12h. à 12h.30 : *Questions aux intervenants.*

**12h.30 à 14h.30 : Déjeuner.**

14h.30 à 15h. : C. Abry : Du soi vocal à la théorie de l'esprit.

15h. à 15h.30 : J.L. Schwartz : Perception de la parole.

15h.30 à 16h. : L.-J. Boë Perceptions et diffusions des théories scientifiques : erreurs et leurres.

16h. à 16h.30. : *Questions aux intervenants.*

16h.30 à 17h. : R. Sock : A propos de la viabilité en production-perception de la parole.

17h. à 17h.30 : J.F.-P. Bonnot : Structuration phonologique dans l'acte de parole.

17h.30 à 18h. : *Questions aux intervenants.*

**Dimanche 17 Octobre 1999**

9h.30 à 10h. : G. Kleiber : L'extra-linguistique dans la construction du sens.

10h. à 10h.30 : J.P. Durafour : La théorie des esquisses et la genèse du sens.

10h.30 à 11h.30 : Table ronde : synthèse et conclusions.

**COMITE D'ORGANISATION**

J.F.-P. Bonnot - J.P. Durafour - D. Keller - R. Sock

Université Marc Bloch de Strasbourg

**INSCRIPTION / RENSEIGNEMENTS**

Renseignements uniquement le matin :

Tél. 03.88.41 78 57

Fax. 03.88.41.74.75

Adresse électronique: sock@umb.u-strasbg.fr

Talon d'inscription au colloque à retourner avant le 15 septembre 1999

Nom :

Prénom :

Adresse :

Ci-joint chèque :

n°:

Repas du Vendredi midi : OUI / NON

Prix : 80.francs

Repas du Vendredi Soir : OUI / NON

Prix : 150 francs

Repas du Samedi midi : OUI / NON

Prix : 80 francs

Repas du Samedi Soir : OUI / NON

Prix : 200 francs.

Réservation et règlement par chèque à l'ordre du Trésorier de l'association :

**Laboratoire de Recherche en Sports**

14, rue Descartes - 67084 Strasbourg Cedex

\*\*\*\*\*

**Séminaire**

**"Le Traitement Automatique de la Parole"**

**AVIGNON - 17/11/99**

**Découvrir et/ou Approfondir les Technologies clés et les Applications  
du Traitement Automatique de la Parole**

Organisé par le  
**Groupe Francophone de la Communication Parlée de la Société  
Française d'Acoustique**

et le

**LIA - Laboratoire Informatique de l'Université d'Avignon  
Avec l'aide du Réseau Jeunes Chercheurs en Parole**

**Programme Provisoire :**

10h20 - 10h30 : Accueil

10h30 - 11h10 : Introduction sur le Traitement Automatique de la Parole

Frédéric Bimbot (IRISA/Rennes)

11h10 - 11h50 : Eléments de traitement et de caractérisation du signal pour la Parole

Pascal Perrier (ICP/Grenoble) et Régine André-Obrecht (IRIT/Toulouse).

11h50 - 12h30 : Les algorithmes classiques en traitement de la parole

Pierre Dupont (Université Saint-Etienne) et Thierry Artieres (LIP6/Paris)

Repas

14h30 - 15h10 : Reconnaissance Automatique de la Parole

Martine Adda-Decker (LIMSI/Paris) et Dominique Fohr (Loria/Nancy)

15h10 - 15h50 : Reconnaissance Automatique du Locuteur

Delphine Charlet (CNET/Lannion) et Jean-françois Bonastre (LIA/Avignon)

15h50 - 16h30 : Synthèse de la Parole

Frédéric Béchét (LIA/Avignon) et François Yvon (ENST/Paris)

Pause

17h00 - 17h20 : Méthodologie d'expérimentation et d'évaluation

Joseph Mariani (LIMSI/Paris)

17h20 - 18h00 : Caractéristiques des applications commerciales en Parole

Claire Waast (IBM/Paris) et Denis Juvet (CNET/Lannion)

**Renseignements :**

JF Bonastre - 04 90 84 35 14 - <mailto:jean-francois.bonastre@lia.univ-avignon.fr> jean-  
francois.bonastre@lia.univ-avignon.fr

[http://www.lia.univ-avignon.fr/heberges/sem\\_GFCP/index.html](http://www.lia.univ-avignon.fr/heberges/sem_GFCP/index.html)

**Inscription :**

<<http://www.eurecom.fr/~delacour/GFCP/inscript.html>>

<http://www.eurecom.fr/~delacour/GFCP/inscript.html> Ou envoyer un mail à

<[perrine.delacourt@eurecom.fr](mailto:perrine.delacourt@eurecom.fr)> [perrine.delacourt@eurecom.fr](mailto:perrine.delacourt@eurecom.fr)

\*\*\*\*\*

**RJC'99**

**TROISIEMES RENCONTRES  
JEUNES CHERCHEURS EN PAROLE**

**Centre d'Enseignement et de Recherche en Informatique  
- Avignon, 17 au 19 novembre 1999 -**

**ATTENTION:** Nouvelles dates d'inscription et de demande de bourses

La troisième édition des Rencontres Jeunes Chercheurs en Parole, RJC'99, se tiendra en Avignon du 18 au 19 Novembre 1999.

Cette manifestation, parrainée par le Groupe Francophone de la Communication Parlée (GFCP), donne aux doctorants ou jeunes docteurs l'occasion de se rencontrer, de présenter leurs travaux et de discuter sur les domaines divers et variés de la Parole.

Deux types de communications seront proposes:

- \* sessions affichées ou les jeunes chercheurs pourront exposer leurs travaux en cours
- \* sessions orales dédiées aux doctorants en fin de thèse désirant exposer plus en détail leurs travaux, en vue de leur soutenance de thèse.

Ces journées sont aussi ouvertes aux jeunes chercheurs désirant simplement assister aux présentations sans proposer eux-mêmes une communication.

En outre, en association avec le GFCP, les participants aux RJC'99 seront invités à assister, gratuitement, au séminaire sur la Parole qui aura lieu en Avignon le 17 Novembre 1999.

### **Communications**

Les propositions de communication, composées d'un résumé de 2 pages A4 maximum, seront à envoyer, avant le 1er octobre 1999, sous forme électronique (format PostScript, PDF ou RTF) ou par courrier à l'adresse suivante:

RJC99@lia.univ-avignon.fr  
ou  
Teva Merlin - RJC'99  
LIA/CERI - Agroparc - BP 1228  
339, chemin des Meinajaries  
F-84911 Avignon Cedex 9

Un recueil de résumés sera publié à l'issue de ces rencontres.

### **Inscriptions**

L'inscription aux RJC'99 est gratuite. Elle s'effectuera directement par le site Web des RJC'99 à l'adresse suivante:

<http://www.lia.univ-avignon.fr/RJC/>

Les personnes ne pouvant accéder au site Web, nous ferons parvenir par courrier électronique ou postal une demande de bulletin de participation (voir adresse ci-dessus).

### **Bourses**

Des bourses couvrant une partie des frais de séjour ou de déplacement pourront être accordées, en nombre limité, sur demande des doctorants.

(Voir modalités à l'adresse: <http://www.lia.univ-avignon.fr/RJC/>)

### **Dates importantes**

Date limite de soumission : 15 octobre 1999  
Date limite d'inscription : 15 octobre 1999  
Date limite pour demande de bourse : 15 octobre 1999  
Séminaire Parole GFCP : 17 novembre 1999  
RJC'99 : 18-19 novembre 1999

### **Partenaires associés**

Centre National d'Études en Télécommunication (CNET)  
Société Française d'Acoustique - Groupe Francophone de la  
Communication Parlée (SFA-GFCP)  
Institut Eurecom  
Institut Universitaire Professionnalise Informatique d'Avignon (IUP)  
Laboratoire Informatique d'Avignon (LIA)  
Université d'Avignon et des Pays de Vaucluse (UAPV)  
Région Provence - Alpes Cote d'Azur

Vous pouvez retrouver tous ces renseignements sur le site du Réseau  
Jeunes Chercheurs en Parole:  
<http://www.lia.univ-avignon.fr/RJC/>  
Si vous n'avez pas accès à ce site ou si vous rencontrez le moindre  
problème, vous pouvez nous contacter :  
[perrine.delacourt@eurecom.fr](mailto:perrine.delacourt@eurecom.fr)  
ou  
[corinne.fredouille@lia.univ-avignon.fr](mailto:corinne.fredouille@lia.univ-avignon.fr)

\*\*\*\*\*

## Nos Thèses, passées et futures...

Thèse en Microélectronique - Institut National Polytechnique de Grenoble.  
Préparée à l'Institut de la Communication Parlée, INPG / CNRS / Université  
Stendhal

46, av. Félix Viallet 38031 Grenoble Cedex 01, France

Directeur de thèse : **Christian Benoît**

**Titre : Conception et évaluation d'un système de suivi automatique des gestes labiaux en parole**

**Auteur : Lionel Revéret, [reveret@icp.inpg.fr](mailto:reveret@icp.inpg.fr)**

### Résumé

Cette thèse présente un système de suivi automatique en 3D des mouvements labiaux d'un locuteur, à partir d'une séquence vidéo, sans utilisation de maquillage. Il combine à la fois une approche ascendante, orientée " image " et une approche descendante, orientée " modèle ", de l'analyse de la forme et du mouvement des lèvres. Dans un premier temps, une analyse statistique apprend la séparation optimale entre les couleurs des lèvres et de la peau du locuteur. L'analyse de n'importe quelle image couleur RGB fournit alors une image en niveaux de gris où le contraste entre lèvres et peau est rehaussé. Parallèlement, la forme et les mouvements des lèvres du locuteur sont représentés avec précision par un modèle linéaire 3D contrôlé par seulement trois paramètres articulatoires. Le modèle est appris à partir d'une Analyse en Composantes Principales (ACP). L'ACP est appliquée sur un étiquetage selon 30 points de contrôle 3D d'un ensemble de formes de lèvres phonétiquement pertinentes : ces formes visent à explorer les extrêmes des réalisations articulatoires du locuteur. Le suivi automatique est ensuite réalisé par synthèse et projection du

modèle 3D sur l'image 2D provenant de l'analyse couleur. Une procédure d'optimisation ajuste les paramètres articulatoires de telle sorte que la projection 2D du modèle corresponde au mieux avec la zone de probabilité estimée pour les lèvres. Ce système combine ainsi de manière hybride la précision de l'analyse chromatique et la robustesse imposée par le modèle. Le système est évalué sous divers aspects : ses capacités à s'adapter à la morphologie labiale et aux stratégies articulatoires de plusieurs locuteurs, la qualité des mesures géométriques délivrées et sa rapidité d'analyse. Le système complet a été implanté et testé en langage C sur une station de travail monoprocesseur. L'exécution est évaluée en nombre d'instructions à partir du code machine généré par le compilateur du système de la station. Ces résultats ont permis d'identifier les zones critiques de traitement pour lesquels des optimisations sont proposées. Compte tenu de ces optimisations, il apparaît que la cadence de 50 images par seconde est alors accessible sans avoir recours à une implantation matérielle spécialisée.

Mots-clés : parole audiovisuelle, suivi automatique des lèvres, analyse chromatique, modélisation articulatoire, analyse / synthèse de modèle 3D, temps réel.

\*\*\*\*\*

Thèse en Signal Image Parole - Institut National Polytechnique de Grenoble.

Préparée à l'Institut de la Communication Parlée, INPG / CNRS / Université

Stendhal

46, av. Félix Viallet 38031 Grenoble Cedex 01, France

**Titre : ANALYSE DE SCENES AUDITIVES COMPUTATIONNELLE (CASA) :  
Un nouvel outil de marquage du plan temps-fréquence par détection  
d'harmonicité exploitant une statistique de passages par zéro**

**Auteur : Francois Gaillard, [gaillard@icp.inpg.fr](mailto:gaillard@icp.inpg.fr)**

**Résumé :**

Pour modéliser notre capacité à structurer notre environnement sonore, l'une des approches proposées par l'Analyse de Scènes Auditives (ASA) consiste considérer que cette capacité de notre système auditif résulte de l'utilisation, en coopération, de plusieurs images du plan temps-fréquence, construites à partir de l'extraction d'indices primitifs des signaux. Dans ce cadre, et au carrefour du traitement du signal, de la physiologie et de la reconnaissance de la parole, ce travail de thèse présente une méthode de marquage du plan temps-fréquence basée sur les propriétés harmoniques des sons voisés. Cette méthode est basée sur le principe d'une méthode ancienne d'extraction de pitch, la " méthode PPZ " (i.e. méthode des passages par zéro), connue pour sa sensibilité à la présence d'interférence. Ce travail de thèse permet de montrer que cette sensibilité peut être tournée en avantage pour la détection d'harmonicité en conditions interférentes. En effet, la statistique des passages par zéro fournit un indice de fiabilité permettant de classer chaque région du plan temps-fréquence en deux catégories, selon qu'elle contient, ou non, une source harmonique et dominante.

A partir de formalisations théoriques et de simulations, un modèle complet de marquage du plan temps-fréquence est alors développé ; ce modèle est ensuite évalué en différents paradigmes d'interférences, incluant les paradigmes de doubles voyelles et de signaux bruités, puis sur des signaux à fortes variations prosodiques. Enfin, la plausibilité physiologique de ce modèle est discutée.

\*\*\*\*\*

**Titre de la thèse :**

**Contribution de l'approche Multi-Bandes à la reconnaissance automatique de la parole.**

**Auteur :**

**Cerisara Christophe**

bât. LORIA

Campus Scientifique - BP 239

54506 VANDOEUVRE-lès-NANCY CEDEX

tel : 03-83-59-20-86

email : cerisara@loria.fr

**Résumé :**

Le travail de recherche présenté dans ce mémoire réalise l'étude d'une nouvelle architecture pour les systèmes de reconnaissance automatique de la parole. Cette architecture est basée sur un découpage du spectre du signal en plusieurs zones fréquentielles, éventuellement recouvrantes, et sur un traitement indépendant de chacune de ces "bandes". Celles-ci sont ensuite recombinaison afin de fournir une réponse unique au problème de la reconnaissance. L'utilisation de cette méthode dite "Multi-Bandes" est à l'origine motivée par les travaux du psycho-acousticien H. Fletcher, travaux qui ont été récemment reconsidérés par J. B. Allen et qui aboutissent à un modèle de l'audition humaine proche de ce principe. L'application de celui-ci à une tâche de reconnaissance automatique de la parole est généralement motivée par sa robustesse aux bruits limités fréquentiellement. Nous montrons dans ce mémoire qu'il possède d'autres avantages, moins attendus, comme la résistance à tous les types de bruits stationnaires, et qu'il peut également surpasser le système de référence dans des environnements non bruités.

Ce mémoire commence par présenter les travaux existant dans le domaine du "Multi-Bandes", puis réalise une étude préliminaire du comportement de chacune des bandes. Ensuite, le problème du choix du module de recombinaison est posé, et plusieurs solutions sont proposées et testées. De même, nous montrons qu'il n'est pas possible d'utiliser l'algorithme classique de Viterbi en reconnaissance continue lorsque les bandes sont asynchrones, et différentes autres possibilités sont étudiées. Nous proposons ainsi deux algorithmes permettant de recombinaison les bandes soit en fin de phrase, soit après des segments temporels associés à des unités de parole, comme les phonèmes.

Nous élaborons ensuite une architecture possible pour un système Multi-Bandes, architecture caractérisée notamment par l'utilisation conjointe du spectre complet et des sous-bandes, ainsi que par l'apprentissage du système réalisé grâce au critère de Minimisation de l'Erreur de Classification. Les tests de ce système sont réalisés en milieu plus ou moins bruité, avec différents types de bruit, et pour plusieurs tâches de traitement de la parole. Nous montrons ainsi que, selon l'environnement considéré, certains composants du système Multi-Bandes doivent être modifiés, comme par exemple l'importance accordée au spectre complet. Finalement, le dernier chapitre de ce mémoire est consacré à un nouvel algorithme d'apprentissage qui permet d'approcher l'optimum global du système. Cet algorithme a également la propriété de modifier les modèles dans chaque bande afin qu'ils soient réellement représentatifs de l'information contenue dans celles-ci, ce qui n'est pas le cas des phonèmes qui sont classiquement utilisés.

**Mots-clefs :** Reconnaissance Automatique de la Parole, Multi-Bandes, Reconnaissance Robuste de la Parole, Combinaison de Classifieurs, Modèles Stochastiques, Apprentissage.

\*\*\*\*\*

**Étude de la fusion des données hétérogènes pour la reconnaissance automatique de la parole audiovisuelle**

**Thèse de doctorat de l'Université de Paris XI, spécialité électronique, capteurs et circuits intégrés, préparée au Laboratoire d'Informatique de l'Université du Maine et soutenue le 9 juillet 1999 par Alexandrina ROGOZAN**

L'utilisation de connaissances supplémentaires conjointement au signal de parole améliore les performances et la robustesse des systèmes de reconnaissance de la parole. L'utilisation de données sur la forme et le mouvement des lèvres du locuteur semble être une voie prometteuse pour la reconnaissance automatique surtout en milieux sonores bruités.

Notre recherche concerne l'intégration des informations acoustiques et visuelles dans un système de reconnaissance de la parole. Si cette intégration est fort séduisante, la problématique qu'elle soulève est loin d'être simple. Se pose la question du niveau d'intégration : est-ce celui de données ou de résultats ? Puis il y a les phénomènes de décalage temporel entre les réalisations auditive et visuelle d'un phonème. Ensuite intervient le problème d'adaptation des contributions des modalités acoustique et visuelle selon leur fiabilité relative. Enfin se pose la question de la pertinence de l'utilisation, pour le traitement de la parole visible, d'unités de décision spécifique, nommées visèmes.

Les systèmes audiovisuels développés au moyen des modèles de Markov cachés respectent successivement une fusion des données par identification directe (ID), une fusion des résultats après identification séparée (IS) et une fusion hybride de type ID+IS. Chaque modalité y est impliquée avec un poids différent adapté dynamiquement en fonction du rapport signal sur bruit et du contenu phonétique des phrases.

Nous avons testé les performances des systèmes sur un corpus monolocuteur de lettres connectées. Les tests réalisés pour différents niveaux de bruit ont montré une amélioration des performances due à la pondération adaptative des modalités acoustique et visuelle. Les meilleures performances sont obtenues avec le système ID+IS. Celles-ci peuvent être améliorées en utilisant un ensemble des visèmes adapté à la tâche de reconnaissance. Nous avons défini cet ensemble au moyen des cartes auto-organisantes de Kohonen. Pour renforcer le rôle de visèmes, nous les avons utilisés avec un apprentissage discriminant fondé sur réseaux de neurones.

Alexandrina Rogozan – Laboratoire d'Informatique de l'Université du Maine, 72085 Le Mans Cedex 9, Tél : +33 2 43 83 38 64 – Fax : + 33 2 43 83 38 68 – Mél : Alexandrina.Rogozan@lium.univ-lemans.fr

\*\*\*\*\*