

Budget Prévisionnel

Par cette opération de recherche nous souhaitons mettre en commun un ensemble de financement afin de recruter un post-doctorant et deux alternants pour accompagner les travaux du projet. Ces financements proviennent de différent contrat de prestation de service sur les thématiques de traitement automatique de la parole.

Recettes

Type	Référence	Montant
Contrat de prestation de service	NXP-1	30.000 €
Contrat de prestation de service	Zenidoc	10.000 €
Contrat de prestation de service	Semaxone	40.500 €
Contrat de prestation de service	ATOO	20.000 €
Contrat de prestation de service	NXP-2	58.000 €
Contrat de prestation de service	Jurispredis	20.000 €
Total		178.500 €

Dépenses 2019

Type	Durée	Montant
Post-doctorat	4 mois	17.000 €
Alternant	4 mois	5.000 €
Alternant	4 mois	5.000 €
Mission d'anonymisation et de pseudonymisation		6.700 €
Total		33.700 €

Dépenses 2020

Type	Durée	Montant
Post-doctorat	8 mois	33.000 €
Alternant	12 mois	15.000 €
Alternant	12 mois	15.000 €
Total		63.000 €

Dépenses 2021

Type	Durée	Montant
Alternant	8 mois	10.000 €
Alternant	8 mois	10.000 €
Total		20.000 €

CV Jean-François Bonastre

Directeur du Laboratoire Informatique d'Avignon

Professeur de Classe Exceptionnelle à l'Université d'Avignon (27^{ème} section)

Né le 4 Décembre 1967, Marié, deux enfants

Doctorat en Informatique de l'Université d'Avignon et des Pays de Vaucluse (1994)

“ Stratégie analytique orientée connaissances pour la caractérisation et la reconnaissance du locuteur ”

Habilitation à Diriger les Recherches (2000)

“ Reconnaissance du locuteur et approche statistique : description, limites et potentialités ”

Membre Junior de l'Institut Universitaire de France (promotion 2006)

Auditeur de la 26^{ème} session des Hautes Etudes Stratégiques de la Méditerranée, IHEDN/FMES (2015-2016)

Président du Conseil Départemental de Concertation du Département de Vaucluse (2014-2015)

Responsabilités administratives

Directeur du LIA (Mars 2016-)

Membre du conseil de gestion de la Fondation Pierre Bergé de l'université d'Avignon

Membre du conseil de direction du LABEX BLRI et de l'institut Convergences ILCB

Membre du comité de pilotage du Carnot “Cognition”

Administrateur Provisoire de l'Université d'Avignon (Août 2015-Décembre 2015)

Vice-Président (conseil d'administration) de l'Université d'Avignon (2008-2015) et **1er Vice-Président** (2012-2015)

Participation aux sociétés savantes

Président de l'International Speech Communication Association/ISCA (2011-2013) et Vice-Président (2007-2011)

Membre Senior de IEEE et du IEEE Speech and language Technical Committee (pendant 2 ans), de l'IEEE Biometric council Technical Committee (pendant 2 ans)

Président de l'Association Francophone de la Communication Parlée/AFCP (2000-2004) et membre du CA de l'AFCP (2000-2006 puis 2014-)

Président du Groupe Francophone de la Communication Parlée/GFCP de la Société Française d'Acoustique/SFA (1999-2000) et **Vice-Président du GFCP** (1997-1999)

Initiateur du Special Interest Group « Speaker and Language Characterization » (SIG/SPLC) de l'ISCA (1999).

Mise en place et co-animation du GT1 "Caractérisation du Locuteur et de la Langue" du GDR-PRC CHM (1995-1998)

Co-animateur du groupe “ Parole ” du GDR-PRC I3 jusqu'en décembre 2005.

Co-Organisateur des Interspeech Special Event “Speaker Comparison for Forensic and Investigative Applications” en 2015, 2016 et 2017

Activités d'expertise et recrutement

Membre du conseil scientifique du Centre de Recherche Informatique de Montréal (11/2015-)

Membre du conseil scientifique du programme européen SEE-ERA.NET PLUS (1 année)

Membre du comité du programme "FIT-IT", FFG (Agence de la recherche, Autriche, 2006/2007) et évaluateur pour le programme Espagnol « Consolider » (2010)

Membre de comités d'évaluation pour l'AERES (GIPSAIab) et le CNRS (Unité internationale de Hanoi : MICA)

Editeur Associé pour la revue *IEEE TSALP*, Editeur invité pour les revues *Speech Communication* et *Signal Processing*. Arbitre pour Pattern Recognition letters, JASA, JASP... Arbitre pour ICASSP, Interspeech, LREC, ASRU, SLT... Membre du comité scientifique du numéro spécial de la revue TAL "TAL et éthique" (2015)

Conférences invitées depuis 2010

- « l'Odyssée des sciences », Montréal, « La voix comme clef, la voix à la barre » », 12/05/2016
- **Spring School on Individual-centered approaches to speech processing, Schloss Dagstuhl, avril 6-9 2014: From automatic speaker verification to forensic phonetics**
- **Afeka speech conference, Tel-Aviv, juillet 2014, Keynote speaker: The Notion of Performance: A Paradigm Shift in Speaker Recognition and Voice Comparison?**
- Technopolice 2013, Paris, Comparaison de voix : notions de performance et de fiabilité
- Workshop on Advances in Speech Technologies, 23/06/2011 (IRCAM/Paris) : Speaker recognition: a new binary representation

Responsabilités dans le cadre de projets coopératifs

Responsable LIA pour les projets européens MTM, COST-275, BIOSECURE, HARTES, MOBIO, Eurostar BIOSPEAK, COST IC1206 et pour le projet ARC-LINKAGE avec le QUT/Brisbane (financement australien), le projet NuSpecs (financement italien)

Porteur de projet pour les projets nationaux BIO_MUL (ACI), ALIZE (Technolangue), MISTRAL (ANR), FABIOLÉ (ANR) et participant aux projets BIOBIMO (ANR), GAFFES (ANR), DesPho-APaDy (ANR)

Rôles dans l'organisation d'événements scientifiques

Organisateur du colloque RLA2C, Avignon 1998

Organisateur des Journées Scientifiques de l'IUF, Avignon 2007

Co-organisateur de JEP-TALN 2008 (Avignon) et Président du Comité de Programme des JEP 2002 et 2004.

Co-Organisateur MMUA 2006 (Toulouse)

Membre permanent du comité de programme des workshops «Speaker Odyssey»-01/04/06/08/10/12/14 et 2016 et Co-Organisateur de « Speaker Odyssey 2018 »

Area Chair ISCA/Interspeech, Speaker and Language Characterization and Recognition (2008, 2015, 2016, 2017)

Rôle dans le cadre de l'utilisation de l'identification vocale dans le milieu judiciaire :

Mise en place et animation du GT1 "Caractérisation du Locuteur et de la Langue" (GDR-PRC CHM, 1995). Sessions spéciales "forensic" Interspeech 2003 et 2015, RLA2C (1998), table ronde JEP 2000, session JEP 2012. Séance du « Café des Sciences d'Avignon » (en collaboration avec P Perrot de l'IRCGN). Multiples exposés dont : « Vio Voice 2015 », Afeka Speech conférence 2014, Schloss Dagstuhl 2014 et Technopolice 2013. Publication d'articles de «

positionnement» dans les revues Langue, Justice, J'essaime, articles dans **Eurospeech 2003**, JEP 2004, JEP2010, « **Signal Processing** » **2009, chapitre de livre en 2015.**

Témoignage dans des tribunaux à propos de la comparaison de voix, à l'occasion de différents procès (~15 et Intervention fréquente dans les médias (dont Radio Télévision Suisse, Sciences et Avenir, Sciences & Vie,...)

Prix et distinctions

Prix de la meilleure thèse AFCP (Juliette Kahn, 2012)

Prix du meilleur papier étudiant (Mitchell MacLaren Interspeech 2008)

Prix du meilleur papier étudiant (Nicolas Scheffer, Odyssey 2006)

Encadrement doctoral

Direction de 18 Thèses de Doctorat soutenues

Direction de deux thèses en cours, dont une soutenance prévue en Juin 2017

+50 participations à des jurys de Thèse ou HDR hors Université d'Avignon. Rapporteur pour 42 thèses ou HDR en France et à l'étranger (Univ. Vigo/Portugal, U. Swansea/UK, IDIAP/Suisse, UPC/Spain, U. East Finland, Finland)

Publications : <https://scholar.google.fr/citations?user=P1-Ar8MAAAAJ&hl=fr&oi=ao>

Mickael Rouvier

PhD in Computer Science

contact

258, Bd Romain Rolland
Batiment 16, Entrée B3
13009, Marseille, France

+33 6 81 59 74 07

mickael.rouvier@gmail.com
www.mickael-rouvier.com

languages

french mother tongue
english & italian fluency

education

- 2008–2011 **PhD** in Computers Science University of Avignon, France
Audiovisual content structuring for automatic summarization
In this thesis, we focus on extraction methods for video summary, based on audio analysis. We treat various scientific problems related to this objective : content extraction, document structuring, definition and estimation of objective function and algorithm extraction.
- 2006–2008 **Masters** in Computers Science University of Avignon, France
Spoken Term Detection
In this dissertation, we treat the problem of spoken term detection. We proposed a term spotting method based on phonetic filtering and request-driven decoding.
- 2005–2006 **Bachelors** in Mathematics and Computers University of Avignon, France
Specialization in Computers.

experience

- 2013–Now **Computer Science Laboratory of Marseille (LIF)** Marseille, France
Postdoc
Worked on speaker adaptation of Deep Neural Network for Automatic Speech Recognition, and on clarification dialogs for speech-to-speech translation (DARPA-BOLT).
- 2011–2013 **Computer Science Laboratory of Le Mans (LIUM)** Le Mans, France
Postdoc
Worked on low latency Automatic Speech Recognition combination and on speaker clustering methods for Speaker Diarization (French REPERE project).
- 2008–2011 **Computer Science Laboratory of Avignon (LIA)** Avignon, France
PhD Student
Worked on automatic speech summarization and on audio document structuring.

technical skills

- **Programming:** C/C++, Java, Ruby, Perl, Shell
- **System:** Linux/Unix, OSX, Windows
- **Architecture:** CVS, SVN, Git
- **Database:** MySQL
- **Scientific tools:** Alize, Kaldi, OpenCV, HTK, libSVM

conferences

- member of the organizing committee of MajecSTIC/RJCP 2012, Avignon
- member of the organizing committee of JEP/TALN 2010, Avignon

publications

articles in peer-reviewed journals

Audio-based video genre identification

Mickael Rouvier, Stanislas Oger, Georges Linares, Driss Matrouf, Bernard Merialdo, Yingbo Li
IEEE Transactions on Audio Speech and Language (TASL) (2014). 2014

Modeling Nuisance Variabilities with Factor Analysis for GMM-based Audio Pattern Classification

Driss Matrouf, Florian Verdet, Mickael Rouvier, Jean-François Bonastre, Georges Linares
Computer Speech and Language (CSL) (2010). 2010

Query driven strategy for onthefly term spotting in spontaneous speech

Mickael Rouvier, Georges Linares, Benjamin Lecouteur
European Association for Signal Processing (EURASIP) (2009). 2009

international peer-reviewed conferences

Multimodal embedding fusion for robust speaker role recognition in video broadcast

Mickael Rouvier, Sebastien Delecraz, Benoit Favre, Meriem Bendris, Frederic Bechet
Automatic Speech Recognition and Understanding (ASRU), 2015

Speaker Diarization Through Speaker Embeddings

Mickael Rouvier, Benoit Favre
European Signal Processing Conference (EUSIPCO), 2015

"Speech is silver, but silence is golden": improving speech-to-speech translation performance by slashing users input

Frederic Bechet, Mickael Rouvier, Benoit Favre
International Speech Communication Association (InterSpeech), 2015

Joint decoding of complementary utterances

Mickael Rouvier, Benoit Favre, Frederic Bechet
Speech Language and Technology (SLT), 2014

Speaker adaptation of DNN-based ASR with i-vectors: Does it actually adapt models to speakers ?

Mickael Rouvier, Benoit Favre
International Speech Communication Association (InterSpeech), 2014

Multimodal understanding for person recognition in video broadcasts

Frederic Bechet, Meriem Bendris, Delphine Charlet, Geraldine Damnati, Benoit Favre, Mickael Rouvier, Remi Auguste, Benjamin Bigot, Richard Dufour, Corinne Fredouille, Georges Linares, Gregory Senay, Pierre Tirilly, Jean Martinet
International Speech Communication Association (InterSpeech), 2014

Scene understanding for identifying persons in TV shows: beyond face authentication

Mickael Rouvier, Meriem Bendris, Benoit Favre, Delphine Charlet, Geraldine Damnati
International Workshop on Content-Based Multimedia Indexing (CBMI), 2014

Reranked aligners for interactive transcript correction

Benoit Favre, Mickael Rouvier, Frederic Bechet
International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2014

An Open-source State-of-the-art Toolbox for Broadcast News Diarization

Mickael Rouvier, Gregor Dupuy, Paul Gay, Elie Khoury, Teva Merlin, Sylvain Meignier
International Speech Communication Association (InterSpeech), 2013

Searching segments of interest in single story web-videos

Mickael Rouvier, Georges Linares, Benoit Favre, Bernard Merialdo
International Workshop on Image and Audio Analysis for Multimedia Interactive services (WIAMIS), 2013

LIUM ASR System for ETAPE French Evaluation Campaign: Experiments on System Combination Using Open-Source Recognizers

Fethi Bougares, Paul Deleglise, Yannick Esteve, Mickael Rouvier
Text, Speech and Dialogue (TSD), 2013

An investigation of single-pass ASR system combination for Spoken Language Understanding

Fethi Bougares, Mickael Rouvier, Nathalie Camelin, Paul Deleglise, Yannick Esteve
International Conference on Statistical Language and Speech Processing (SLSP), 2013

i-vectors and ILP clustering adapted to cross-show speaker diarization

Gregor Dupuy, Mickael Rouvier, Sylvain Meignier, Yannick Esteve
International Speech Communication Association (InterSpeech), 2012

Subspace Gaussian Mixture Models Based on Noise Compensation for Speech Recognition

Mohamed Bouallegue, Mickael Rouvier, Driss Matrouf, Georges Linares
International Speech Communication Association (InterSpeech), 2012

Low latency combination of parallelized single-pass LVCSR systems

Fethi Bougares, Mickael Rouvier, Yannick Esteve, Georges Linares
International Speech Communication Association (InterSpeech), 2012

A Global Optimization Framework For Speaker Diarization

Mickael Rouvier, Sylvain Meignier
Speaker Odyssey, 2012

Factor Analysis Based Session Variability Compensation for Automatic Speech Recognition

Mickael Rouvier, Mohamed Bouallegue, Driss Matrouf, Georges Linares
Automatic Speech Recognition and Understanding Workshop (ASRU), 2011

Subspace Gaussian Mixture Models for Vectorial HMM-states Representation

Mohamed Bouallegue, Mickael Rouvier, Driss Matrouf, Georges Linares
Automatic Speech Recognition and Understanding Workshop (ASRU), 2011

Static and Dynamic Video Summaries

Yingbo Li, Bernard Merialdo, Mickael Rouvier, Georges Linares
ACM Multimedia, 2011

Speaker Role Recognition using question detection and characterization

Thierry Bazillon, Benjamin Maza, Mickael Rouvier, Frederic Bechet, Alexis Nasr
International Speech Communication Association (InterSpeech), 2011

A Language identification inspired method for spontaneous speech detection

Mickael Rouvier, Richard Dufour, Georges Linares, Yannick Esteve
International Speech Communication Association (InterSpeech), 2010

On-the-fly video genre classification by combination of audio features

Mickael Rouvier, Georges Linares, Driss Matrouf
International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2010

Transcription-based video genre classification

Stanislas Oger, Mickael Rouvier, Georges Linares
International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2010

Robust Audio-based Classification of Video Genre

Mickael Rouvier, Georges Linares, Driss Matrouf
International Speech Communication Association (InterSpeech), 2009

Factor Analysis for Audio-based Video Genre Classification

Mickael Rouvier, Driss Matrouf, Georges Linares
International Speech Communication Association (InterSpeech), 2009

On-the-fly term spotting by phonetic filtering and request-driven decoding

Mickael Rouvier, Georges Linares, Benjamin Lecouteux
Speech Language and Technology (SLT), 2009

book in French

Le systeme du LIA pour la campagne DEFT'10 : datation et localisation d'articles de presse francophones

Stanislas Oger, Mickael Rouvier, Nathalie Camelin, Remy Kessler, Fabrice Lefevre, Juan-Manuel Torres-Moreno

Hermes, 2012

French peer-reviewed conferences

Identification de personnes dans des flux multimédia

Frederic Bechet, Meriem Bendris, Delphine Charlet, Geraldine Damnati, Benoit Favre, Mickael Rouvier, Remi Auguste, Benjamin Bigot, Richard Dufour, Corinne Fredouille, Georges Linares, Gregory Senay, Pierre Tirilly, Jean Martinet

Conférence en Recherche d'Information et Applications (CORIA), 2015

TALEP @ DEFT'15 : Le plus coool des systèmes d'analyse de sentiment

Mickael Rouvier, Benoit Favre, Balamurali Andiyakkal Rajendran

Traitement Automatique du Langage Naturel (TALN), 2015

Correction interactive de transcriptions de parole par fusion de phrases

Mickael Rouvier, Benoit Favre, Frederic Bechet

Journées d'Etudes sur la Parole (JEP), 2014

Segmentation et Regroupement en Locuteurs d'une collection de documents audio

Gregor Dupuy, Mickael Rouvier, Sylvain Meignier, Yannick Esteve

Journées d'Etudes sur la Parole (JEP), 2012

Nouvelle approche pour le regroupement des locuteurs dans des émissions radiophoniques et télévisuelles

Mickael Rouvier, Sylvain Meignier

Journées d'Etudes sur la Parole (JEP), 2012

Avances dans le domaine de la transcription automatique par decodage guide

Fethi Bougares, Yannick Esteve, Paul Deleglise, Mickael Rouvier, Georges Linares

Journées d'Etudes sur la Parole (JEP), 2012

Qui etes vous ? Categoriser les questions pour determiner le role des locuteurs dans des conversations orales

Thierry Bazillon, Benjamin Maza, Mickael Rouvier, Frederic Bechet, Alexis Nasr

Traitement Automatique du Langage Naturel (TALN), 2011

Systeme du LIA pour la campagne DEFT'10 : datation et localisation d'articles de presse francophones

Stanislas Oger, Mickael Rouvier, Nathalie Camelin, Remy Kessler, Fabrice Lefevre, Juan-Manuel Torres-Moreno

Defi Fouille de Textes (DEFT), 2010

Classification du genre video reposant sur des transcriptions automatiques

Stanislas Oger, Mickael Rouvier, Georges Linares

Traitement Automatique du Langage Naturel (TALN), 2010

Identification du genre video à la volée par combinaison de parametres acoustiques

Mickael Rouvier, Georges Linares, Driss Matrouf

Journées d'Etudes sur la Parole (JEP), 2010

Identification robuste du genre video par l'audio

Mickael Rouvier, Georges Linares, Driss Matrouf

Manifestation des Jeunes Chercheurs en Sciences et Technologies de l'Information et de la Communication (MajecSTIC), 2010

other publications

LIF @ TASS 2015 : The deep models are very cooooooIIIII for sentiment analysis

Mickael Rouvier, Benoit Favre

Technical Report, 2015

PERCOLATTE : a multimodal person discovery system in TV broadcast for the Medieval 2015 evaluation campaign

Meriem Bendris, Delphine Charlet, Benoit Favre, Gregory Senay, MinYoung Kim, Mickael Rouvier, Benoit Favre, Géraldine Damnati, Frederic Bechet

MediaEval, 2015

LIA @ MediaEval 2011 : Compact Representation of Heterogeneous Descriptors for Video Genre Classification

Mickael Rouvier, Georges Linares

MediaEval, 2011

Curriculum vitæ

Né le 10 janvier 1973 à Toulouse - Nationalité française

Professeur 1^{ère} classe à Avignon Université en section CNU 27 (Informatique)

En **délégation** à Avignon Université depuis septembre 2018 et définitivement rattaché à Avignon Université après mutation au 1^{er} septembre 2019

Parcours professionnel :

- ▶ 2019 - **Professeur 1^{ère} classe** à Avignon Université
- ▶ 2015 - **Professeur 1^{ère} classe** à Le Mans Université
- ▶ 2015-2018 - **Directeur de l'Institut d'Informatique Claude Chappe (Le Mans Université)**
- ▶ 2012-2016 - **Directeur du Laboratoire d'Informatique de l'Université du Mans**
- ▶ 2015 à août 2018 - **Directeur scientifique adjoint de la structure régionale RFI AtlanSTIC 2020** (structure régionale multi-établissement + acteurs industriels pour la recherche/formation/innovation en informatique en Pays de la Loire, Nantes - Angers - Le Mans)¹
- ▶ 2010-2012 - **Directeur du département d'informatique de l'UFR STS de l'Université du Mans**
- ▶ 2010 - **Professeur 2^{ème} classe** à l'Université du Mans
- ▶ 2003 - **Maître de conférences** à l'Université du Mans
- ▶ 2003 - **Post-doc** au LIA, Université d'Avignon et des Pays de Vaucluse
- ▶ 2002 - **Ingénieur de recherche**, France Télécom R&D, Lannion
- ▶ 2001 - 2002 - **ATER** au LIA, Université d'Avignon
- ▶ 1998 - 2001 - **Doctorant/Moniteur**, LIA, Université d'Avignon et des Pays de Vaucluse

Diplômes universitaires :

- ▶ 2009 - **Habilitation à Diriger les Recherches**, spécialité informatique - Université du Maine, Le Mans. Titre : « *Transcription automatique de la parole : contributions* ».
- ▶ 2002 - **Doctorat en informatique** (félicitations du jury), Université d'Avignon et des Pays de Vaucluse. Directeur de Thèse : Professeur Renato De Mori ; co-directeur : M. Frédéric Béchet
- ▶ 1998 - **DEA d'informatique** (mention bien), Université Aix-Marseille II, Faculté des Sciences de Luminy, Marseille
- ▶ 1997 - **Maîtrise d'informatique** (mention assez bien), Université Aix-Marseille II, Faculté des Sciences de Luminy, Marseille
- ▶ 1996 - **Licence d'informatique** (mention assez bien), Université Aix-Marseille II, Faculté des Sciences de Luminy, Marseille
- ▶ 1995 - **DEUG MASS** (option mathématiques fondamentales, mention assez bien), Université Aix-Marseille II, Faculté des Sciences de Luminy, Marseille

¹ <https://atlanstic2020.fr>

Avignon, le 2 juillet 2019

A l'attention de
Georges Linarès, Vice-Président Recherche, Avignon Université

Affaire suivie par
Jean-François BONASTRE

Téléphone
+33 0490843514

Courriel
jean-francois.bonastre@univ-
avignon.fr

Objet : Soutien du LIA à la demande d'opération pluriannuelle de
recherche « Les réseaux de neurones profonds utilisé dans le
cadre du traitement automatique de la parole »

Ref : DIRLIA_2019_022

Monsieur le Vice-Président,

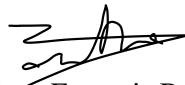
Trois enseignants-chercheurs du LIA, mes collègues Mickael Rouvier et Yannick Estève ainsi que moi-même, déposent conjointement une demande de création d'une opération pluriannuelle de recherche sur le thème des réseaux de neurones profonds utilisés dans le cadre du traitement automatique de la parole.

J'atteste par ce courrier que : 1/ ce thème prend beaucoup d'importance dans notre domaine et dans le cadre de la politique scientifique du LIA 2/ il a vocation à être récurrent pour notre laboratoire, sur un horizon de 10 ans.

De plus, le LIA a obtenu récemment des collaborations ou prestations de recherche avec différents industriels comme NXP, ATOO, Semaxone et Juripredis, ce qui montre le potentiel de ressources propres pour développer cette thématique.

Pour toutes ces raisons, je soutiens sans réserve cette demande de création.

Cordialement,



Jean-François Bonastre
Professeur, Université d'Avignon
Directeur du LIA

Descriptif scientifique

Les réseaux de neurones profonds utilisés dans le cadre du traitement automatique de la parole

I - Positionnement général

Depuis une dizaine d'années, on constate que les modèles neuronaux ont pris l'ascendant dans de nombreux domaines (tels que l'image, le texte et l'audio) et ceci grâce à l'accès à des quantités de données de plus en plus importantes, des nouveaux algorithmes plus performants issus de réseaux de neurones profonds et l'utilisation de GPU (carte graphiques) pour effectuer des calculs intensifs.

C'est aussi le cas dans les domaines d'activités de recherche en traitement automatique de la parole, sur lesquels le LIA travaille, où les modèles neuronaux connaissent aussi un grand succès : modélisation acoustique, modélisation du langage, traduction automatique, compréhension de la parole...

Même si les réseaux de neurones profonds sont utilisés dans plusieurs domaines, les problématiques de recherche liées aux réseaux de neurones profonds restent les mêmes. Ce projet de recherche vise à résoudre ces problématiques et à utiliser le traitement automatique de la parole comme cadre applicatif.

II - Objectifs et problématiques

- **Adaptation des réseaux de neurones profonds**

Problématique : De manière générale, l'adaptation des modèles permet de réduire l'écart entre le modèle initial et un corpus. Contrairement aux précédentes approches statistiques, les réseaux de neurones n'ont pas de structure claire et identifiable, mais ont plus de paramètres (conséquence des larges et profondes couches cachées). L'adaptation de ces nombreux paramètres à des données n'est pas simple.

Méthodologie : Des approches ont été proposées dans la littérature, permettant d'adapter les réseaux de neurones profonds en modifiant la couche d'entrée et la couche de sortie. Le but est d'essayer de comprendre la structure des réseaux de neurones profonds.

- **Augmentation de données**

Problématique : Cette approche devient incontournable, elle permet d'améliorer la robustesse et peut être combinée avec toute autre approche. Cela consiste à intégrer la (ou les) perturbation(s) dans les données d'apprentissage afin qu'elles soient prises en compte dans la décision par le réseau de neurones. Il n'existe pas une manière universelle pour bruite les données. Alors comment le faire et avec quel corpus ?

Méthodologie : Des « recettes » ont été proposées dans chaque domaine pour brouter les données du corpus. L'objectif serait de trouver une approche qui s'adapterait au corpus de développement afin de brouter les données avec les meilleurs paramètres.

III - Cadre applicatif

Nos efforts actuels visent à répondre à ces différents verrous scientifiques et à les appliquer à différents cadres :

- **Augmentation de données** : Un système de *wake word* consiste à déclencher un événement lorsqu'un mot ou une expression a été prononcé. C'est notamment le cas des assistants personnels de Siri ou Google lorsqu'on prononce les mots « Dis Siri » ou « Ok Google ». Lorsque cette expression a été prononcée, le système va déclencher un événement. Le système doit être assez robuste pour pouvoir détecter cette expression dans tous les environnements. Une approche possible est de brouter de manière artificielle le corpus pour le rendre le plus robuste possible. Cette problématique de recherche est traitée dans le projet de collaboration que le LIA a avec l'entreprise NXP.
- **Adaptation des réseaux de neurones profonds** : Un système de reconnaissance automatique de la parole consiste à transcrire de manière automatique sous forme de texte ce qui a été dit oralement. Le système est généralement composé de plusieurs modèles statistiques dont le modèle de langage qui consiste à donner une probabilité de suite de mots. Lorsque nous sommes dans une thématique, par exemple : nautique, nous donnerons plus de poids aux mots ou suites de mots dépendant de ce thème (bateau, navire, catamaran...). Le but est de pouvoir adapter le modèle de langage en fonction de la thématique. Cette problématique de recherche est traitée dans les projets de collaboration que le LIA a avec Zenidoc (adapter le modèle de langage à un domaine médical), ATOO (adapter le modèle de langage à un domaine de restauration) et Semaxone (adapter le modèle de langage à un domaine d'aviation).

IV - Équipe de recherche

Cette opération sera portée par : Mickael Rouvier (McF), Yannick Esteve (Pr) et Jean-François Bonastre (Pr). Elle pourra impliquer tous les chercheurs du LIA travaillant sur les problématiques du traitement de la parole, de l'intelligence artificielle et de l'apprentissage automatique.