

# Projet de thèse : Fiabilité en Biométrie Vocale

## 1. DESCRIPTIF

Ce projet s'inscrit dans le cadre du projet ANR Blanc 2012 « Fabiole ». Il concerne le domaine de l'authentification biométrique vocale et s'attache plus particulièrement à explorer l'apport des « connaissances phonétiques ».

Depuis près de 15 années, des systèmes de Reconnaissance Automatique du Locuteur (RAL) sont évalués par le NIST à travers les campagnes SRE. Ces évaluations montrent une progression importante des performances, amenant à envisager de nombreuses applications notamment dans le domaine judiciaire. Les conséquences que peuvent avoir de telles applications nécessitent de s'interroger sur la fiabilité des mesures d'évaluation effectuées. En effet, les critères actuellement employés pour mesurer la performance d'un système sont estimés globalement sur l'ensemble des tests effectués. Il s'agit de mesures de Taux d'Égale Erreur (EER) ou de Decision Cost Function (DCF). Ces mesures moyennes calculées sur un grand nombre de tests provenant de locuteurs différents ne tiennent qu'imparfaitement compte du cadre applicatif pour lequel une réponse doit être retournée pour un cas précis lié à des données impliquant seulement un ou deux locuteurs donnés. En particulier, ce mode d'évaluation ne prend en considération ni la différence entre un enregistrement et un locuteur, ni les différences entre locuteurs. Des travaux récents montrent clairement les limites des mesures de performance utilisées.

L'objet de ce projet de thèse est double. Dans un premier temps, il s'agit de caractériser et de mesurer l'influence des facteurs acoustiques et phonétiques amenant la variabilité interlocuteur. Il ne s'agit plus ici de mesurer une réussite globale en termes de reconnaissance du locuteur mais de préciser la part de variabilité interlocuteur expliquée par chacun des facteurs ainsi caractérisés. Dans un deuxième temps, ces éléments serviront à établir une mesure de confiance en comparaison de voix, basée uniquement sur les deux enregistrements de voix à comparer soit indépendamment de la technique utilisée par la suite.

Il s'agit ici d'identifier les informations caractéristiques du locuteur présentes dans chacun des deux enregistrements et de mesurer la cohérence de ces informations, tout en croisant ces informations avec la capacité des systèmes automatiques à traiter celles-ci.

Outre les apports en termes de connaissances sur les caractéristiques individuelles portées par la voix, cette thèse permettra, pour la comparaison de voix, de dépasser le cadre de la performance pour entrer dans le domaine de la fiabilité.

## 2. ENVIRONNEMENT ET ORGANISATION

Ce projet de thèse s'inscrit dans de la collaboration entre le LIA et le LIG qui a mené au co-encadrement de Juliette Kahn (dont la thèse a été soutenue fin décembre 2011), sur la notion de « parole et de locuteur ». Cette collaboration a permis de définir la méthodologie qui sera réutilisée dans ce projet. Elle a également démontré la pertinence du rapprochement entre les aspects « RAL » et « Sciences du langage » et de l'approche proposée ainsi que la faisabilité de celle-ci.

Le ou la doctorante bénéficiera de l'expérience du LIA en reconnaissance automatique du locuteur (plateforme ALIZE, participation continue à NIST-SRE depuis 1998...). De même, la thèse s'appuiera sur les compétences du LIG en ce qui concerne les éléments acoustiques et phonétiques. Enfin, le troisième partenaire du projet Fabiole, le LNE, amènera ses compétences en termes de protocoles d'évaluation et d'exploitation des données.

Le ou la doctorante sera basé principalement au LIA ou au LIG. Des séjours longs (6 mois) seront organisés dans le deuxième laboratoire pour bénéficier pleinement de l'apport pluridisciplinaire du sujet et de l'encadrement. Enfin, des séjours courts et fréquents (un jour à une semaine) seront organisés au LNE.

### 3. COMPÉTENCES SOUHAITÉES ET ACQUISES

Au vu de la nature fondamentalement pluridisciplinaire du sujet proposé, les candidatures attendues doivent montrer a minima une bonne connaissance du traitement automatique (et/ou de la classification automatique) ou des sciences du langage. Un intérêt démontré pour le deuxième domaine est également indispensable.

Le cas échéant, les encadrants s'attacheront à compléter les connaissances du doctorant en début de thèse, sur un des deux domaines considérés.

Durant tout le travail de thèse, les encadrants veilleront à l'équilibre entre les deux grandes disciplines, tant au niveau du travail à réaliser que de la valorisation de celui-ci, de manière à permettre au jeune docteur de faire valoir des compétences de haut niveau dans chacune des disciplines.

### 4. RÉFÉRENCES BIBLIOGRAPHIQUES

- (Bimbot, 2004) F. Bimbot, J. F. Bonastre, C. Fredouille, G. Gravier, I. Magrin-Chagnolleau, S. Meignier, T. Merlin, J. Ortega-Garcia, D. Petrovska-Delacretaz, et D. A. Reynolds, 2004. A tutorial on Text-Independent speaker verification. *Eurasip Journal on applied signal processing* 4, 430-451
- (Bonastre, 2003) J. Bonastre, F. Bimbot, L. Boë, J. Campbell, D. Reynolds, et I. Magrin-Chagnolleau, 2003. Person authentication by voice : A need for caution. Dans les actes de EUROSPEECH
- (Campbell, 2009) Campbell J.P, Shen W., Campbell W.M., Schwartz R., [J-F. Bonastre, D. Matrouf, "Forensic Speaker Recognition: A need for caution"](#), 2009, *IEEE Signal Processing Magazine*, March 2009, vol 26 pp 95-103
- (Doddington, 1998) G. Doddington, W. Liggett, A. Martin, M. Przybocki, et D. Reynolds, 1998. Sheep, goats, lambs and wolves : A statistical analysis of speaker performance in the NIST 1998 speaker recognition evaluation. Dans les actes de International Conference on Spoken Language Processing (ICSLP), Sydney, 1351-1354
- (Doddington, 2000) G. R. Doddington, M. A. Przybocki, A. F. Martin, et D. A. Reynolds, 2000. The NIST speaker recognition evaluation - overview, methodology, systems, results, perspective. *Speech Communication* 31(2-3), 225-254
- (Eriksson, 2006) A. Eriksson, 2006. Charlatanry and fraud—an increasing problem for forensic phonetics ? Dans les actes de Conference of the International Association for Forensic Phonetics and Acoustics (IAFPA), Göteborg, 10-11.

- (Fauve, 2007) Fauve, [D. Matrouf](#), D. Scheffer, [J-F. Bonastre](#), J.-F. Mason, J. S. D.  
 "[State-of-the-Art Performance in Text-Independent Speaker Verification Through Open-Source Software](#)", 2007, IEEE Transactions on Audio, Speech and Language Processing, Vol. 15, Issue 7, pp 1960-1968, September
- (Kahn, 2011a) J. Kahn, 2011, Parole de Locuteur : Performance et confiance en identification biométrique vocale. Thèse de Doctorat de l'Université d'Avignon et des Pays de Vaucluse
- (Kahn, 2011b) J. Kahn, N. Audibert, J-F Bonastre, S. Rossato, 2011, « Inter and intra-speaker variability: a formant analysis for french speakers and its implication on automatic speaker verification», International Conference in Phonetic Sciences 2011, Hong Kong, p
- (Kahn, 2010a) J. Kahn, N. Audibert, S. Rossato, J-F Bonastre , 2010, « Intra-speaker variability effects on Speaker Verification performance », Odyssey 2010, Brno, p
- (Kahn, 2010b) J. Kahn, S. Rossato, J-F Bonastre, 2010, « Beyond Doddington menagerie, a first step towards », International Conference on Acoustic, Speech and Signal Processing 2010, Dallas, p
- (Kahn, 2009) J. Kahn, S. Rossato, 2009, « Do Humans and Automatic systems use the same information to recognize voices? », Interspeech 2009, Brighton, p.2375-p.2378
- (Magrin-Chagnolleau) I. Magrin-Chagnolleau, J. Bonastre, et F. Bimbot, 1995. Effect of utterance duration and phonetic content on speaker identification using second-order statistical methods. Dans les actes de European Conference on Speech Communication and Technology (Eurospeech), Madrid, 337-340
- (Stoll, 2010) L. Stoll et G. Doddington, 2010. Hunting for wolves in speaker recognition. Dans les actes de ISCA-IEEE Speaker Odyssey, Brno, 159-162