**PROPOSITION SUJETS DE THESES**

**CONTRATS DOCTORAUX 2013-2016**

**Directeur de thèse** :MATROUF Driss

**Co-directeur** :

**Correspondant** :

Nom : MATROUF Prénom : Driss

Mail : driss.matrouf@univ-avignon.fr Téléphone : 0490843526

**Titre en français** :Reconnaissance du locuteur en milieu bruité

**Titre en anglais** : Speaker recognition in noisy environment

**Mots-clés** : speaker recognition, noise, i-vectors, UBM-GMM, HMM

**Co tutelle** : Non

Pays :

**Profil du candidat :** Titulaire d'un Master Recherche ou diplôme d'Ingénieur Grandes écoles avec des compétences en informatique, mathématiques, modélisation statistique et traitement du signal.

**Présentation détaillée en français :**

Nous avons atteint ces dernières années de très bonnes performances en reconnaissance du locuteur. Ces progrès ont été obtenus malgré la présence de la variabilité session. En effet, cette variabilité est prise en compte lors du scoring en utilisant une matrice de covariance modélisant cette dernière. Ce processus est effectué dans l'espace des i-vectors [1]. Le concept des i-vectors est devenu un standard en reconnaissance du locuteur.

Dans la dernière évaluation internationale NIST 2012, nous avons été confrontés à une nouvelle difficulté qui est le bruit additif [2], c'est à dire le bruit ambiant. La recherche pour réduire l'impact du bruit dans les systèmes de reconnaissance du locuteur est motivée en grande partie par le besoin d'appliquer les technologies de reconnaissance du locuteur sur des appareils portables ou sur l'Internet. Alors que les technologie promet un niveau supplémentaire de sécurité biométrique pour protéger l'utilisateur, la mise en œuvre pratique de ces systèmes doit faire face à de nombreux défis. Un des plus importants défis à surmonter est le bruit environnemental. En raison de la mobilité de ces systèmes, les sources de bruit peuvent être très variables dans le temps et potentiellement inconnus.

Nous proposons de travailler dans ce cadre : proposer des stratégies permettant de compenser l'effet du bruit additif. Ces stratégies peuvent intervenir à différents niveaux du processus de reconnaissance: au niveau du signal, au niveau des modèles acoustiques, au niveau des i-vectors et au niveau du scoring.....

- Débruitage des signaux
- Effet du bruit sur la VAD (Voice activity detection)
- Bruitage des modèles
- Intégration des caractéristiques statistiques du bruit dans la phase du scoring

Dans une deuxième partie du travail, nous proposons de nous mettre dans les meilleures conditions pour que le système soit le plus robuste au bruit. Par exemple, le choix de l'énoncé à prononcer par le locuteur peut avoir de l'influence sur les performances du système [3]. Faut-il avoir avoir le même énoncé pour tous les locuteurs, ou au contraire chaque locuteur se distingue des autres locuteur sur un ensemble bien précis d'unités acoustiques ? Dans ce dernier cas, il faut trouver une stratégie, qui permet de déterminer l'ensemble des unités acoustiques qui différencient le plus possible un locuteur (des autres locuteurs). D'autres stratégies de robustesse au bruit doivent être proposées et étudiées dans le cadre de cette thèse. Une des pistes à explorer est l'utilisation de la théorie des caractéristiques manquantes (*missing-feature theory*), qui a été utilisée dans le domaine du traitement de la parole [4][5][6].

Les systèmes de reconnaissance du locuteur de l'état de l'art sont fondamentalement basés sur l'utilisation de l'UBM (Universal Backgroud Model), il s'agit d'un modèle trop simple pour le traitement et la modélisation de la parole. Dans le cas de la reconnaissance en milieu bruité, la tâche devient plus complexe, il est donc légitime de se reposer la question sur l'adéquation de ce modèle pour cette tâche. Nous proposons d'adapter une approche utilisant des HMM (ou autre modèle) à cette tâche tout en profitant des avancées récemment proposées ( Factor analysis, I-vectors, ...).

**Présentation détaillée en anglais:**

Despite the presence of the session variability (convolutive noise), we reached in recent years a very good performance in speaker recognition. The session variability is taken into account during the scoring phase by using a covariance matrix modeling the session variability. This process is done in the i-vector space [1]. The concept of i-vectors has become a standard in speaker recognition.

In the latest NIST Speaker Recongnition Evaluation (NIST 2012), we faced a new problem which is the additive noise [2], ie the ambient noise. Research to reduce the impact of noise in speaker recognition systems is motivated mainly by the need to apply the technologies of speaker recognition on mobile devices or on the Internet. While the technology promises a new level of biometric security to protect the user, the practical implementation of these systems faces many

challenges. One of the greatest challenges to overcome is related to environmental noise. Because of the mobility of these systems, the noise sources can be highly variable in time and potentially unknown.

We propose to work in this context: speaker recognition in noisy envoronment. Different strategies to compensate the effect of additive noise have to be proposed. These strategies can perform at different levels of the recognition process: at the signal level, at the acoustic models level, at the i-vectors level and at the scoring level ....:

- Speech Enhancement
- Study of the effect of noise on VAD (Voice Activity Detection)
- Integration of noise in the acoustic models
- Use of the statistical noise characteristics during the scoring phase
- .....

In a second step we propose to optimize the use conditions of the system to be the most robust to noise. For example, are all sounds (phonemes or other) have the same discriminatory power for all the speakers, or on the contrary, each speaker has its own a subset of sounds which characterise him? If the answer is that the power of discrimination of phonemes depends on the speaker, so we have to use systems where each speaker have to produce a different utterance (from the other speaker). In this case, a detailed study on the choice of the most discriminating sounds for a given speaker have to be performed [3].

Another direction that we want to explore is the use of missing feature theory. This theory has often been used in the speech processing and especially in a noisy environment [4][5][6].

Finally, the speaker recognition systems of the state of the art are fundamentally based on the use of UBM (Universal Backgroud Model), it is too simple a model for processing and modeling speech. In the case of speaker recognition in a noisy environment, the task becomes more complex, so it is legitimate to ask the question about the adequacy of this model for this task. We propose to adapt an approach based on HMM (or other model) to this task while exploiting the recently proposed advanced (Factor analysis, I-vectors, ...).

**Encadrement :**

La thèse se fera au sein de l'équipe reconnaissance du locuteur, le thésard sera encadré principalement par Driss MATROUF. Mais des possibilités de collaboration avec les différents membres de l'équipe restent ouvertes et bienvenues.

**Contrat / Partenariat :**

**Thématique :**

**Domaine :**

**Objectif :**

Améliorer le robustesse des systèmes de reconnaissance du locuteur en milieu bruité

**Contexte :**

**Méthode :** modélisation statistique, traitement de signal, théorie de l'information, etc.

**Résultat attendu :**

**Références bibliographiques :**

**[1] Bousquet Pierre-Michel, Matrouf Driss and Bonastre Jean-François, «Intersession compensation and scoring methods in the i-vectors space for speaker recognition » Interspeech 2011, Florence.**

**[2] Miranti Indar Mandasari, Mitchell McLaren and David A. van Leeuwen,** « The Effect of noise on modern automatic speaker recognition systems » , ICASSP 2012.

**[3]** Anthony Larcher**,** Pierre-Michel Bousquet**,** Kong-Aik Lee**, Driss Matrouf,** Haizhou Li**,** Jean-François Bonastre**,** « I-vectors in the context of phonetically-constrained short utterances for speaker verification. » ICASSP 2012: 4773-4776.

**[4] M.P. Cooke, P.G. Green, L. Josifovski, and A. Vizinho,** « Robust ASR with unreliable data and minimal assumptions, » in Proc., Robust'99, 1999

**[5] M.P. Cooke, P.G. Green, L. Josifovski, and A. Vizinho,** « Robust Automatic Speech Recognition with missing and unreliable acoustic data, » Speech Communication,, 2000.

**[6] B. Raj, M.L. Seltzer, and R.M. Stern,** « Reconstruction of missing features for robust speech recognition, » Speech Communication, 2004.