PROPOSITION SUJETS DE THESES

CONTRATS DOCTORAUX 2014-2017

Directeur de thèse : Jean-François Bonastre

Correspondant :

Nom : Bonastre Prénom : Jean-François

Mail : jean-francois.bonastre@univ-avignon.fr

Titre en français :Un espace binaire de représentation des caractéristiques vocales individuelles

Mots-clés : Reconnaissance du locuteur, traitement automatique de la parole

Co tutelle : Non

Pays :France

Profil du candidat :Informatique, développement logiciel, statistiques

Présentation détaillée en français :

L'identification biométrique vocale, est un domaine suscitant un très fort intérêt dans plusieurs grands domaines d'application. Le travail de thèse proposé s'intéresse majoritairement aux applications commerciales, comme la sécurisation de transactions bancaires et la protection de la vie privée (accès à des données personnelles par téléphone ou sur internet). Ce secteur est en plein développement et de nombreux acteurs s'y intéressent (IBM, Nuance, Agnitio, Loquendo, Validsoft, Mobbeel, Geol Semantics...).

Ce projet de thèse s'inscrit dans la continuité des travaux menés au LIA en reconnaissance automatique du locuteur (RAL) et en caractérisation de voix. Le LIA est un acteur majeur au niveau international en RAL, il participe de manière continue aux différentes campagnes d'évaluation du domaine, dont les campagnes NIST-SRE (depuis 1998). Il a dans ce cadre développé une plateforme de reconnaissance du locuteur distribuée en logiciel libre, ALIZE, à travers différents projets (RNTL/ALIZE, ANR/MISTRAL et BIOBIMO, PCRD hArtes et MOBIO, Eureka BIOSPEAK). Cette plateforme est utilisée par plus de 50 laboratoires académiques et privés de par le monde (près de 120 utilisateurs inscrits, représentant plus de 80 institutions différentes).

Plus précisément, le LIA a développé une approche originale basée sur un espace binaire de représentation de la voix. Un individu est représenté par un vecteur binaire à l'intérieur de cet espace, un vecteur de grande à très grande dimension mais fortement parcimonieux (un vecteur binaire très majoritairement composé

de 0). Chaque coefficient du vecteur indique si une caractéristique donnée est présente ou non. A contrario des approches traditionnelles, cette approche permet de modéliser des éléments de la voix peu fréquents mais susceptibles de caractériser finement les locuteurs. De plus, elle permet d'expliciter les ressemblances et les différences relevées entre deux voix, en fonction de chacun des éléments présents ou non dans les deux vecteurs, chaque coefficient de l'espace de représentation binaire étant –comme nous l'avons vu-associé à une caractéristique précise. Ce point revêt une grande importance. En effet, la reconnaissance du locuteur suit actuellement le paradigme de la performance, celle ci étant mesurée globalement, à travers des protocoles standardisés. Malheureusement, il existe toujours des facteurs de variabilité peu adressés dans ces protocoles et déterminer le potentiel d'une approche pour une application pratique donnée reste un verrou important pour la diffusion applicative de la reconnaissance du locuteur. La possibilité, unique, offerte par notre approche de spécifier explicitement les éléments menant à la décision permet d'envisager un changement de paradigme, ouvrant la porte à la notion de fiabilité. Dans ce paradigme, chaque élément menant à la décision est étudié séparément et son domaine d'application est défini. Lorsqu'une application pratique est envisagée, cette description fine permet de prédire la performance à attendre ET la fiabilité de celle-ci, en fonction du recouvrement entre le domaine d'application de chacun des éléments décisionnels par rapport au domaine visé en pratique.

Outre cette caractéristique fondamentale, cette approche offre plusieurs avantages annexes, comme une représentation très compacte des données acoustiques et des processus de traitement peu coûteux, ce qui la désigne comme un candidat intéressant dans le cadre de systèmes d'authentification biométrique vocale pour des transactions électroniques, adressant des millions d'utilisateurs potentiels ou -à l'opposé- de systèmes embarqués, par exemple sur des smartphones.

Cette approche a fait l'objet de travaux préliminaires qui ont permis d'en valider le concept. Ces travaux ont mené à plusieurs publications et à un brevet protégeant le principe général.

Le travail de thèse proposé consiste à explorer les possibilités de la représentation binaire proposée. Le travail portera à la fois sur les aspects théoriques, liés à la théorie de l'information et à la théorie bayésienne de la décision et sur des aspects plus pratiques, liés aux spécificités algorithmiques de la représentation binaire.

Trois objectifs immédiats sont identifiés :

- L'approche proposée repose sur la notion de « spécificité ». Une spécificité est une zone de l'espace acoustique qui n'est activée que pour un sous groupe de personnes. Les spécificités sont apprises à partir d'un très large ensemble d'enregistrements de parole. Si, dans la preuve de concept, une première approche, plutôt empirique, a été développée sur la base des

outils statistiques disponibles, cet élément doit être revu de façon plus théorique (ce point peut être vu comme un problème de partitionnement d'espace dirigé par l'objectif -task-driven).

- L'approche proposée comporte deux phases de binarisation. L'une permet de projeter chaque vecteur acoustique d'entrée dans l'espace binaire quand la deuxième accumule les vecteurs binaires associés à un ensemble de vecteurs d'entrée pour résumer l'ensemble par un nouvel et unique vecteur binaire. Si les procédés en eux-même sont simples, ils comportent des méta-paramètres importants, actuellement fixés par des heuristiques. La théorie de l'information offre un cadre solide pour remplacer ces heuristiques. Le concept de l'espace binaire facilite grandement, de plus, cette mise en œuvre.

- De même, toutes les approches « état de l'art » intègrent un dispositif de réduction des variabilités nuisibles, crucial pour les performances. Là encore, une première étude a démontré que de tels procédés pouvaient être transposés dans notre espace binaire. Cependant, dans cette première ébauche, les avantages intrinsèques de l'approche proposée n'ont pas été étudiés ni mis en œuvre. Ce point représente un facteur d'innovation et d'amélioration important, tant en termes de performance que d'utilisation de ressource.

La validation des travaux se fera dans le cadre des évaluations internationales NIST-SRE.

Bibliographie :

P-M. Bousquet, J-F. Bonastre, Typicality extraction in a Speaker Binary Keys model, 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp 1713-1716

J-F. Bonastre, X. Anguera, G. H. Sierra, P-M. Bousquet,"Speaker modeling using local binary decisions", 2011, Interspeech 2011, Florence

X. Anguera, J-F. Bonastre, "Fast speaker diarization based on binary keys", 2011, ICASSP 2011, May 2011, Prague

J.-F. Bonastre (UAPV), X. Anguera (Telefonica), FR 10/57732 - "Procédé de classification de données biométriques", déposé le 24 Septembre 2010, PCT/FR2011/052151

P-M. Bousquet, D. Matrouf, J-F. Bonastre, "Intersession compensation and scoring methods in the i-vectors space for speaker recognition", 2011, Interspeech 2011, Florence