

Sujet de thèse : Un espace binaire de représentation des caractéristiques vocales individuelles

Proposé par J.F Bonastre, Professeur, LIA, Université d'Avignon

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.

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'explicitier 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. 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 gros systèmes d'identification du locuteur ou 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. Enfin, l'intérêt de l'espace de représentation binaire proposé au sein de l'approche classique dominante (iVectors) sera également étudié, en impliquant d'acquérir des compétences en modélisation statistique. La validation des travaux se fera dans le cadre des évaluations internationales NIST-SRE.

Contact :

jean-francois.bonastre@univ-avignon.fr

Bibliographie :

1. 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
2. J-F. Bonastre, X. Anguera, G. H. Sierra, P-M. Bousquet, "Speaker modeling using local binary decisions", 2011, Interspeech 2011, Florence
3. X. Anguera, J-F. Bonastre, "Fast speaker diarization based on binary keys", 2011, ICASSP 2011, May 2011, Prague
4. J.-F. Bonastre (UAPV), X. Anguera (Telefonica), FR 10/57732 - "Procédé de classification de données biométriques", déposé le 24 Septembre 2010, PCT/FR2011/052151
5. 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