Sujet de thèse

Authentification des personnes par la voix :

Un espace binaire de représentation des caractéristiques vocales individuelles

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

L'authentification des personnes par la voix est un domaine suscitant un très fort intérêt tant au niveau industriel que pour ses applications liées à la lutte contre la criminalité et le terrorisme. Au niveau des applications commerciales, le point focal se déplace vers le secteur de la protection des applications individuelles, comme les réseaux sociaux (protection des données privées mais aussi sécurité des systèmes d'information, au sens large), sans que la sécurisation de transactions réalisées à distance (téléphone ou internet) soit laissée pour compte. Du coté des applications liées à la lutte anti criminalité, si l'authentification vocale représente un besoin identifié depuis longtemps, la pression s'est considérablement renforcée ces dernières années, avec l'importance prise par les réseaux de téléphonie mobile.

Si l'authentification par la voix a enregistré des progrès très significatifs lors de la dernière décennie, avec des systèmes montrant des taux d'erreurs pouvant être de l'ordre d'un pourcent lors des campagnes d'évaluation de grande ampleur comme les campagnes SRE organisées par le NIST, peu de solutions sont déployées à ce jour. Plusieurs points expliquent ce paradoxe apparent. D'une part, l'évaluation de la performance lors des campagnes SRE montre des faiblesses, en s'attachant à une évaluation globale quand différents travaux ont montré l'importante variation que masque cette mesure moyenne. Hors, dans une application pratique –et particulièrement dans le cadre général de la lutte anti criminalité, le contexte n'est pas contrôlable et la moyenne n'est pas un estimateur fiable. D'autre part, les méthodes employées (comme l'approche iVector qui obtient les meilleurs résultats) sont basées sur le concept de l'apprentissage statistique. Malgré son intérêt largement démontré, ce concept montre deux défauts majeurs : un élément est jugé important lorsqu'il est fréquent dans une population, ce qui interdit de tenir compte d'une particularité réellement propre à un seul individu et la réponse d'un système est un nombre brut, les éléments ayant menés à ce nombre restent non explicités. Ce dernier point est un désavantage important dans un contexte judiciaire dans lequel il est primordial d'expliquer tout résultat.

Le projet de thèse décrit dans ce document vise à répondre aux limitations exposées ci-dessus. Il s'inscrit dans la continuité des travaux menés au LIA en reconnaissance automatique du locuteur (RAL) et en caractérisation de voix. Le LIA, Equipe d'accueil évaluée A+ et membre du LABEX *Brain and Language Research Institute*, 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 près de 150 laboratoires académiques et privés de par le monde. La thèse proposée s'inscrira également en complément du projet ANR Fabiole, dédié à la notion de fiabilité en reconnaissance du locuteur. Le ou la doctorante bénéficiera ainsi de la méthodologie d'évaluation de la fiabilité ainsi que des ressources associées (dont une base de données dédiée à ces questions).

Récemment, le LIA a proposé 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. Elle permet également 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.

Outre les deux caractéristiques fondamentales de cette approche explicitées ci-dessus et aptes à apporter une rupture réelle avec l'état de l'art, le concept proposé offre également plusieurs avantages annexes pouvant favoriser la mise en application. En premier, l'espace binaire de travail proposé est très compact et permet des processus de traitement peu couteux. Cette caractéristique peut s'avérer déterminante pour le passage à l'échelle dans des systèmes de suivi travaillant sur plusieurs milliers de locuteurs. La représentation binaire proposée implique la prise de multiples décisions locales, une pour chaque élément binaire de l'espace. En présence de bruits, seuls quelques éléments binaires seront perturbés et la dégradation est connue par avance (elle portera sur un nombre borné de bits et chaque dégradation locale est limitée à 1 changement d'état) alors que dans l'approche classique un bruit limité dans le temps peut avoir une influence drastique sur le résultat.

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. Des travaux plus récents ont montré que l'approche propose pouvait tirer profit des dernières avancées obtenues en termes de prise en compte de la « variabilité session » et ainsi proposer un niveau de performance comparable à l'état de l'art.

Le travail de thèse proposé consiste à développer l'approche par représentation binaire. Si quelques solutions ont déjà été proposées pour démontrer la faisabilité de l'approche, le modèle théorique est encore à développer ainsi que les algorithmes d'optimisation correspondant. En particulier, il sera primordial de tenir compte du « feedback », cette possibilité d'expliciter de manière détaillée les décisions du système en listant chacun des éléments ayant menés à la décision. 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 tout en demandant des connaissances en phonétique et en acoustique, 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. La validation des travaux se fera dans le cadre du projet Fabiole ainsi qu'à travers une participation continue aux campagnes d'évaluations nationales et internationales.

Contact :

jean-francois.bonastre@univ-avignon.fr

Bibliographie :

1. G Hernandez Sierra, J. Calvo Jose, J.F. Bonastre, Session compensation using binary speech representation for speaker recognition, Pattern Recognition Letters, 2014
2. 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
3. J-F. Bonastre, X. Anguera, G. H. Sierra, P-M. Bousquet,"Speaker modeling using local binary decisions", 2011, Interspeech 2011, Florence
4. X. Anguera, J-F. Bonastre, "Fast speaker diarization based on binary keys", 2011, ICASSP 2011, May 2011, Prague
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
6. J. Kahn, N. Aubibert, S. Rossato, J.F. Bonastre, SPEAKER VERIFICATION BY INEXPERIENCED AND EXPERIENCED LISTENERS VS. SPEAKER VERIFICATION SYSTEM, ICASSP 2011, May 2011, Prague, 2011
7. J. Kahn, S. Rossato, J.F. Bonastre, Beyond Doddington menagerie, a first step towards, ICASSP, Dallas, 2010
8. J.-F. Bonastre (UAPV), X. Anguera (Telefonica), FR 10/57732 - "Procédé de classification de données biométriques", déposé le 24 Septembre 2010, PCT/FR2011/052151
9. N. Dehak, P. Kenny, R. Dehak, P. Dumouchel, P. Ouellet, Front-End Factor Analysis for Speaker Verification,", IEEE Transactions on Audio, Speech, and Language Processing, vol. 19, no. 4, pp.788–798, 2011
10. J. F Bonastre, F. Wils, S. Meignier, ALIZE, a free toolkit for speaker recognition. In *ICASSP 2005,* pp. 737-740, 2005
11. J.F. Bonastre, N. Scheffer, D. Matrouf, C. Fredouille and al., ALIZE/spkdet: a state-of-the-art open source software for speaker recognition. In *Odyssey* (p. 20), 2008.
12. A. Larcher, J.F. Bonastre, B.G. Fauve, K.A. Lee, C. Lévy, H. Li and al., ALIZE 3.0-open source toolkit for state-of-the-art speaker recognition. In *INTERSPEECH* 2013.