

PROPOSITION SUJETS DE THESES

CONTRATS DOCTORAUX 2020-2023

Appel ciblé (merci de cocher la case correspondante):

Contrat doctoral ministériel ED 536

Contrat doctoral ministériel ED 537

Directeur de thèse : Jean-François Bonastre

Co-directeur éventuel :

Co-encadrant éventuel :

Titre en français : Estimation profonde, conjointe, explicite et interprétable des caractéristiques individuelles portées par la voix

Titre en anglais : Deep joint and explainable estimation of speaker specific information

Mots-clés : AI, speech processing, deep learning, human-machine vocal interactions

Co tutelle : XXX - Non **Pays :**

Opportunités de mobilité à l'international du doctorant dans le cadre de sa thèse : oui

Profil du candidat :

Le candidat doit avoir un master en informatique, en traitement du signal, en mathématiques appliquées ou dans une spécialité proche, avec une composante obligatoire sur les méthodes d'apprentissage automatique. La bourse de thèse fera l'objet d'un concours au sein de l'Ecole Doctorale 536 de l'université d'Avignon, avec une audition du candidat le cas échéant (d'autres financements/modalités de thèse sont envisageables).

Date limite 15 Mai 2020. La candidature est composée a minima de : CV, lettre de motivation, relevés de notes, références des professeurs.

Pour tout renseignement et pour postuler merci d'envoyer un mail à J.F. Bonastre (jean-francois.bonastre@univ-avignon.fr).

Présentation détaillée du sujet : Estimation profonde, conjointe, explicite et interprétable des caractéristiques individuelles portées par la voix

Domaine / Thématique : IA, Traitement automatique de la parole, Reconnaissance du locuteur, identification vocale, apprentissage profond, réseaux de neurones

Objectif :

Cette thèse vise à construire un système d'extraction profond des caractéristiques individuelles portées par la voix qui soit par construction explicite et interprétable.

Il s'agit de déterminer un jeu de caractéristiques de la voix, chacune étant propre à un sous groupe de la population. La reconnaissance vocale d'une personne se fera de manière simple, comme dans l'identification par l'ADN : pour chaque caractéristique estimée présente dans l'échantillon de voix considéré, on accumulera la probabilité de présence de cette caractéristique dans l'échantillon vocal avec la typicalité de cette caractéristique (soit le pourcentage de personnes dans la population présentant cette caractéristique).

La nature explicite de l'approche proposée vient de la décomposition de la décision finale par caractéristique et de la disponibilité, pour chaque caractéristique utilisée de sa probabilité de présence et de sa typicalité.

La méthode proposée devra également être interprétable : chaque caractéristique devra être décrite explicitement de façon compréhensible pour un opérateur non expert.

La méthode proposée utilise l'apprentissage profond pour être à même de gérer au mieux les différentes variabilités : type de parole, bruits, canaux de transmission, etc. De plus, la méthode estimera de manière conjointe la présence des caractéristiques de manière à ce que chacune profite de la somme des connaissances extraites par les réseaux de neurones employés.

Contexte et enjeux :

La reconnaissance des individus par la voix est un domaine en fort développement et avec des répercussions sociétales importantes. De très nombreuses applications sont proposées, dans le cadre bancaire, des assistants vocaux ou de l'IOT par exemple. L'identification par la voix est également utilisée dans le cadre des applications de sécurité, privée ou nationale. Enfin, elle touche aussi le domaine judiciaire, avec l'expertise en comparaison de voix.

Si les systèmes de reconnaissance du locuteur font preuve d'un niveau de performance très élevés lors d'évaluation scientifiques, ils souffrent toujours de quelques défauts. En premier, comme tout système d'apprentissage automatique, ils peuvent admettre des biais d'apprentissage qui amènent à des décisions non appropriées (par exemple, la faible présence d'un accent régional dans la base d'apprentissage peut pousser le système à confondre caractéristique individuelle et caractéristique régionale). De plus, les systèmes fonctionnent

sous forme de boîte noire : ils retournent un score numérique en réponse à un stimuli et cela dans toutes circonstances, même si l'enregistrement sonore ne contient que peu d'information caractéristique du locuteur. Enfin, les scores proposés par les systèmes de reconnaissance du locuteur n'ont pas de signification en tant que tel. Pour prendre une décision, il faut encore normaliser, ou « calibrer » le score, pour tenir compte du contexte applicatif et des conditions locales d'utilisation. Un défaut de cette calibration, due par exemple à des conditions pas encore rencontrées peut donner au système un comportement erratique.

Cette thèse permettra à la fois de développer des systèmes de reconnaissance du locuteur efficaces et de protéger les utilisateurs, par une illustration explicite des caractéristiques extraites. Elle facilitera également les régulations nécessaires pour éviter les discriminations, en permettant au législateur d'interdire l'usage de certaines caractéristiques, comme cela est fait pour l'ADN.

Le LIA est un acteur reconnu pour ses compétences en reconnaissance du locuteur, comme son rôle dans les projets ANR VoxCrim, ROBOVOX ou VoicePersonae le montre. Il possède toutes les connaissances et les ressources nécessaires (logiciels, bases de données et calculateurs) pour mener à bien ce projet. Le LIA est partie prenante de plusieurs collaborations internationales autour de ce sujet, ce qui facilitera les échanges et séjours durant la thèse et après.