



Offre n°2024-07504

Doctorant F/H Synthèse de la parole pour l'alsacien et les langues de France

Type de contrat : CDD

Niveau de diplôme exigé : Bac + 5 ou équivalent

Fonction : Doctorant

Contexte et atouts du poste

Cette thèse se place dans le cadre du Défi Inria COLaF "Corpus et Outils pour les Langues de France", dont l'objectif est de créer des corpus, des modèles et des logiciels ouverts et inclusifs pour les langues de France. Cela inclut les langues régionales (alsacien, breton, corse, occitan, picard, etc.) et d'outre-mer (créoles, langues polynésiennes, langues kanakes, mahorais, etc.) et les langues d'immigration non-territoriales (arabe dialectal, arménien occidental, berbère, judéo-espagnol, romani, yiddish).

Le/la doctorant.e sera co-encadré.e par [Vincent Colotte](#), [Pascale Erhart](#) et [Emmanuel Vincent](#). Il/elle bénéficiera de l'expertise de l'équipe [Multispeech](#) en traitement de la parole et celle de [LiLPa](#) en dialectologie, en phonétique de corpus et en TAL. Il/elle collaborera avec les ingénieurs chargés de la création et la distribution des corpus et des briques logicielles et avec les autres partenaires du projet.

Mission confiée

Les technologies linguistiques sont clés pour la protection, la valorisation et l'enseignement des langues régionales et d'outre-mer et pour l'inclusion des locuteurs de langues d'immigration non-territoriales. Ces langues restent cependant largement ignorées des fournisseurs de technologies linguistiques [1]. Cela est vrai en particulier des systèmes de synthèse vocale, qui sont classiquement appris sur un jeu de données de haute qualité enregistré en studio par un petit nombre d'acteurs professionnels. Cette méthode induit un coût élevé pour chaque langue et limite le nombre de voix et leur expressivité.

L'objectif de la thèse est de concevoir une approche générale de développement de systèmes de synthèse vocale multi-voix pour ces langues peu dotées à partir d'archives sonores existantes (radios, télévisions, web, etc.). Il s'agit d'une approche radicalement différente de l'état de l'art, qui pose deux difficultés : ces archives sont de qualité variable et pour l'essentiel non transcrites sous forme textuelle. Pour résoudre ces difficultés, nous nous appuyerons sur la possibilité d'apprendre un système de synthèse vocale de haute qualité à partir d'un jeu de données de qualité variable [2] et sur l'émergence de systèmes de reconnaissance de la parole *few-shot* [3] permettant d'envisager la transcription automatique des données.

L'approche développée sera validée pour l'alsacien, qui est la deuxième langue régionale parlée en France en nombre de locuteurs tout en restant une langue sous-dotée en termes de données [4]. Elle sera ensuite étendue à une ou deux autres langues de France, selon les compétences et les souhaits du candidat. Le travail de recherche s'appuiera sur les jeux de données collectés par les ingénieurs du Défi COLaF.

[1] DGLFLF, Rapport au Parlement sur la langue française 2023, <https://www.culture.gouv.fr/Media/Presse/Rapport-au-Parlement-sur-la-langue-francaise-2023>

[2] S. Ogun, V. Colotte, E. Vincent, "Can we use Common Voice to train a Multi-Speaker TTS system?", in 2022 IEEE Spoken Language Technology Workshop (SLT), 2023, pp. 900-905.

[3] A. Radford, J.W. Kim, T. Xu, G. Brockman, C. McLeavey, I. Sutskever, "Robust speech recognition via large-scale weak supervision", in 40th International Conference on Machine Learning, 2023, pp. 28492-28518.

[4] D. Bernhard, A.-L. Ligozat, M. Bras, F. Martin, M. Vergez-Couret, P. Erhart, J. Sibille, A. Todirascu, P. Boula de Mareüil, D. Huck, "Collecting and annotating corpora for three under-resourced languages of France: Methodological issues", *Language Documentation & Conservation*, 2021, 15, pp.316-357.

Principales activités

Outre la faible quantité de données et potentiellement le faible nombre de locuteurs disponibles, i) toutes les langues régionales ne sont pas écrites ou ne le sont pas de façon standardisée ii) les enregistrements n'ont pas toujours été réalisés avec une qualité adéquate à la synthèse vocale iii) la transcription textuelle est souvent indisponible ou diffère des mots prononcés (erreurs de transcription, sous-titres, etc.). Il s'agira donc i) de concevoir une méthodologie de choix et de préparation des données, qui pourra s'appuyer sur l'estimation automatique de la qualité du signal [2] et de la transcription [5], sur des méthodes de transcription et de correction semi-automatiques et/ou sur l'apprentissage actif, ii) de concevoir une méthode de synthèse vocale multi-voix capable d'exploiter ces données, qui pourra

s'appuyer sur la proximité phonétique et/ou morphologique entre les langues ciblées et des langues proches bien dotées (français, allemand, etc.) [6], ainsi que sur des ressources uniquement textuelles [7], iii) de la coupler avec des approches de transfert de style pour l'expression des émotions [8].

[5] K. Fan, J. Wang, B. Li, S. Zhang, B. Chen, N. Ge, Z. Yan, "Neural zero-inflated quality estimation model for automatic speech recognition system", in Interspeech, 2020, pp. 606-610.

[6] Z. Cai, Y. Yang, M. Li, "Cross-lingual multi-speaker speech synthesis with limited bilingual training data", Computer Speech and Language, 2023, 77, pp. 101427.

[7] N. San, M. Bartelds, B. Billings, E. de Falco, H. Feriza, J. Safri, W. Sahrozi, B. Foley, B. McDonnell, D. Jurafsky, "Leveraging supplementary text data to kick-start automatic speech recognition system development with limited transcriptions", in 6th Workshop on Computational Methods for Endangered Languages, 2023, pp. 1-6.

[8] A. Kulkarni, V. Colotte, D. Juvet, "Analysis of expressivity transfer in non-autoregressive end-to-end multispeaker TTS systems", in Interspeech, 2022, pp. 4581-4585.

Compétences

Master en traitement de la parole, TAL, machine learning, linguistique informatique ou dans un domaine lié.

Solides compétences en programmation Python/Pytorch.

Une expérience préalable en traitement de la parole ou en TAL sera un atout.

La connaissance d'une langue régionale, d'outre-mer ou non-territoriale de France est un plus.

Avantages

- Restauration subventionnée
- Transports publics remboursés partiellement
- Congés: 7 semaines de congés annuels + 10 jours de RTT (base temps plein) + possibilité d'autorisations d'absence exceptionnelle (ex : enfants malades, déménagement)
- Possibilité de télétravail (après 6 mois d'ancienneté) et aménagement du temps de travail
- Équipements professionnels à disposition (visioconférence, prêts de matériels informatiques, etc.)
- Prestations sociales, culturelles et sportives (Association de gestion des œuvres sociales d'Inria)
- Accès à la formation professionnelle
- Sécurité sociale

Rémunération

2100 € brut/mois (la 1ère année)

Informations générales

- **Thème/Domaine** : Langue, parole et audio
- **Ville** : Villers lès Nancy
- **Centre Inria** : [Centre Inria de l'Université de Lorraine](#)
- **Date de prise de fonction souhaitée** : 2024-10-01
- **Durée de contrat** : 3 ans
- **Date limite pour postuler** : 2024-05-19

Contacts

- **Équipe Inria** : [MULTISPEECH](#)
- **Directeur de thèse** :
Vincent Emmanuel / emmanuel.vincent@inria.fr

A propos d'Inria

Inria est l'institut national de recherche dédié aux sciences et technologies du numérique. Il emploie 2600 personnes. Ses 215 équipes-projets agiles, en général communes avec des partenaires académiques, impliquent plus de 3900 scientifiques pour relever les défis du numérique, souvent à l'interface d'autres disciplines. L'institut fait appel à de nombreux talents dans plus d'une quarantaine de métiers différents. 900 personnels d'appui à la recherche et à l'innovation contribuent à faire émerger et grandir des projets scientifiques ou entrepreneuriaux qui impactent le monde. Inria travaille avec de nombreuses entreprises et a accompagné la création de plus de 200 start-up. L'institut s'efforce ainsi de répondre aux enjeux de la transformation numérique de la science, de la société et de l'économie.

Attention: Les candidatures doivent être déposées en ligne sur le site Inria. Le traitement des candidatures adressées par d'autres canaux n'est pas garanti.

Consignes pour postuler

Sécurité défense :

Ce poste est susceptible d'être affecté dans une zone à régime restrictif (ZRR), telle que définie dans le décret n°2011-1425 relatif à la protection du potentiel scientifique et technique de la nation (PPST).

L'autorisation d'accès à une zone est délivrée par le chef d'établissement, après avis ministériel favorable,

tel que défini dans l'arrêté du 03 juillet 2012, relatif à la PPST. Un avis ministériel défavorable pour un poste affecté dans une ZRR aurait pour conséquence l'annulation du recrutement.

Politique de recrutement :

Dans le cadre de sa politique diversité, tous les postes Inria sont accessibles aux personnes en situation de handicap.