Offre n°2024-08319

Doctorant F/H Synthèse de la parole multilingue appliquée aux langues régionales

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

La synthèse vocale est une technologie-clé pour la valorisation des langues régionales et d'immigration. Ces langues restent cependant largement ignorées des fournisseurs de technologies linguistiques [1], qui entraînent classiquement les systèmes de synthèse vocale sur des jeux de données monolingues de haute qualité enregistrés 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 un système de synthèse vocale multilingue et multi-voix applicable aux langues régionales. Parmi les systèmes de synthèse vocale multilingue existants, IMS Toucan [2] est le seul à couvrir plus de 7000 langues. Il s'appuie sur le phonétiseur multilingue transphone [3], l'encodeur articulatoire PanPhon [4], le synthétiseur FastSpeech 2 [5] conditionné sur des plongements de locuteur et de langue et le vocodeur HiFi-GAN [6] entraînés sur un corpus de 18000 heures de parole en 462 langues. Deux défis subsistent: réduire le caractère haché de la voix et les erreurs de phonétisation, tous deux plus prononcés pour les langues peu dotées non vues à l'apprentissage.

[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] F. Lux, S. Meyer, L. Behringer, F. Zalkow, P. Do, M. Coler, E.A.P. Habets, N.T. Vu, "Meta learning text-to-speech synthesis in over 7000 languages", in *Interspeech*, 2024, pp.4958-4962.
[3] X. Li, F. Metze, D. Mortensen, S. Watanabe, and A. Black, "Zero-shot learning for grapheme to phoneme conversion with language ensemble", in *Findings of ACL*, 2022, pp.2106-2115.
[4] D.R. Mortensen, P. Littell, A. Bharadwaj, K. Goyal, C. Dyer, L.S. Levin, "PanPhon: A resource for mapping IPA segments to articulatory feature vectors", in *26th International Conference on Computational Linguistics (COLING)*, 2016, pp.3475-3484.
[5] Y. Ren, C. Hu, X. Tan, T. Qin, S. Zhao et al., "FastSpeech 2: Fast and high-quality end-to-end text to speech", in *9th International Conference on Learning Representations (ICLR)*, 2021.
[6] J. Kong, J. Kim, J. Bae, "HiFi-GAN: Generative adversarial networks for efficient and high fidelity speech synthesis", in *NeurIPS*, 2020, pp.17022-17033.

Principales activités

Pour réduire le caractère haché, nous exploiterons les enregistrements vocaux disponibles pour les langues régionales et d'immigration visées et pour d'autres langues proches sur le plan phonétique et/ou morphologique. Ces enregistrements issus d'archives ouvertes ou privées (radios, télévisions, web, etc.) n'ont pas toujours été réalisés et transcrits avec une qualité adaptée à la synthèse vocale. Nous nous appuierons sur des systèmes d'estimation de qualité sonore [7] et de transcription [8] pour sélectionner automatiquement les données de haute qualité.

Pour améliorer la phonétisation, nous exploiterons en sus de ces données vocales les connaissances phonologiques et phonétiques disponibles, avec une attention particulière à l'alternance codique et à la variabilité des prononciations. Une méthode d'apprentissage actif permettant la correction itérative des prononciations pourra être évaluée.

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 [9], et étendue à d'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.

[7] 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.
[8] 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.
[9] 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.

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 :** 2025-01-01
- **Durée de contrat :** 3 ans
- **Date limite pour postuler :** 2024-12-06

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.