Offer #2024-07504

PhD Position F/M Speech synthesis for Alsatian and the languages of France

Contract type : Fixed-term contract

Level of qualifications required : Graduate degree or equivalent

Fonction : PhD Position

Context

This PhD is part of the Inria COLaF Challenge "Corpora and Tools for the Languages of France", which aims to create open, inclusive corpora, models and software for the languages of France. These include regional languages (Alsatian, Breton, Corsican, Occitan, Picard, etc.), overseas languages (Creoles, Polynesian, Kanak, Mahorese languages, etc.), and non-territorial immigrant languages (dialectal Arabic, Western Armenian, Berber, Judeo-Spanish, Romani, Yiddish).

The PhD student will be co-supervised by Vincent Colotte, Pascale Erhart, and Emmanuel Vincent. They will benefit from the expertise of the Multispeech team in speech processing and that of LiLPa in dialectology, corpus phonetics and NLP. They will collaborate with the engineers responsible for the creation and distribution of corpora and software building blocks and with other project partners.

Assignment

Linguistic technologies are key for the protection, promotion and teaching of regional and overseas languages and for the inclusion of speakers of non-territorial immigrant languages. However, these languages remain largely ignored by language technology providers [1]. This is particularly true of text-to-speech (TTS) systems, which are typically trained on high-quality datasets recorded in the studio by a few professional actors. This method induces a high cost for each language and limits the number of voices and their expressiveness.

The objective of the PhD is to design a general approach to the development of multi-speaker TTS systems for these under-resourced languages based on existing sound archives (radio, television, web, etc.). This is a radically different approach from the state of the art, which poses two difficulties: these archives are of variable quality and mostly untranscribed. To address these difficulties, we will rely on the possibility of learning a high-quality TTS system from a variable-quality dataset [2] and on the emergence of few-shot speech recognition systems [3] allowing automatic transcription of data.

The developed approach will be validated for Alsatian, which is the second regional language spoken in France in terms of number of speakers while remaining an under-resourced language in terms of data [4]. It will then be extended to one or two other languages of France, depending on the skills and wishes of the candidate. The research work will be based on the datasets collected by the engineers of the COLaF Challenge.

[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.

Main activities

In addition to the small amount of data and potentially the low number of speakers available, i) not all regional languages are written or admit a standard writing, ii) speech has not always been recorded with sufficient quality for TTS, iii) the text transcript is often unavailable or differs from the spoken words (transcription errors, subtitles, etc.). Work will therefore include i) designing a methodology for choosing and preparing data, which can be based on the automatic estimation of the quality of the signal [2] and its transcription [5], on semi-automatic transcription and correction methods and/or on active learning, ii) designing a multi-speaker TTS method capable of exploiting this data, which can rely on the phonetic and/or morphological proximity between languages targeted and similar well-resourced languages (French, German, etc.) [6], as well as on text-only resources [7], iii) coupling it with style transfer approaches for the expression of emotions [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. Jouvet, "Analysis of expressivity transfer in non-autoregressive end-to-end multispeaker TTS systems", in Interspeech, 2022, pp. 4581-4585.

Skills

MSc degree in speech processing, NLP, machine learning, computational linguistics, or in a related field.
Strong programming skills in Python/Pytorch.
Prior experience with speech processing or NLP is an asset.
Knowledge of a French regional, overseas or non-territorial language is a plus.

Benefits package

- 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

Remuneration

2100 € brut/mois (la 1ère année)

General Information

- **Theme/Domain** : Language, Speech and Audio
- **Town/city** : Villers lès Nancy
- **Inria Center** : Centre Inria de l'Université de Lorraine
- **Starting date** : 2024-10-01
- **Duration of contract** : 3 years
- **Deadline to apply** : 2024-05-19

Contacts

- **Inria Team** : MULTISPEECH
- **PhD Supervisor** :
  Vincent Emmanuel / emmanuel.vincent@inria.fr

About Inria

Inria is the French national research institute dedicated to digital science and technology. It employs 2,600 people. Its 200 agile project teams, generally run jointly with academic partners, include more than 3,500 scientists and engineers working to meet the challenges of digital technology, often at the interface with other disciplines. The Institute also employs numerous talents in over forty different professions. 900 research support staff contribute to the preparation and development of scientific and entrepreneurial projects that have a worldwide impact.

Warning : you must enter your e-mail address in order to save your application to Inria. Applications must be submitted online on the Inria website. Processing of applications sent from other channels is not guaranteed.

Instruction to apply

Defence Security :
This position is likely to be situated in a restricted area (ZRR), as defined in Decree No. 2011-1425 relating to the protection of national scientific and technical potential (PPST).Authorisation to enter an area is granted by the director of the unit, following a favourable Ministerial decision, as defined in the decree of 3 July 2012 relating to the PPST. An unfavourable Ministerial decision in respect of a position situated in a ZRR would result in the cancellation of the appointment.

Recruitment Policy :
As part of its diversity policy, all Inria positions are accessible to people with disabilities.