Offre n°2022-05013

PhD Position F/M Fine-grained, multimodal speech anonymization

Le descriptif de l'offre ci-dessous est en Anglais

Type de contrat : CDD

Niveau de diplôme exigé : Bac + 5 ou équivalent

Fonction : Doctorant

Contexte et atouts du poste

This PhD is part of the "Personal data protection" project of PEPR Cybersécurité, which aims to advance privacy preservation technology for various application sectors. It will be co-supervised by Emmanuel Vincent and Marc Tommasi. The PhD student will have the opportunity to spend time in both the Multispeech and Magnet teams, to collaborate with 9 other research teams in France and with the French data protection authority CNIL, and to contribute to the project's overall goals including the organization of an anonymization challenge.

Mission confiée

Large-scale collection, storage, and processing of speech data poses severe privacy threats [1]. Indeed, speech encapsulates a wealth of personal data (e.g., age and gender, ethnic origin, personality traits, health and socio-economic status, etc.) which can be linked to the speaker's identity via metadata or via automatic speaker recognition. Speech data may also be used for voice spoofing using voice cloning software. With firm backing by privacy legislations such as the European general data protection regulation (GDPR), several initiatives are emerging to develop and evaluate privacy preservation solutions for speech technology. These include voice anonymization methods [2] which aim to conceal the speaker's voice identity without degrading the utility for downstream tasks, and speaker re-identification attacks [3] which aim to assess the resulting privacy guarantees, e.g., in the scope of the VoicePrivacy challenge series [4].

[1] A. Nautsch, A. Jimenez, A. Treiber, J. Kolberg, C. Jasserand, E. Kindt, H. Delgado, M. Todisco, M. A. Hmani, M. A. Mtibaa, A. Abdelraheem, A. Abad, F. Teixeira, M. Gomez-Barrero, D. Petrovska, N. Chollet, G. Evans, T. Schneider, J.-F. Bonastre, B. Raj, I. Trancoso, and C. Busch, "Preserving privacy in speaker and speech characterisation," Computer Speech and Language, vol. 58, pp. 441–480, 2019.

[2] B. M. L. Srivastava, M. Maouche, M. Sahidullah, E. Vincent, A. Bellet, M. Tommasi, N. Tomashenko, X. Wang, and J. Yamagishi, "Privacy and utility of x-vector based speaker anonymization," IEEE/ACM Transactions on Audio, Speech, and Language Processing, to appear.

[3] B. M. L. Srivastava, N. Vauquier, M. Sahidullah, A. Bellet, M. Tommasi, and E. Vincent, "Evaluating voice conversion-based privacy protection against informed attackers," in 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2802–2806, 2020.

[4] N. Tomashenko, X. Wang, E. Vincent, J. Patino, B. M. L. Srivastava, P.-G. Noé, A. Nautsch, N. Evans, J. Yamagishi, B. O'Brien, A. Chanclu, J.-F. Bonastre, M. Todisco, and M. Maouche, "The VoicePrivacy 2020 Challenge: Results and findings," Computer Speech and Language, vol. 74, pp. 101362, 2022.

Principales activités

The first objective of this PhD is to improve the privacy-utility tradeoff by better disentangling speaker identity from other attributes, and better decorrelating the underlying dimensions. Solutions may rely on suitable generative or self-supervised models [5, 6] or on adversarial learning [7]. The resulting privacy guarantees will be evaluated via stronger attackers, e.g., taking metadata into account.

The second objective is to extend the proposed audio-only approach to multimodal speech (audio, facial video, and gestures). Solutions will exploit existing facial anonymization technology [8]. A key difficulty will be to preserve the correlations between modalities, which are essential for training multimodal voice processing systems.

Depending on the PhD student's skills, additional directions may also be explored, e.g., evaluating the proposed anonymization solutions in the context of federated learning.

[5] L. Girin, S. Leglaive, X. Bie, J. Diard, T. Hueber, and X. Alameda-Pineda, "Dynamical variational autoencoders: A comprehensive review," Now Foundations and Trends, 2021.

[6] A. Baevski, H. Zhou, A. Mohamed, and M. Auli, “wav2vec 2.0: A framework for self-supervised learning of speech representations,” in Advances in Neural Information Processing Systems, pp. 12449–12460, 2020.

[7] B. M. L. Srivastava, A. Bellet, M. Tommasi, and E. Vincent, “Privacy-preserving adversarial representation learning in ASR: Reality or illusion?” in Interspeech, pp. 3700–3704, 2019.

[8] T. Ma, D. Li, W. Wang, and J. Dong, “CFA-Net: Controllable face anonymization network with identity representation manipulation,” arXiv preprint arXiv:2105.11137, 2021.

Compétences

MSc in computer science, machine learning, or signal processing.
Strong programming skills in Python/Pytorch.
Prior experience in speech and video processing will be an asset.

Avantages

- Subsidized meals
- Partial reimbursement of public transport costs
- Leave: 7 weeks of annual leave + 10 extra days off due to RTT (statutory reduction in working hours) + possibility of exceptional leave (sick children, moving home, etc.)
- Possibility of teleworking (after 6 months of employment) and flexible organization of working hours
- Professional equipment available (videoconferencing, loan of computer equipment, etc.)
- Social, cultural and sports events and activities
- Access to vocational training
- Social security coverage

Rémunération

1982€ gross/month for 1st and 2nd year.
2085€ gross/month for 3rd year.
Monthly salary after taxes : around 1596,05€ for 1st and 2nd year, 1678,99€ for 3rd year (medical insurance included).

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 :** 2022-10-01
- **Durée de contrat :** 3 ans
- **Date limite pour postuler :** 2022-07-07

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.