

Masquage en continu des informations personnelles dans la parole



Mots-clés : Anonymization de la parole, informations personnelles identifiantes, traitement du langage naturel en streaming

Description

Au sein du Pôle Recherche de l'AMIAD (Agence Ministérielle pour l'IA de Défense), vous rejoindrez notre équipe, où vous développerez un système d'anonymization de la parole capable de détecter les informations personnelles et modifier les mots d'un locuteur avec une latence minimale.

La tâche d'anonymisation de la parole consiste à rendre l'association d'un extrait de parole à son locuteur impossible, tout en conservant les signaux utiles (la parole doit rester intelligible et conserver les marqueurs d'émotions et d'intention ; rester naturelle). Ces applications sont multiples : de la protection de données personnelles, des données médicales, de la vie privée. Cette tâche est maintenant bien implantée dans la communauté du traitement de la parole, grâce notamment à des challenges récurrent comme le Voice Privacy Challenge¹.

Les solutions technologiques en réponse à ce challenge suivent généralement deux grands axes² :

- **Les solutions de conversion de voix (VC)** qui modifient directement le signal audio, permettant une parole en sortie plus naturelle, un traitement plus rapide de la parole qui permet le développement de systèmes en streaming, mais au coût d'une anonymization réduite.
- **Les solutions en cascades (ASR+TTS)** qui transcrivent puis re-génèrent la parole, permettant une anonymization presque parfaite, la possibilité d'utiliser des systèmes de traitement du langage pour retirer les informations personnelles avant reconstruction, mais au coût d'une latence bien plus élevée interdisant pour l'instant de passer en streaming.

Des travaux récents dans les modèles de langage en parole conversationnels (Conversational Speech LLMs) montrent qu'il est possible d'avoir un traitement et une reconstruction de la parole en un temps très réduit, par l'utilisation de représentations intermédiaires de la parole, en utilisant des codecs ou tokenizers audios³.

Objectifs

L'objectif principal du stage sera l'implémentation et la mise en œuvre d'un système d'anonymization de la parole en streaming incluant un masquage/remplacement des informations personnelles. Cette implémentation sera basée sur un système d'anonymisation en streaming existant, les objectifs principaux étant l'ajout du traitement des informations personnelles sans perte de latence. Le stagiaire bénéficiera du soutien de l'équipe du support technique du Pôle Recherche de l'AMIAD, notamment pour les différents choix technologiques.

Profil recherché

Vous suivez une formation d'ingénieur ou de master spécialité machine learning / linguistique informatique / data science et vous disposez :

- de connaissances préalables dans les domaines du traitement de la parole, de l'apprentissage automatique et de l'apprentissage profond ;
- d'un intérêt prononcé pour la recherche et les nouvelles technologies et êtes à l'aise en anglais ;
- de créativité et de motivation ;
- de solides compétences en programmation python et familier l'un des frameworks IA (idéalement PyTorch).

Informations générales

- Durée : 6 mois
- Niveau : Master 2 (ou équivalent en école d'ingénieur)
- Début : dès que possible
- Lieu : AMIAD - X-Innovation Center, École Polytechnique, Palaiseau
- Ressources mises à disposition : l'AMIAD dispose de ressources computationnelles et dédiées (Serveurs et GPU) ainsi qu'un accès à Jean-Zay, le super-calculateur du CNRS.
- Contacts : Thomas Thebaud, CR AMIAD

1. <https://www.voiceprivacychallenge.org/>

2. <https://www.voiceprivacychallenge.org/vp2024/docs/VPC-2024-.pdf>

3. <https://developers.openai.com/api/docs/models/gpt-realtime-2.1>

Point de contact

Thomas Thebaud : thomas.thebaud@polytechnique.edu