Anonymisation de parole par quantification vectorielle

Pierre Champion[1,2] Denis Jouvet[1] Anthony Larcher[2]

[1]Université de Lorraine, CNRS, Inria, LORIA, Nancy, France. [2]Le Mans Université, LIUM, France

Problématique

La reconnaissance de la parole est de plus en plus répandue notamment via **les assistants virtuels**.

- Ils **collectent, traitent et stockent** des données vocales sur des serveurs centralisés.
- Cette **transmission de données personnelles** ne respecte pas des contraintes légales et éthiques.

Anonymisation de la parole

L'anonymisation de la parole a pour but de supprimer des informations paralinguistiques personnelles contenues dans les signaux de parole.

Objectifs:

- **Dissimuler l'identité du locuteur** avant d'envoyer les signaux de parole aux fournisseurs de services.
- **Garder le contenu linguistique** afin de reconnaître la parole et de construire de grands corpus.

Contexte

Plutôt que de transmettre ses données de parole brutes, **chaque utilisateur anonymise sa parole** sur son terminal. Seule la représentation anonymisée est transmise aux fournisseurs de services.

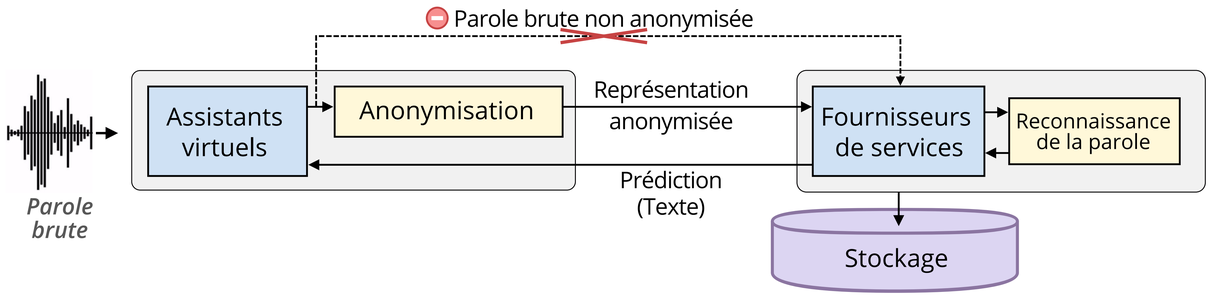

Figure 1. Processus de transmission avec anonymisation [1].

Extraction de représentations anonymisées

L'objectif est de **séparer** les caractéristiques exposant l'identité du locuteur du contenu linguistique.

La méthode de référence [2] consiste à extraire la représentation depuis la couche *bottleneck* d'un **modèle acoustique** de reconnaissance de la parole.

Amélioration de l'anonymisation par quantification vectorielle

- La **séparation** entre l'identité et le contenu n'est pas assez forte avec la méthode de référence.
- Nous proposons d'ajouter une **contrainte** sur la couche *bottleneck* de référence via une quantification vectorielle.

La quantification vectorielle remplace les ***bottlenecks* continus** par d'autres de même dimension, mais appartenant à un **ensemble discret et fini** de N vecteurs $\{e_1, e_2, \ldots, e_N\}$.

$$QV : \mathbb{R}^{256} \rightarrow \{e_1, e_2, \ldots, e_N\} \ \ avec \ \ e_i \in \mathbb{R}^{256}$$

Évaluation

Notre méthode est évaluée et comparée avec la méthode de référence sur le jeu de données **LibriSpeech** *test-clean*.

Métriques de privacité et d'utilité:

1. ***Equal Error Rate*** ($\text{EER}_{\%}$), mesure la capacité de bien dissimuler l'identité du locuteur.
2. ***Word Error Rate*** ($\text{WER}_{\%}$), mesure l'intelligibilité du contenu linguistique.

Nombre de vecteurs N de quantification	Privacité $\text{EER}_{\%} \uparrow$	Utilité $\text{WER}_{\%} \downarrow$
Méthode de référence (sans QV)	4.2	5.8
1024	18.1	**7.2**
256	19.4	7.6
48	22.3	8.7
32	26.5	9.8
16	**31.2**	15.9
Idéal théorique	50.0	0.0

Table 1. Performances de reconnaissance du locuteur et de la parole.

Conclusion

Sans quantification, les représentations encodent à la fois l'information de l'identité du locuteur et le contenu linguistique. La quantification vectorielle permet de contrôler le compromis entre la privacité et l'utilité.

Référence

[1] S. A. Osia, A. Shahin Shamsabadi, S. Sajadmanesh, *et al.*, "A hybrid deep learning architecture for privacy-preserving mobile analytics," *IEEE Internet of Things Journal*, 2020.

[2] L. Sun, K. Li, H. Wang, S. Kang, and H. Meng, "Phonetic posteriorgrams for many-to-one voice conversion without parallel data training," in *IEEE International Conference on Multimedia and Expo*, 2016.