Méthodes avancées d'anonymisation pour le contenu textuel et vocal industriel

Nassim Boutikar*, Asma Trabelsi* Emmanuel Helbert* Sebastien Warichet* Hamed Kallo*

* 260 Rue Léon Foucault, 67400 Illkirch Graffenstaden
https://www.openrainbow.com/

1 Introduction

La confidentialité des données est cruciale dans les enregistrements vocaux et textuels, particulièrement pour les outils comme $Rainbow^{TM}$. L'anonymisation préserve la vie privée tout en permettant une utilisation légale des données, conforme au RGPD. Cet article explore des techniques d'anonymisation robustes pour l'audio et le texte, avec un accent sur la reconnaissance d'entités nommées (NER) et la protection contre le clonage vocal.

2 Importance dans l'industrie

L'anonymisation des données est essentielle dans des secteurs comme la santé, la finance et le juridique. Elle protège les informations personnelles tout en permettant leur exploitation sécurisée (ex. : dossiers médicaux anonymisés pour la recherche). Dans les plateformes de collaboration en temps réel comme $Rainbow^{TM}$, anonymiser les voix et les contenus sensibles (noms, lieux, données bancaires) renforce la confidentialité et améliore la conformité réglementaire.

2.1 Anonymisation par pipeline

Les approches par pipeline pour l'anonymisation vocale suivent une séquence d'étapes modulaires. D'abord, la reconnaissance automatique de la parole (ASR), via des modèles comme WhisperX ou Wav2Vec 2.0, convertit l'audio en texte avec des horodatages précis. Ensuite, des modèles NER comme Flair ou SpaCy identifient et anonymisent les entités sensibles du texte. Enfin, l'audio est modifié en fonction des segments anonymisés, soit par insertion de bips, transformation vocale, ou synthèse vocale.

Cette méthode est flexible et débogable, mais peut souffrir de propagation d'erreurs (ex. : ASR vers NER) et être plus lente en raison de ses multiples étapes. Elle offre cependant une personnalisation à chaque stade, influençant directement la confidentialité et la qualité des résultats.

TAB. 1 – *Comparaison des outils NER pour l'anonymisation*

Outil	Langue supportée	Points forts	Nb entités détectées
Flair	Multilingue	Haute précision	10+
SpaCy	Multilingue	Rapidité	18+
Presidio	Anglais	Intégration facile	12+
Stanford NER	Anglais	Haute précision académique	7
Duckling	Multilingue	Extraction d'expressions spécifiques	5+

2.2 Approches de bout en bout (E2E)

Les approches E2E anonymisent directement les entrées audio en s'appuyant sur des architectures neuronales avancées, évitant les étapes intermédiaires comme l'ASR et le NER. Une méthode notable est la solution Signal-to-Entity (S2E), combinant extraction d'embeddings multimodaux (via Whisper pour l'audio et BERT pour les entités textuelles), un mécanisme de co-attention pour aligner efficacement les données audio et textuelles, et des pertes multimodales pour l'optimisation. Cette approche optimise directement la reconnaissance et l'anonymisation des entités, réduisant la propagation des erreurs et la complexité du modèle, tout en offrant des performances améliorées.

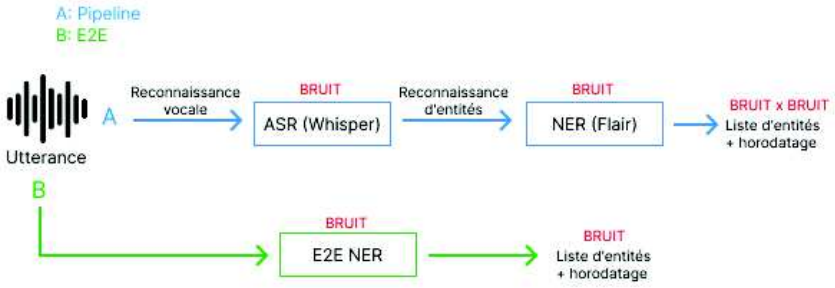

FIG. 1 – *Architecture*

3 Conclusion

Les approches par pipeline utilisant l'ASR et la NER offrent modularité et facilité de débogage, mais souffrent de propagation d'erreurs et de latence. À l'inverse, les modèles de bout en bout (E2E) anonymisent directement le signal audio, réduisant les erreurs et améliorant précision et efficacité. Des techniques comme Wav2Vec 2.0 et HuBERT exploitent des représentations audio robustes pour ces tâches, bien qu'elles nécessitent davantage de données et de calculs. La Figure 1 illustre les points forts et les compromis entre ces deux approches.