# Une architecture légère de reconnaissance vocale pour les langues peu dotées : application au poular et au soussou

Ibrahima Sory Bah, Mariame Touré
Laboratoire d'Intelligence Artificielle Appliquée, Institut Supérieur des Technologies, Conakry, Guinée

3e Conférence Internationale sur le Traitement Automatique des Langues Africaines (CITALA 2025), Dakar, Sénégal | Actes de conférence, pp. 142-151 | DOI: 10.9182/citala.2025.0142

## Résumé

Les langues à faibles ressources numériques, telles que le poular et le soussou, restent largement absentes des systèmes modernes de reconnaissance vocale. Nous présentons une architecture légère, basée sur un encodeur convolutionnel suivi d'un module d'attention réduit, conçue pour fonctionner avec un corpus d'entraînement limité à moins de 20 heures d'audio annoté par langue. Évaluée sur un corpus collecté auprès de locuteurs natifs de la région de Labé et de Boffa, notre approche atteint un taux d'erreur de mots de 31,2% pour le poular et 28,7% pour le soussou, des performances comparables à des modèles nécessitant dix fois plus de données d'entraînement.

**Mots-clés :** reconnaissance vocale, langues peu dotées, poular, soussou, apprentissage profond, traitement automatique des langues

## 1. Introduction

Le développement de technologies vocales pour les langues africaines peu dotées se heurte à la rareté des corpus annotés. Le poular et le soussou, parlés respectivement par plusieurs millions de locuteurs en Guinée et dans la sous-région, ne disposent à ce jour d'aucun système de reconnaissance vocale robuste. Cette absence limite considérablement l'accès aux services numériques pour une part importante de la population, notamment en zone rurale où le taux d'alphabétisation en français reste faible.

## 2. Approche proposée

Notre architecture repose sur trois blocs convolutionnels suivis d'un module d'attention à deux têtes, conçu pour limiter le nombre de paramètres entraînables tout en conservant une capacité de généralisation suffisante. Un pré-entraînement par apprentissage auto-supervisé sur un corpus multilingue ouest-africain non annoté précède la phase de fine-tuning sur les données spécifiques au poular et au soussou.

## 3. Expérimentations et résultats

Le corpus d'évaluation comprend 18 heures d'audio pour le poular et 15 heures pour le soussou, collectées dans des conditions d'enregistrement variées (studio, extérieur, téléphone mobile). Le tableau 1 (non reproduit ici) compare nos résultats à deux systèmes de référence : un modèle Whisper fine-tuné et un modèle CTC classique. Notre approche surpasse le modèle CTC de 6,4 points de pourcentage tout en nécessitant 60% de temps d'entraînement en moins.

## 4. Conclusion et perspectives

Cette étude démontre qu'une architecture légère, combinée à un pré-entraînement auto-supervisé adapté au contexte régional, permet d'obtenir des performances satisfaisantes pour la reconnaissance vocale de langues peu dotées avec des ressources limitées. Les travaux futurs porteront sur l'extension du système à d'autres langues guinéennes, notamment le maninka et le kissi, ainsi que sur le déploiement d'une application mobile pilote en zone rurale.

## Références

Camara, L., & Diop, A. (2023). Corpus vocal pour les langues mandingues : constitution et défis. Actes de TALAF 2023, 88-97.

Kaba, S. (2022). Self-supervised speech representation learning for African languages. Proceedings of InterSpeech, 2201-2205.