

# Reconnaissance vocale pour le handicap moteur

**Mots-clés** : Transcription automatique, réseaux de neurones, adaptation, handicap moteur

## Référence client

Rouas, Jean-Luc [rouas@labri.fr](mailto:rouas@labri.fr) , Lacroix, Pierre [pierre.lacroix@u-bordeaux.fr](mailto:pierre.lacroix@u-bordeaux.fr)

## Domaine

Le domaine du projet est la transcription automatique de la voix, en utilisant des modèles basés sur des réseaux de neurones pré-entraînés, comme par exemple Whisper [1] mais également des modèles développés au LaBRI pour lesquels nous connaissons précisément les données d'entraînement. Il s'agit de proposer des adaptations de ce(s) modèle(s) pour qu'ils fonctionnent sur de la parole de personnes ayant un handicap moteur. Une étape de collecte de données auprès d'un locuteur ayant ce handicap sera également réalisée.

## Problématique

La problématique est l'adaptation de modèles de réseaux de neurones profonds (typiquement transformers [2] ou conformers [3]) en utilisant une faible quantité de données. Différentes techniques d'adaptation pourront être explorées (voir par exemple [4]) afin de comparer leurs performances en fonction du volume de données d'entraînement utilisé (qui restera de toute manière très faible). Une fois cette étape réalisée, l'objectif final est la création d'une application simple d'utilisation permettant la dictée vocale et éventuellement le contrôle de l'ordinateur.

## Environnement technique

Ce projet nécessite d'avoir accès à des calculateurs munis de GPUs pour l'adaptation des modèles. Comme il s'agit uniquement d'adaptation, les temps de calculs seront raisonnables. Le développement des méthodes d'adaptation et de l'application pourra se faire en python. Le système d'exploitation cible sera certainement Windows (à confirmer avec les utilisateurs potentiels)

## Objectifs

1. Collecte de données : il s'agira de mettre au point en concertation avec les utilisateurs potentiels une liste de phrases/commandes et de procéder à leur enregistrement vocal.
2. Évaluation des performances de modèles génériques de transcription automatique en français sur les données collectées : inférence de systèmes en libre accès (par ex. whisper) et/ou de modèles entraînés au LaBRI.
3. Adaptation des modèles avec une partie des données collectées et évaluation
4. Développement de l'application finale, interface et tests utilisateurs.

## Références

- [1] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey, and I. Sutskever, "Robust speech recognition via large-scale weak supervision," in *Proceedings of the 40th International Conference on Machine Learning*, in ICML'23, vol. 202. Honolulu, Hawaii, USA: JMLR.org, July 2023, pp. 28492–28518.
- [2] L. Dong, S. Xu, and B. Xu, "Speech-transformer: A no-recurrence sequence-to-sequence model for speech recognition," in *2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, 2018, pp. 5884–5888. doi: 10.1109/ICASSP.2018.8462506.
- [3] A. Gulati *et al.*, "Conformer: Convolution-augmented Transformer for Speech Recognition," in *Interspeech 2020*, ISCA, Oct. 2020, pp. 5036–5040. doi: 10.21437/Interspeech.2020-3015.
- [4] Q. Wang *et al.*, "Dysarthric Speech Conformer: Adaptation for Sequence-to-Sequence Dysarthric Speech Recognition," in *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Apr. 2025, pp. 1–5. doi: 10.1109/ICASSP49660.2025.10889046.