Classification des phonèmes à base d'apprentissage profond - Projection dans un contexte de parole dégradée

Sondes ABDERRAZEK[1], Corinne FREDOUILLE[1], Alain GHIO[2], Muriel LALAIN[2], Christine MEUNIER[2], Virginie WOISARD[3]
[1]LIA, Avignon Université
[2]Aix-Marseille Université, LPL, CNRS, Aix-en-Provence
[3]UT2J, Octogone-Lordat, Toulouse Université & Toulouse Hôpital

PROJET FINANCÉ PAR L'ANR - PROJECT FUNDED BY THE ANR

01 INTRODUCTION

Contexte:
En pratique clinique, les mesures perceptives restent la méthode la plus utilisée pour évaluer les troubles de parole.

Problématique:
Les mesures perceptives :
- sont subjectives
- manquent d'interprétation
- manquent de reproductibilité

=> Besoin d'un outil d'évaluation objectif

ETAPE1:
Mise en place d'un CNN pour une tâche de classification automatique des phonèmes

ETAPE2:
Etendre le CNN pour une tâche de prédiction d'intelligibilité

ETAPE3:
Interpréter le modèle en terme de capacité à rapporter de l'information sur la contribution des unités phonémiques au maintien/perte de l'intelligibilité

PROJET

Identification des unités linguistiques porteuses d'intelligibilité chez des patients atteints de troubles de la parole via du Deep Learning

02 DATA & ARCHITECTURE

2.1 Datasets (Train/Val/Test)

Données normales

Signal parole + Alignement en phonème

Données de parole pathologique

⚠ Mesures perceptives

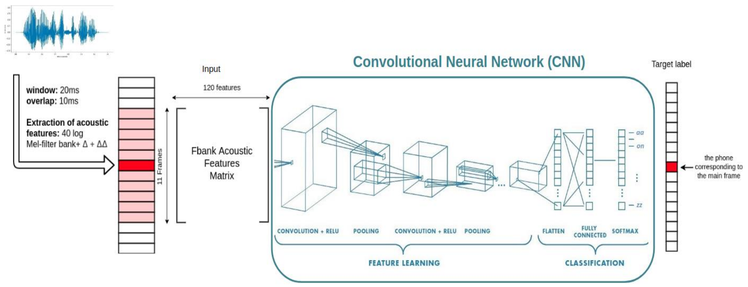
Utilisées pour le *train*, la *validation* et le *test*
- Parole lue (extrait du journal Le Monde)
- 120 locuteurs
- Recrutés dans la région de **Paris**

Utilisées pour *test*
- Parole lue (La chèvre de Mr. Seuguin)
- 82 patients (traités pour des cancer de la tête/ du cou) + 24 témoins
- Recrutés dans la région **Sud-Ouest**

⚠ **Mesures perceptives:** scores attribués par un jury pour évaluer la qualité de parole des patients/ contrôles.
- **Sévérité & Intelligibilité :** (0-Ininteligible; 10- Parfaitement intelligible)
- **Altération phonémique :** (0-Pas d'altération; 3-Altération majeure)

2.2 Architecture du modèle

- **Architecture neuronale :** Convolutional Neural Network (CNN) pour la tâche de classification
- **Input :** Paramètres Banc de filtre d'un contexte de 11 trames (la trame centrale + les 5 trames avant + les 5 trames après)
- **Output :** 32 classes (31 phonèmes français + le silence)



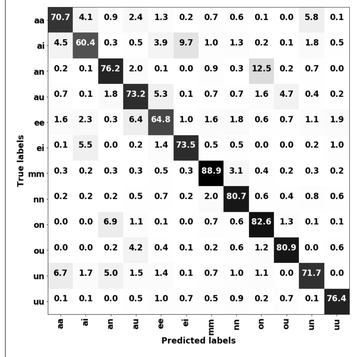
03 RÉSULTATS

3.1 Analyse des performances

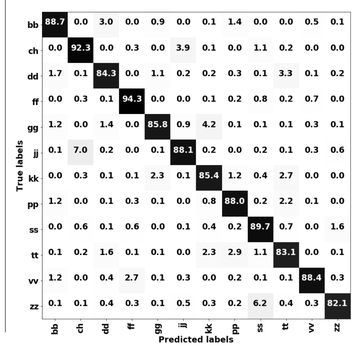
Evaluation de la performance du modèle sur de la parole normale: Effet de changement des données de test (comparaison des matrices de confusion issues de BREF et celles issues des témoins de C2SI)

	Données de Test BREF	Données témoins de C2SI
Accuracy	82%	74%

Matrice de confusion (Voy. Orales + Voy. Nasales + Cons. Nasales)



Matrice de confusion (Cons. Voisées + Cons. Non voisées)



Confusions : "an" avec "aa" et "nn" "un" avec "ai" et "nn"
Due à la nasalisation moins complète des voyelles nasales dans l'accent de sud-ouest

Confusions : "j" avec "ch" "gg" avec "kk"
Due à la perte du trait distinctif de voisement ("jj": fricative voisée Vs "ch": fricative non voisée)

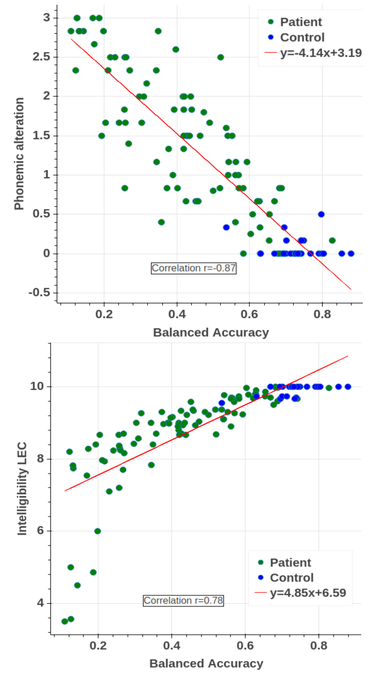
Confusions : "ff" avec "ss" "pp" avec "tt"
Due à la perte du trait distinctif lié au lieu d'articulation

3.2 Analyse des corrélations

Les matrices de confusion montrent que la performance du modèle reste stable lorsqu'il est confronté à de nouvelles données (C2SI dataset).

On a supposé alors que toute dégradation significative des performances du modèle est en relation avec le degré de dégradation de la qualité de la parole.

Dans cette étude, on analyse la corrélation entre les taux de classification obtenus par le modèle pour chacun des contrôles/ patients avec leurs mesures perceptives.



- Les mesures perceptives corrélant le mieux avec les performances de notre modèle sont **la sévérité (r=0.91)** et **l'altération phonémique (r=-0.87)**

=> Ces dernières s'approchent le plus de la tâche de classification de phonème en tenant en compte l'altération acoustique globale, resp. locale perçue des unités linguistiques.

- **L'intelligibilité** corrèle moins **(r=0.78)**

=> **Surestimation** de la mesure par les experts due au phénomène d'habituation au texte lu utilisé (l'intelligibilité devient plutôt une mesure de compréhensibilité)

04 CONCLUSION

Ce travail est la première étape d'un projet à long terme qui vise à déterminer les unités linguistiques contribuant au maintien ou la perte de l'intelligibilité dans un contexte de parole dégradée.

Un modèle d'architecture CNN était entrainé sur de la parole normale lue (BREF) pour la tâche de classification de phonème. Ensuite, une étude des matrices de confusion a révélé que le modèle est assez stable lorsqu'il est exposé à des nouvelles données (témoins de C2SI).
Confronté à la parole dégradé, le modèle a démontré une capacité d'encodage des traits phonémiques et ceci à travers une forte corrélation entre les taux de classification par patient/témoin et leurs mesures perceptives.

Remerciements:
Ce travail a été réalisé dans le contexte du projet RUGBI financé par l'ANR (Contrat n° ANR-18-CE45-0008-04).

Présenté par : Sondes ABDERRAZEK
Publié à : Interspeech 2020
DATE :27/05/2021