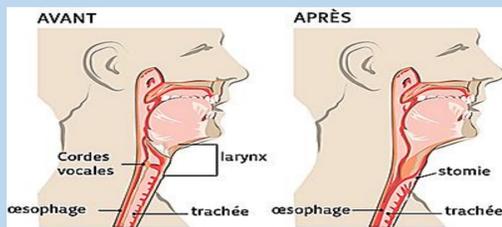


Mouad Laribia<sup>1</sup>, Imen Ben Othmane<sup>2</sup>, Joseph Di Martino<sup>3</sup>, Elhassan Ibn Elhaj<sup>1</sup>  
 mouadlary@gmail.com, imen.benothmen01@gmail.com, joseph.di-martino@loria.fr, ibnelhaj@inpt.ac.ma

- (1) Laboratoire de Recherche Systèmes de Télécommunications, Réseaux et Services. Institut National des Postes et Télécommunications, INPT, Rabat, Maroc.  
 (2) Laboratoire de Recherche Electricité Intelligente & TIC. Ecole Nationale d'Ingénieurs de Carthage, ENICarthage, Université de Carthage, Tunisie.  
 (3) Laboratoire Lorrain de Recherche en Informatique et ses Applications, LORIA, Vandœuvre-lès-Nancy, France.

## Problématique

Suite à une laryngectomie totale et l'ablation du larynx, organe responsable de la parole, le patient ne sera plus en mesure de produire une voix laryngée. Ce qui le mène à apprendre une voix de remplacement qui est la voix œsophagienne.



La voix œsophagienne est caractérisée par :

- Une mauvaise qualité audio.
- Des perturbations dans le signal de parole très élevées.
- Une enveloppe spectrale instable.
- Une faible et chaotique fréquence fondamentale F0.

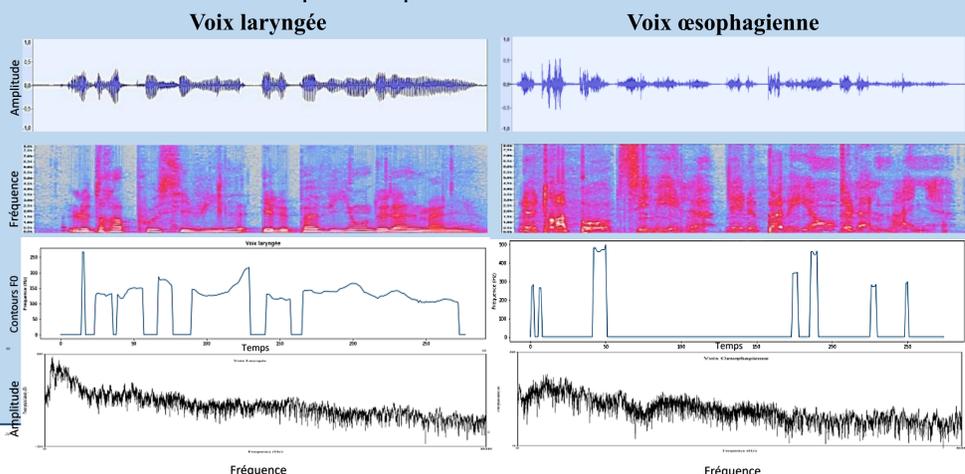


FIGURE 1: Exemple de caractéristiques acoustiques d'une voix laryngée et d'une voix œsophagienne (Forme du signal, Spectrogramme, Contour F0 et Spectre d'amplitude).

## Conversion de Voix

La conversion de la voix est un processus qui transforme la voix d'un locuteur source en celle d'une voix d'un locuteur cible tout en préservant le contenu linguistique du signal original.



## Références

- [1] BEN OTHMANE I., DI MARTINO J., & OUNI K. (2018). Vers la transformation de la parole œsophagienne en voix laryngée à l'aide de techniques de conversion vocale. *7ème Journées de Phonétique Clinique - JPC7*, 2017.  
 [2] BEN OTHMANE I., DI MARTINO J., & OUNI K. (2018). Enhancement of esophageal speech using statistical and neuromimetic voice conversion techniques. *Journal of International Science and General Applications*, Vol. 1, No. 1.  
 [3] BEN OTHMANE I., DI MARTINO J., & OUNI K. (2019). Enhancement of esophageal speech obtained by a voice conversion technique using time dilated Fourier cepstra. *International Journal of Speech Technology* 22, p. 99-110.

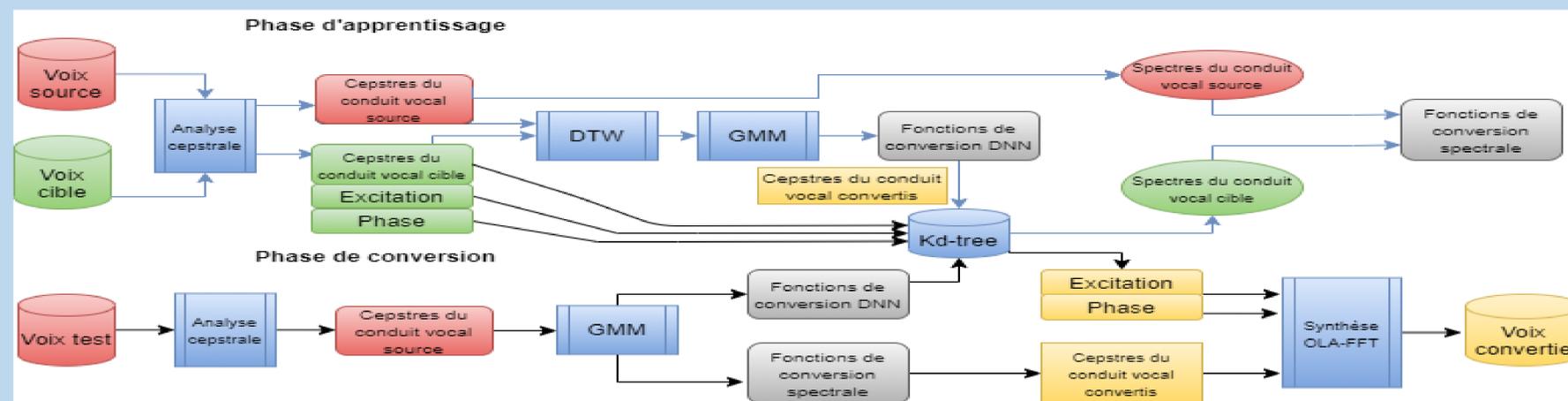
## Objectif

Afin d'améliorer la qualité et l'intelligibilité de la voix œsophagienne, ce travail vise à développer un système de rehaussement en utilisant les techniques de conversion de la voix tout en tenant compte de la particularité de l'appareil vocal pathologique.

## Contribution

Le système de conversion de voix œsophagienne vers une voix laryngée est modélisé par deux phases:

- **Une phase d'apprentissage :**
  - Analyse cepstrale et alignement temporel par DTW des vecteurs cepstraux du conduit vocal source et cible.
  - Classification des vecteurs alignés et concaténés par le modèle de mélange gaussien GMM.
  - Estimation d'une fonction de transformation non linéaire par réseaux de neurones pour chaque classe.
  - Estimation d'une fonction de transformation spectrale pour chaque classe à partir des vecteurs spectraux.
- **Une phase de conversion :**
  - Conversion de nouveaux vecteurs relatifs au conduit vocal.
  - Prédiction de l'excitation et de la phase à partir de l'espace d'apprentissage cible préalablement codé sous la forme d'un arbre binaire.
  - Synthèse en utilisant les premiers paquets cepstraux convertis par les fonctions de transformation spectrale.



## Résultats

- ✓ La méthode proposée nous permet de :
  - ➔ Produire un son audio de meilleure qualité.
  - ➔ Reconstruire les informations prosodiques des signaux.

- ✓ L'évaluation objective par le rapport Signal sur Erreur (SER) indique une amélioration des signaux transformés:

Corpus	Source	DNN_src	DNN_fcs
PC	1.94	2.18	4.33
MH	2.69	3.24	3.45

TABLE 1: Estimation du rapport signal sur erreur (SER) des sons tests resynthésés avec les conduits vocaux source (DNN\_src) et les conduits vocaux convertis par fonction de conversion spectrale (DNN\_fcs).

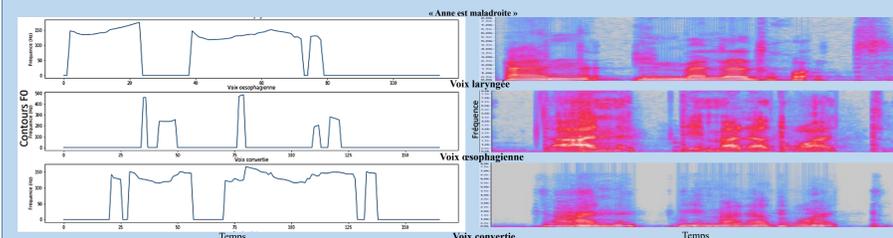


FIGURE 2: Comparaison des caractéristiques acoustiques entre une voix laryngée, une voix œsophagienne et une voix convertie (Contour F0 (gauche) et Spectrogramme (droite)).

## Conclusion

Ce travail présente une nouvelle méthode de rehaussement de la voix œsophagienne en estimant d'abord des fonctions de transformation par réseaux de neurones pour la prédiction de l'excitation et de la phase. Ensuite, des fonctions de conversion spectrale sont estimées afin de réduire les irrégularités observées au niveau formantique.