La reconnaissance automatique de la parole dans le contexte médical

Imed LAARIDH

Introduction (1/2)

La Maladie de Parkinson (MP):

- La maladie neuro-dégénérative la plus courante après la maladie d'Alzheimer, touchant environ 1,5% de la population de plus de 65 ans et environ 170 000 Français *[Tison et al., 1994]*.

L'Atrophie Multi-Systématisée (AMS) :

- Maladie neuro-dégénérative rare d'étiologie inconnue. Elle se caractérise par une combinaison variable de parkinsonisme, atteinte cérébelleuse, troubles dysautonomiques et syndromes pyramidaux *[Tison et al., 2000]*.

Introduction (2/2)

- En stade précoce, les symptômes de la MP et de l'AMS sont très similaires

  => **diagnostic différentiel** souvent très difficile mais très important en raison du pronostic divergent et de la gravité du pronostic de l'AMS

- Pas de **marqueur objectif** validé actuellement disponible pour guider ce diagnostic
  => Besoin de tels marqueurs dans la communauté neurologique.

- La **dysarthrie** est un symptôme **commun** et **précoce** dans les deux maladies
  => Utiliser la dysarthrie pour caractériser et rechercher des différences entre les patients atteints de MP et d'AMS aux premiers stades des maladies.

Objectifs

- Projet ANR **Voice4PD-MSA:** projet pluridisciplinaire (professionnels de la médecine, de l'orthophonie, de l'informatique et de la statistique).

  => Proposer des **marqueurs** extraits à partir de la **parole** pour l'assistance dans le **diagnostic différentiel** entre la MP et l'AMS.

  => Utiliser la **reconnaissance automatique de la parole** pour la recherche de tels marqueurs.

  => La caractérisation de la parole dysarthrique liée à la MP et l'AMS en utilisant la reconnaissance automatique de la parole.

Corpus (1/2)

- Corpus **Voice4PD-MSA** en cours de construction contenant des patients souffrant de stage précoce de MP et AMS. Enregistrements aux CHU Toulouse et CHU Bordeaux sous le contrôle de phoniatres.
- Durant la session d'enregistrement, tous les locuteurs ont réalisé les mêmes tâches de production de parole :
    - Lecture de texte
    - Parole spontanée
    - Voyelle /a/ tenue
    - Répétition de syllabes /pataka/ et /badaga/
    - Lecture de logatomes (pseudo-mots)

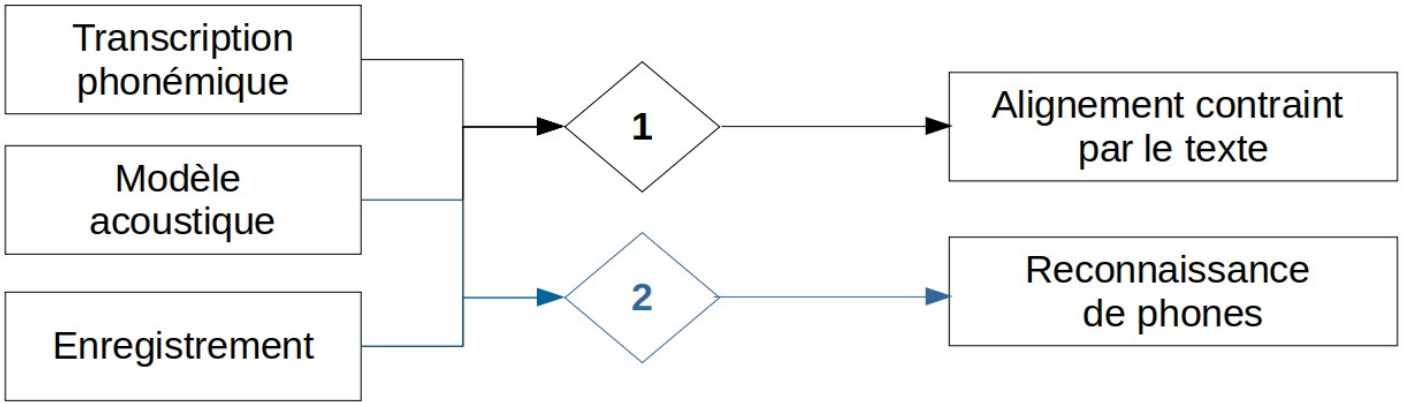
Corpus (2/2)

- Dans ce travail, nous nous concentrons sur la tâche de lecture du texte "La chèvre de monsieur Séguin" d'environ 70 mots.

Population	# de locuteurs	Durée moyenne (écart-type) (s)
Maladie de Parkinson (MP)	20	22.22 (3.56)
Atrophie Multi-Systématisée (AMS)	14	23.86 (4.51)
Témoins	5	22.00 (1.82)

Méthodologie

La méthodologie proposée pour la recherche de marqueurs pour le l'aide au diagnostic différentiel entre la MP et l'AMS consiste en deux traitements automatiques *[Povey et al., 2011]*.



Alignement contraint par le texte

L'alignement contraint par le texte prend comme entrée:

- **Un modèle acoustique** : appris sur ~200h de parole normale radiophonique *[Galliano et al., 2005].*

- **La transcription de l'enregistrement**: une transcription manuelle, après écoute, où sont introduits les ajouts, suppressions et substitutions de mots, réalisés par les patients par rapport au texte de référence.

- **Un lexique phonetisé**: des variantes de prononciation pour chaque mot sont introduites dans le lexique, conformément aux règles de prononciation standard et aux règles de liaison potentielles.

=> Cet alignement résulte en 2 segmentations temporelles des enregistrements : l'une correspond à la localisation des phonèmes avec, pour chacun, ses frontières de début et de fin dans le signal ; l'autre permet de localiser les mots et les pauses.

Reconnaissance de phonème

- Le système de reconnaissance automatique de phones utilise comme élément de base **le phonème** et non le traditionnel "mot".

  => La parole produite par les locuteurs est transcrite en une suite de phonèmes.

- Utilisation d'un modèle de langage **phonétique 1-gram équiprobable** :
  - Utiliser l'information acoustique uniquement lors de la reconnaissance de phones sans correction éventuelle introduite par le lexique de mots et/ou le modèle de langage associé.
  - Pouvoir utiliser le même modèle sur la tâche de production de logatomes (pseudo-mots) où lexique et modèle de langage n'ont pas de sens.

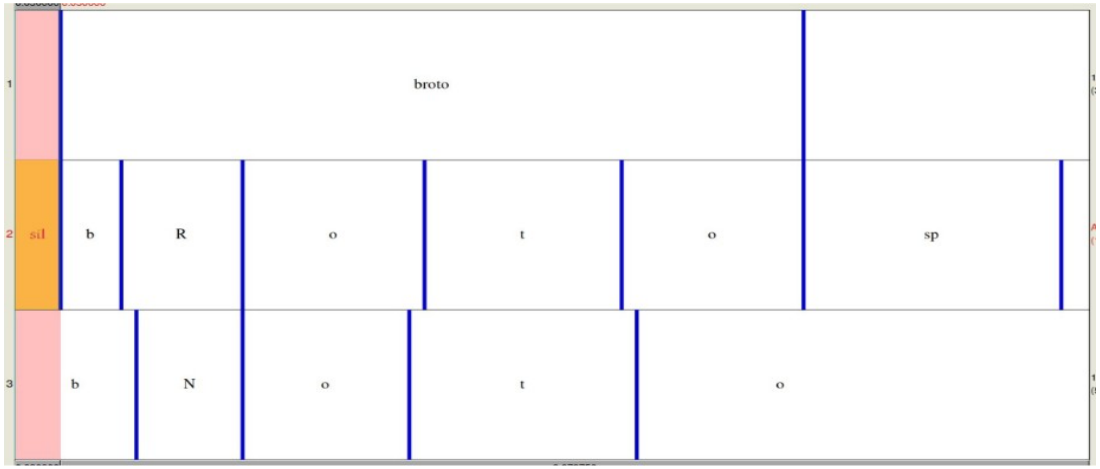
=> Ce processus résulte en une localisation temporelle des phones, **indépendante de la transcription** du texte lu, ainsi que leur labellisation.

Qualité acoustique au niveau phonème (1/2)

- Aligner temporellement les sorties de l'alignement contraint par le texte et de la reconnaissance automatique phonétique au niveau trame (10 ms).

Alignement contraint par le texte

Reconnaissance automatique de phones



$$TR = 100 * \frac{\#\ trames\ bien\ reconnues\ par\ la\ reconnaissance\ automatique}{\#\ de\ trames\ de\ référence}$$

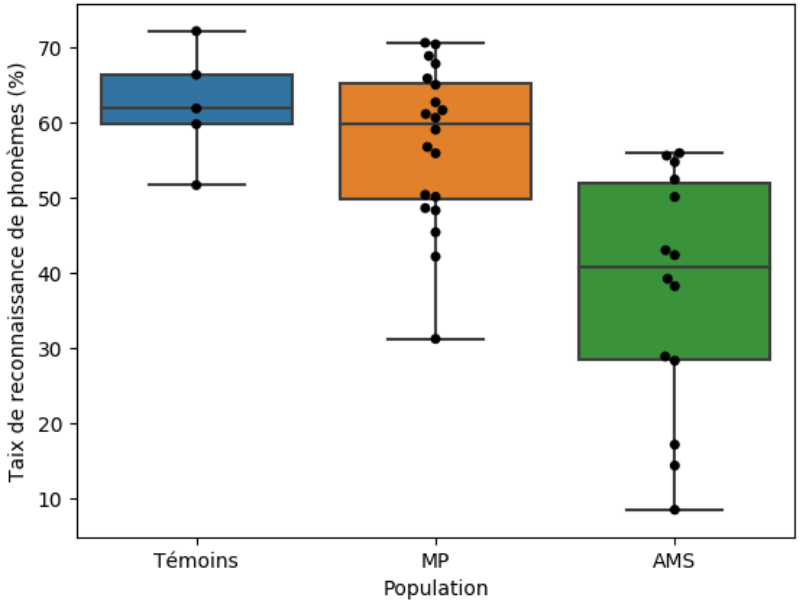
Qualité acoustique au niveau phonème (2/2)

- Faible taux de reconnaissance chez les patients atteints d'AMS (38%).

- Taux comparables entre les patients atteints de la MP (57%) et les locuteur témoins (62%).

=> La dysarthrie liée à l'AMS est souvent **plus sévère** que celle associée à la MP.

=> ***TR*** présente **un potentiel** pour son utilisation pour **la différenciation** entre les deux pathologies.



Durée moyenne des voyelles (1/2)

- L'allongement des phonèmes, plus particulièrement celui des voyelles, est souvent associé à la **dysarthrie ataxique** *[Darley et al., 1975]*.

- La **dysarthrie mixte** associée à l'AMS comporte une composante **ataxique**, absente de la dysarthrie associée à la MP.

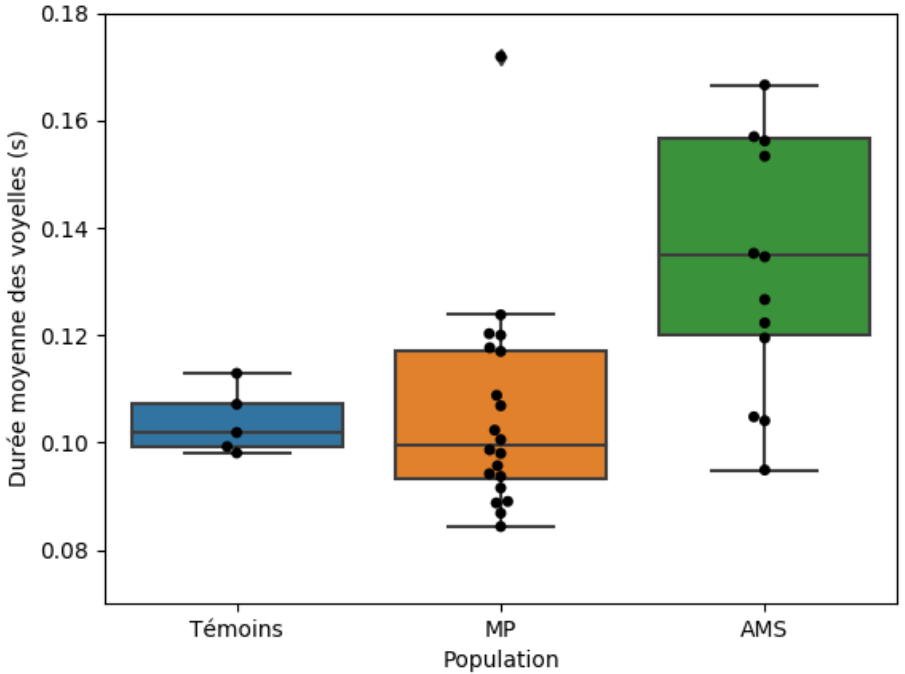
- L'indicateur ***Dur_voy*** est calculé sur chaque enregistrement à l'issue de la reconnaissance automatique de phonèmes.

$$Dur_{voy} = \frac{Durée\ totale\ des\ voyelles}{\#\ de\ voyelles}$$

Durée moyenne des voyelles (2/2)

- Les patients atteints d'AMS présentent un **comportement distinctif**, avec des prononciations de voyelles significativement plus longues que la normale ($p<0.001$).

=> Ce résultat confirme la pertinence de la mesure de **durée moyenne des voyelles** qui reflète **l'allongement des voyelles** associé à la dysarthrie ataxique présente chez les patients atteints d'AMS.



Débit de parole (1/2)

- La dysarthrie ataxique est caractérisée par un **ralentissement du débit** de la parole, contrairement à la **dysarthrie hypokinétique** associée à la MP caractérisée par des **irrégularités** (voire des accélérations) du débit.

- Calcul du nombre de voyelles prononcées par seconde; c'est un estimateur proche du nombre de syllabes par seconde et de la vitesse d'élocution *[Rouas et al., 2004]*.

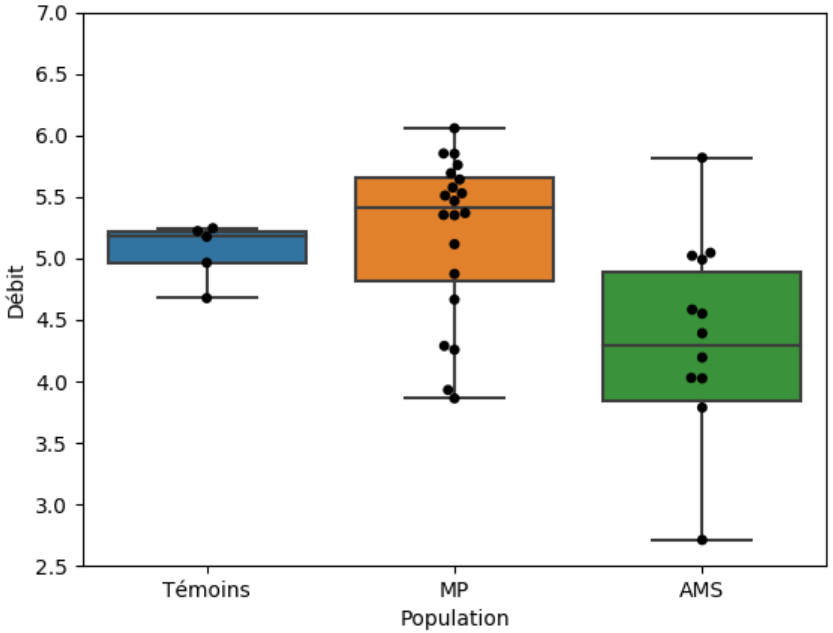
- Le paramètre ***Débit*** est extrait à partir des sorties de la reconnaissance automatique de phonèmes. Pour éviter tout biais, les pauses et respirations ne sont pas prises en compte pour le calcul.

$$Débit = \frac{\#\ de\ voyelles}{Durée\ totale\ de\ lecture - (Durée\ pause\ + Durée\ respiration)}$$

Débit de parole (2/2)

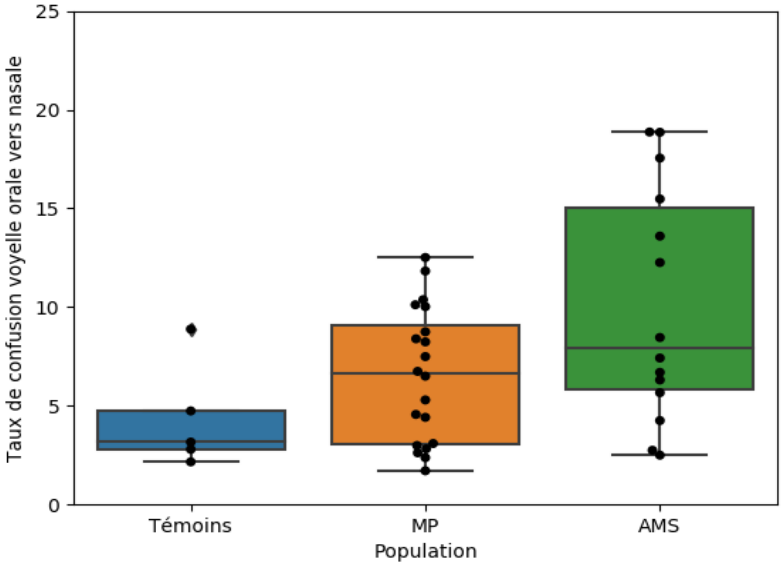
- Les patients atteints d'AMS présentent un débit de parole significativement **plus lent** que les locuteurs témoins ou atteints de la MP ($p<0.01$).

- La variance élevée peut être liée à:
  - Variabilité intra-population entre les individus (plus que probable)
  - Instabilité éventuelle d'un débit "local": irrégularité du débit au cours du temps pour chaque individu.

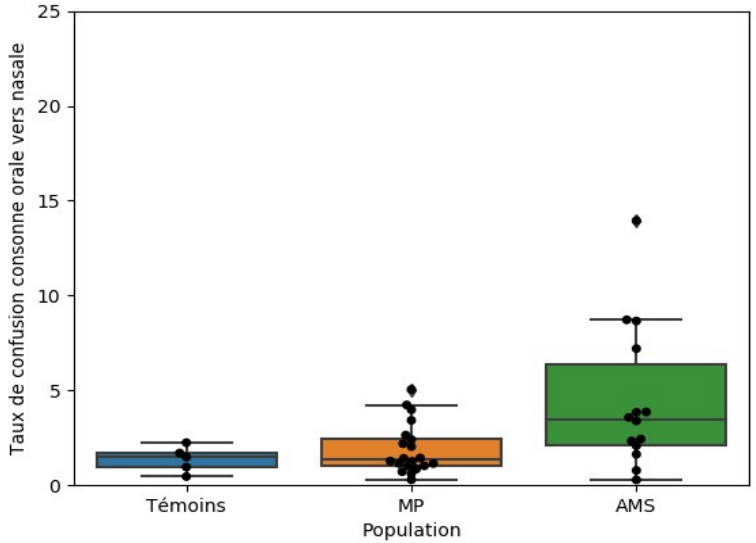


Nasalisation

Voyelles



Consonnes



Plus de confusion oral → nasal chez les patients atteints d'AMS

Distance de confusion (1/4)

- **La distance entre 2 phonèmes est basée sur les traits acoustiques:**
  - 5 traits pour les voyelles et 6 pour les consonnes [Ghio 2016], [Chomsky et Halle 68]
  - Voyelles : nasale; arrière; haut ;arrondi; ouvert
  - Consonnes : vocalique; continuité ;nasalité ;voisement; compact/diffus (velars vs Palatals); grave/Aigu (labial vs dental et velars vs Palatals)

Distance de confusion (2/4)

- Chaque phonème est représenté par un vecteur unique

- En utilisant ces vecteurs, on calcule une matrice de confusion avec la distance entre 2 phonèmes
  - On utilise la distance de norme entre 2 vecteurs pour calculer la dissimilarité entre eux: ça représente le nombre de traits différents entre les deux phonèmes

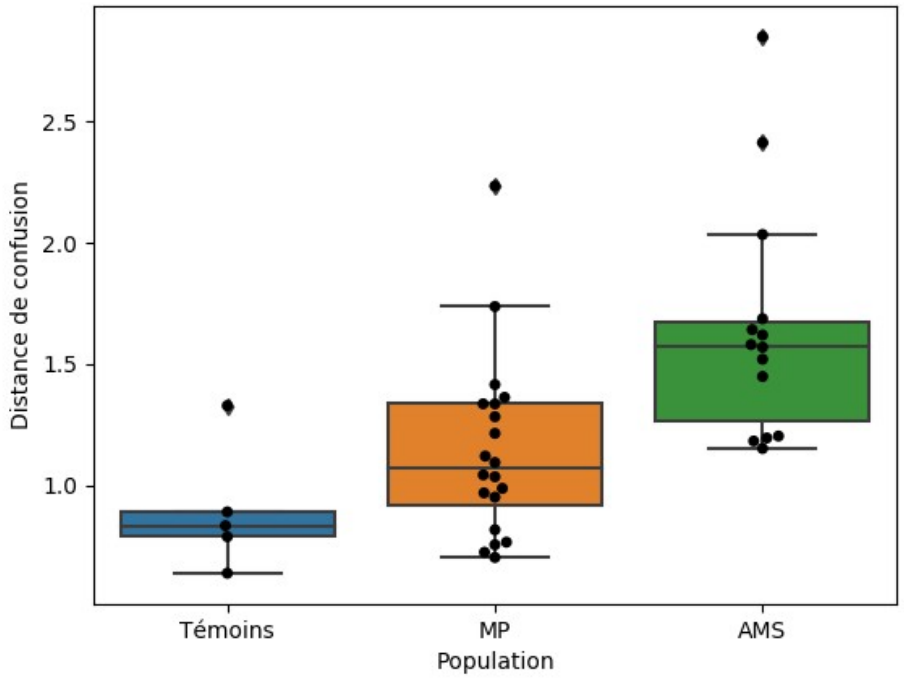
$$\left(d = \sum_i |x_i - y_i|\right).$$

Distance de confusion (3/4)

Phone	@	E	G	N	O	^	a	deux	i	u	un	y	R	S	Z	b	d	f	g	huit	j	k	l	m	n	p	s	t	v	w	z	Sil
@	0	2	2	1	2	2	1	4	4	4	2	5	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6
E	2	0	0	3	2	0	1	2	2	4	2	3	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6
G	2	0	0	3	2	0	1	1	1	3	2	2	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6
N	1	3	3	0	1	3	2	3	5	3	1	4	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6
O	2	2	2	1	0	2	1	2	4	2	2	3	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6
^	2	0	0	3	2	0	1	1	1	3	2	2	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6
a	1	1	1	2	1	1	0	3	3	3	3	4	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6
deux	4	2	1	3	2	1	3	0	2	2	2	1	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6
i	4	2	1	5	4	1	3	2	0	2	4	1	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6
u	4	4	3	3	2	3	3	2	2	0	4	1	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6
un	2	2	2	1	2	2	3	2	4	4	0	3	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6
y	5	3	2	4	3	2	4	1	1	1	3	0	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6
R	7	7	7	7	7	7	7	7	7	7	7	7	0	3	2	3	2	3	3	1	1	4	1	2	2	4	2	3	2	1	1	7
S	7	7	7	7	7	7	7	7	7	7	7	7	3	0	1	4	3	2	2	4	2	1	4	6	5	3	1	2	3	4	2	7
Z	7	7	7	7	7	7	7	7	7	7	7	7	2	1	0	3	2	3	1	3	1	2	3	5	4	4	2	3	2	3	1	7
b	7	7	7	7	7	7	7	7	7	7	7	7	3	4	3	0	1	2	2	2	4	3	2	2	3	1	3	2	1	2	2	7
d	7	7	7	7	7	7	7	7	7	7	7	7	2	3	2	1	0	3	1	3	3	2	1	3	2	2	2	1	2	3	1	7
f	7	7	7	7	7	7	7	7	7	7	7	7	3	2	3	2	3	0	4	4	2	3	4	4	5	1	1	2	1	2	2	7
g	7	7	7	7	7	7	7	7	7	7	7	7	3	2	1	2	1	4	0	4	2	1	2	4	3	3	3	2	3	4	2	7
huit	7	7	7	7	7	7	7	7	7	7	7	7	1	4	3	2	3	2	4	0	2	5	2	2	3	3	3	4	1	0	2	7
j	7	7	7	7	7	7	7	7	7	7	7	7	1	2	1	4	3	4	2	2	0	3	2	4	3	5	3	4	3	2	2	7
k	7	7	7	7	7	7	7	7	7	7	7	7	4	1	2	3	2	3	1	5	3	0	3	5	4	2	2	1	4	5	3	7
l	7	7	7	7	7	7	7	7	7	7	7	7	1	4	3	2	1	4	2	2	2	3	0	2	1	3	3	2	3	2	2	7
m	7	7	7	7	7	7	7	7	7	7	7	7	3	6	5	2	3	4	4	4	2	5	2	0	1	3	5	4	3	2	4	7
n	7	7	7	7	7	7	7	7	7	7	7	7	2	5	4	3	2	5	3	3	3	4	1	1	0	4	4	3	4	3	3	7
p	7	7	7	7	7	7	7	7	7	7	7	7	4	3	4	1	2	1	3	3	5	2	3	3	4	0	2	1	2	3	3	7
s	7	7	7	7	7	7	7	7	7	7	7	7	2	1	2	3	2	1	3	3	3	2	3	5	4	2	0	1	2	3	1	7
t	7	7	7	7	7	7	7	7	7	7	7	7	3	2	3	2	1	2	2	4	4	1	2	4	3	1	1	0	3	4	2	7
v	7	7	7	7	7	7	7	7	7	7	7	7	1	3	2	1	2	1	3	1	3	4	2	3	4	2	2	3	0	1	1	7
w	7	7	7	7	7	7	7	7	7	7	7	7	1	4	3	2	3	2	4	0	2	5	2	2	3	3	3	4	1	0	2	7
z	7	7	7	7	7	7	7	7	7	7	7	7	1	2	1	2	1	2	2	2	2	3	2	4	3	3	1	2	1	2	0	7

Distance de confusion (4/4)

- Distance plus importante chez les patients que les témoins
- Différence entre MP et AMS ($p<0.01$)
- Plus de confusion chez les AMS → distance de confusion plus importante



Conclusions et perspectives

- Recherche de marqueurs issus du traitement automatique de la parole pour le diagnostic différentiel entre la MP et l'AMS.
- Dess indicateurs issus de la reconnaissance automatique de phonèmes, en adéquation avec les observations effectuées par le corps médical, sont pertinents :
    - La **qualité de la reconnaissance phonétique**.
    - La **durée des voyelles** reconnues (plus longues pour l'AMS).
    - Le **débit de parole** (plus lent pour l'AMS).
    - La **distance de confusion** (plus élevé pour l'AMS)
- Étudier l'évolution du **débit local** au cours du temps