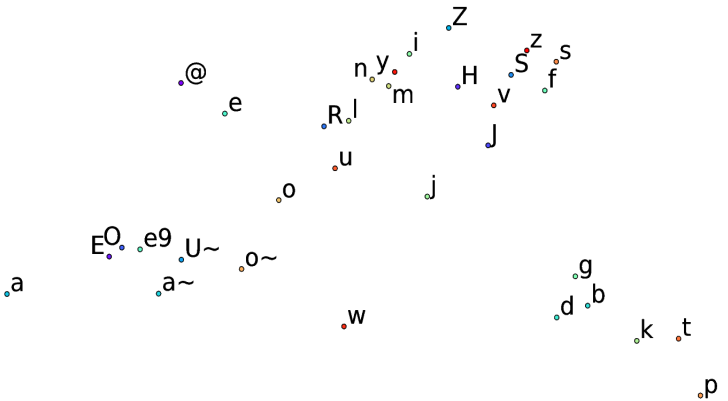
La parole « non-standard » : un défi pour les outils de traitement automatique de la parole

Thomas Pellegrini

Université de Toulouse; UPS; IRIT; Toulouse, France



Plan de la présentation

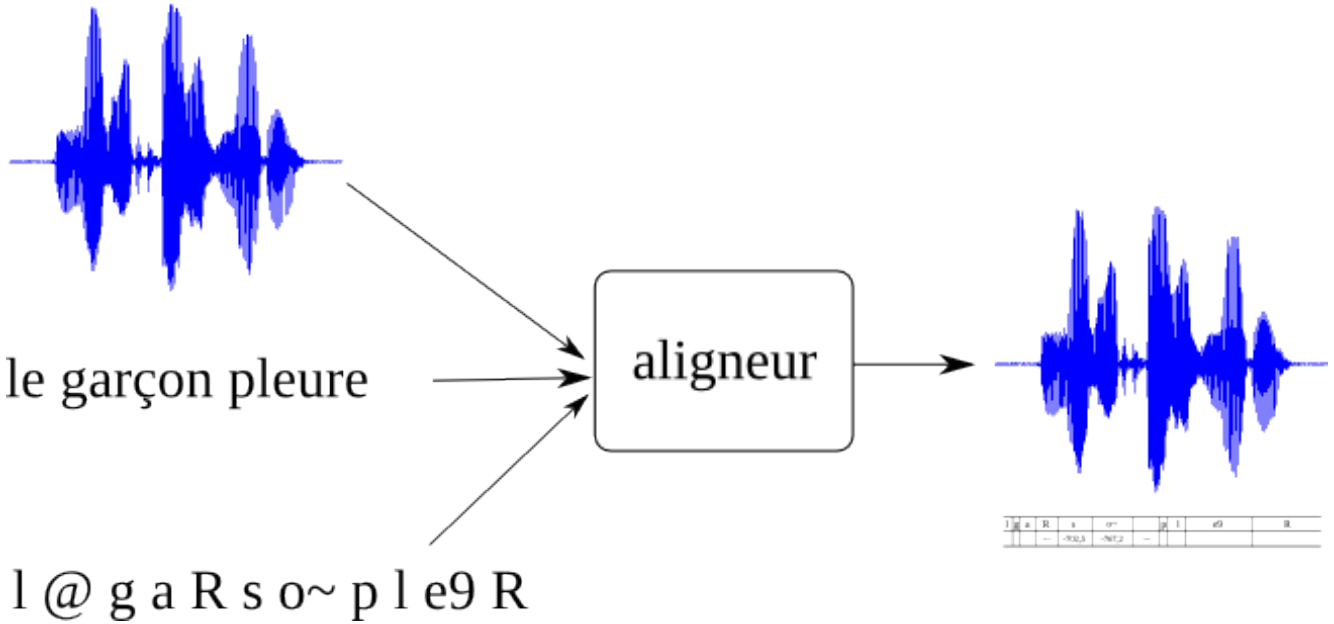
Introduction : quelles outils automatiques pour quels objectifs ?

- Outils de traitement automatique de la parole (TAP) :
  - reconnaissance de la parole
  - alignement automatique
  - détection de mots clés
- Niveau avancé des technologies vocales
- Certes mais faites pour de la parole dite « standard » **du point de vue de la machine** : parole d'adultes sains autour de 30 ans
- Parole « non-standard » ou « atypique » : parole L2, pathologique, d'enfants, de personnes âgées, dans le bruit, sous le coup d'émotions, etc.

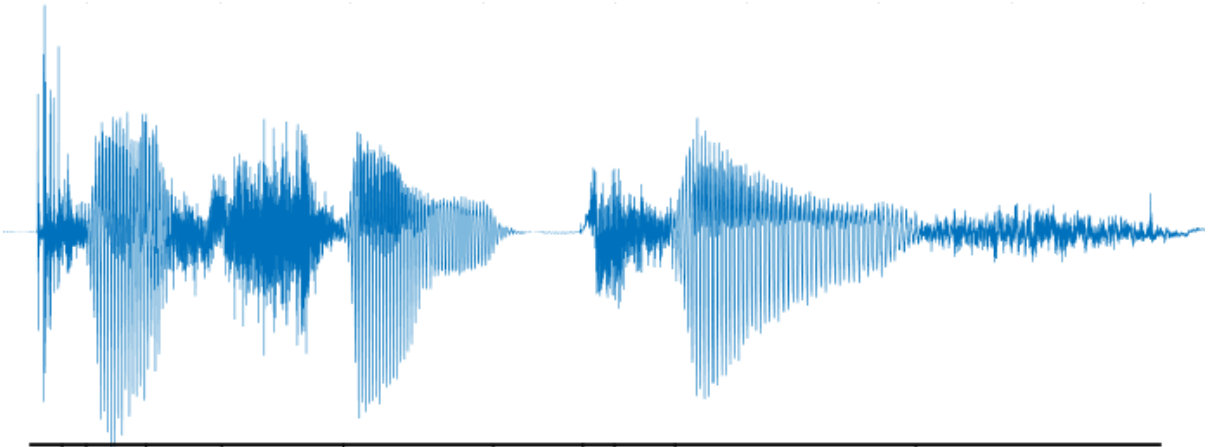
Plan

Caractérisation au niveau segmental : alignement

- Alignements phonétiques d'enregistrements de parole avec des séquences de phonèmes (*phones*) attendus : on connaît le message prononcé à l'avance



Caractérisation au niveau segmental : exemple



l	g	a	R	s	o~		p	l	e9	R
			...	-702,5	-767,2	...				

Le garçon pleure

Caractérisation au niveau segmental : monophones ou triphones

Monophones (phones)

regards R @ g a R

Triphones : un groupe G-X+D

regards R+@ R-@+g @-g+a g-a+R a-R

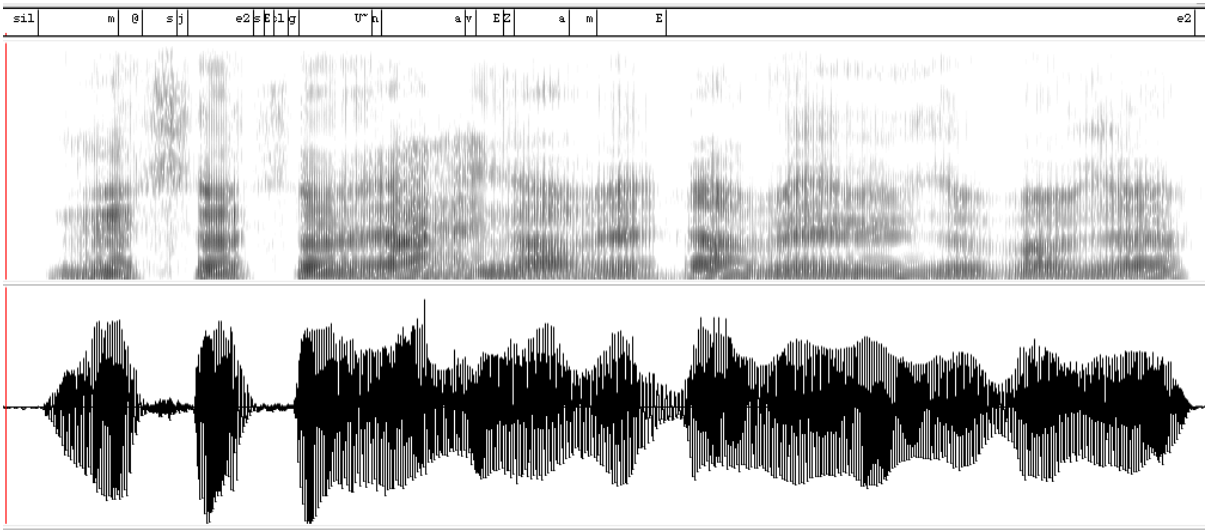
- Lexique de prononciation FR typique : 100000 mots, 37 phones, 5000 triphones
- Choix usuel en parole atypique : monophones

Caractérisation au niveau segmental : liste de logiciels disponibles

- EasyAlign, plugin de Praat (Goldman, 2011)
- SPPAS (Bigi, 2012)
- Train&Align (Brognaux, 2014)
- Praaline (Christodoulides, 2014)
- CLAN (MacWhinney, 2015)
- « plus techniques » : HTK, Julius, Sphinx, Kaldi

Caractérisation au niveau segmental : exemple problématique

Monsieur Séguin n'avait jamais...
Projet C2SI : Carcinologic Speech Severity Index (LPL, PETRA et Octogone, LIA, IRIT)



Caractérisation au niveau segmental : parole dysarthrique

Travaux de Imed Laaridh *et al.* (2015, 2016)

- Corpus Typaloc (Meunier *et al.*, 2015) : 40 enregistrements réalisés par 28 patients dysarthriques et 12 contrôles sains

Caractérisation au niveau segmental : parole dysarthrique

Travaux de Imed Laaridh *et al.* (2015, 2016)

- Corpus Typaloc (Meunier *et al.*, 2015) : 40 enregistrements réalisés par 28 patients dysarthriques et 12 contrôles sains
- 3 pathologies différentes : maladie de Parkinson, une ataxie cérébelleuse, sclérose latérale amyotrophique

Caractérisation au niveau segmental : parole dysarthrique

Travaux de Imed Laaridh *et al.* (2015, 2016)

- Corpus Typaloc (Meunier *et al.*, 2015) : 40 enregistrements réalisés par 28 patients dysarthriques et 12 contrôles sains
- 3 pathologies différentes : maladie de Parkinson, une ataxie cérébelleuse, sclérose latérale amyotrophique

	Contrôles	Park.	Céréb.	SLA
Taux Reco	81%	67%	67%	58%
Fricatives	+ fric.	- occ.	- occ.	+ fric.
Voyelles	+	+	+	- nasales

Mesures issues de la reconnaissance automatique de la parole

- Mesures issues des sorties brutes du moteur de reconnaissance / alignement

Mesures issues de la reconnaissance automatique de la parole

- Mesures issues des sorties brutes du moteur de reconnaissance / alignement
  - Sortie brute : séquence de phones correspondant à la meilleure hypothèse ou bien aux n meilleures hypothèses

Mesures issues de la reconnaissance automatique de la parole

- Mesures issues des sorties brutes du moteur de reconnaissance / alignement
  - Sortie brute : séquence de phones correspondant à la meilleure hypothèse ou bien aux n meilleures hypothèses
  - Score usuel : *Phone Error Rate* (PER)

Mesures issues de la reconnaissance automatique de la parole

- Mesures issues des sorties brutes du moteur de reconnaissance / alignement
  - Sortie brute : séquence de phones correspondant à la meilleure hypothèse ou bien aux n meilleures hypothèses
  - Score usuel : *Phone Error Rate* (PER)
  - Scores plus fins : distance prenant en compte le nombre de traits partagés par 2 phonèmes substitués (Fontan *et al.*, 2014)
- Mesures issues de comparaisons avec des modèles de prononciation

Mesures issues de la reconnaissance automatique de la parole

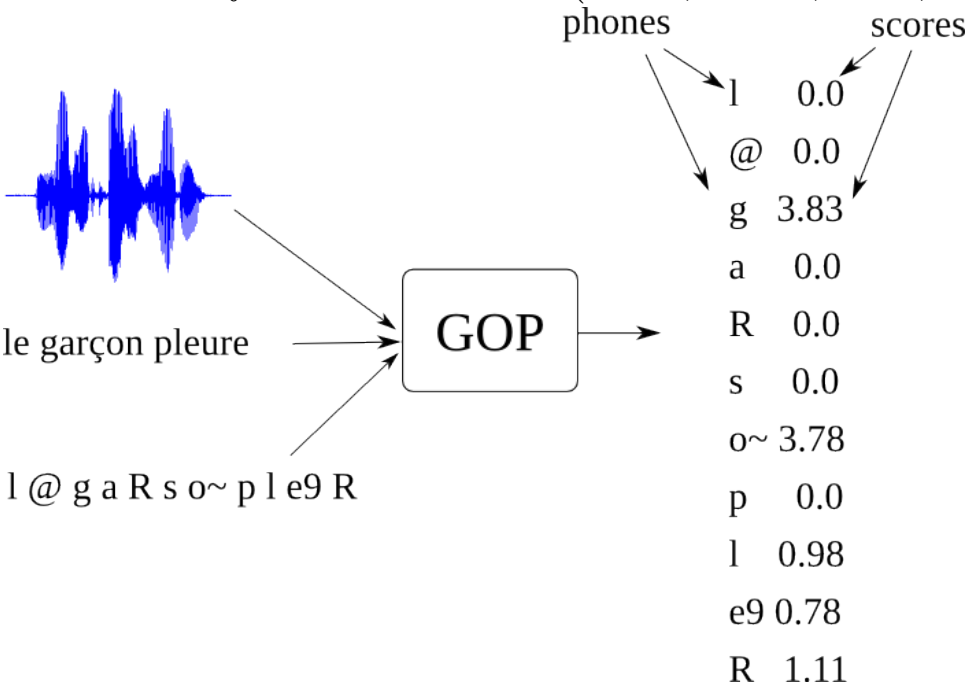
- Mesures issues des sorties brutes du moteur de reconnaissance / alignement
  - Sortie brute : séquence de phones correspondant à la meilleure hypothèse ou bien aux n meilleures hypothèses
  - Score usuel : *Phone Error Rate* (PER)
  - Scores plus fins : distance prenant en compte le nombre de traits partagés par 2 phonèmes substitués (Fontan *et al.*, 2014)
- Mesures issues de comparaisons avec des modèles de prononciation
  - Natifs et non-natifs : *native-likeness*

Mesures issues de la reconnaissance automatique de la parole

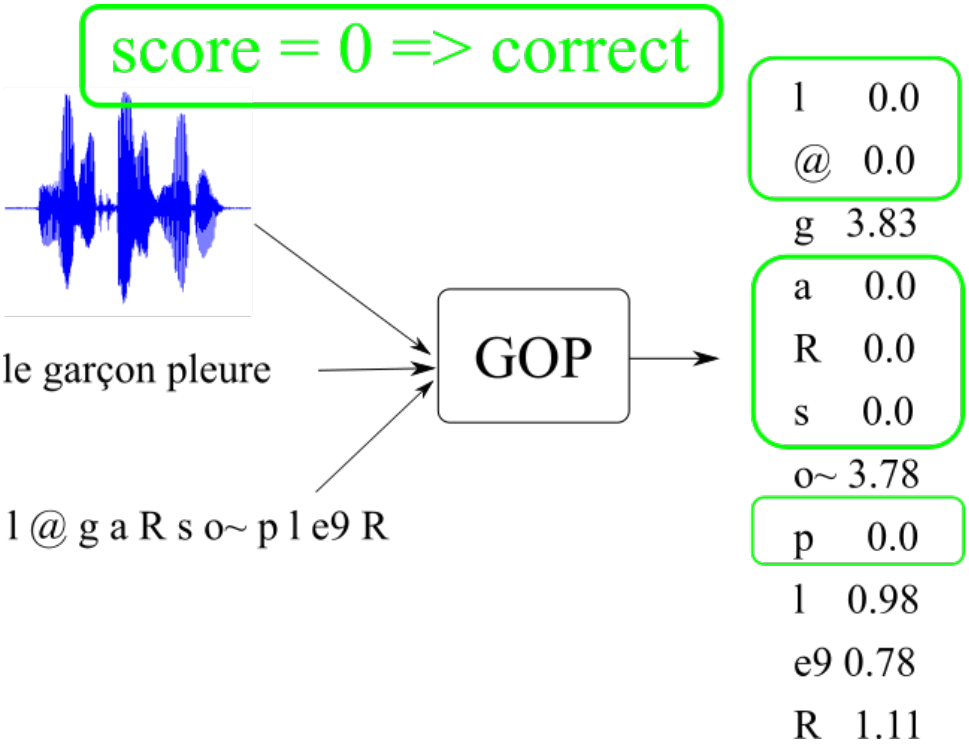
- Mesures issues des sorties brutes du moteur de reconnaissance / alignement
  - Sortie brute : séquence de phones correspondant à la meilleure hypothèse ou bien aux n meilleures hypothèses
  - Score usuel : *Phone Error Rate* (PER)
  - Scores plus fins : distance prenant en compte le nombre de traits partagés par 2 phonèmes substitués (Fontan *et al.*, 2014)
- Mesures issues de comparaisons avec des modèles de prononciation
  - Natifs et non-natifs : *native-likeness*
  - Natifs : ***Goodness of Pronunciation*** – **GOP**

Goodness of Pronunciation – GOP

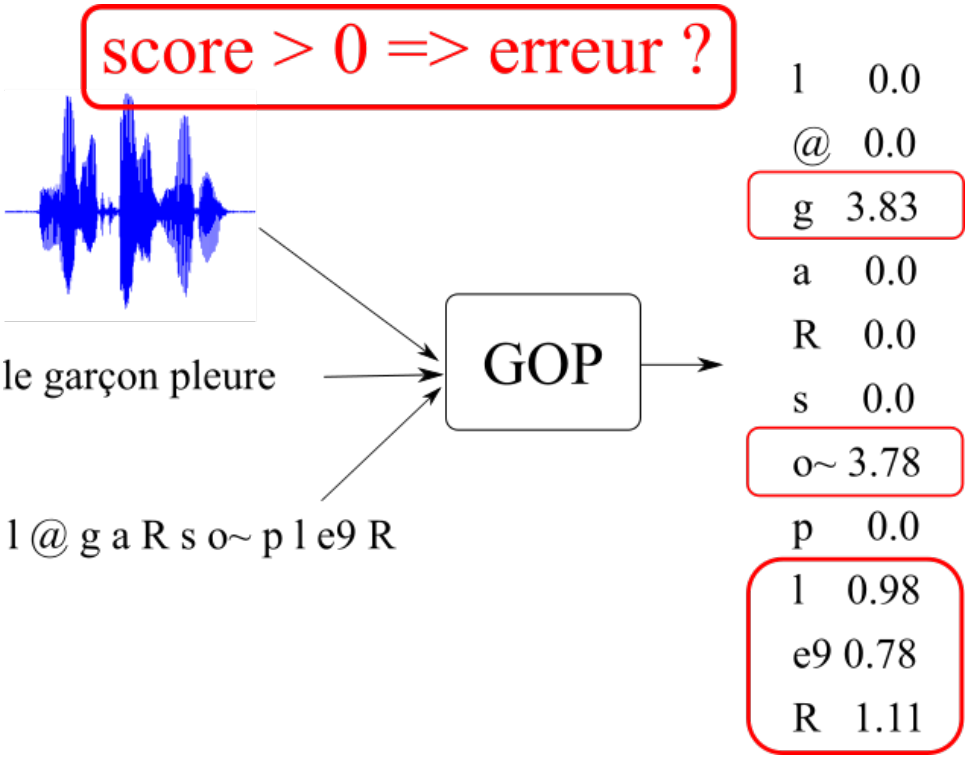
Algorithme *Goodness of Pronunciation* (Witt, 1999 ; Luo, 2009)



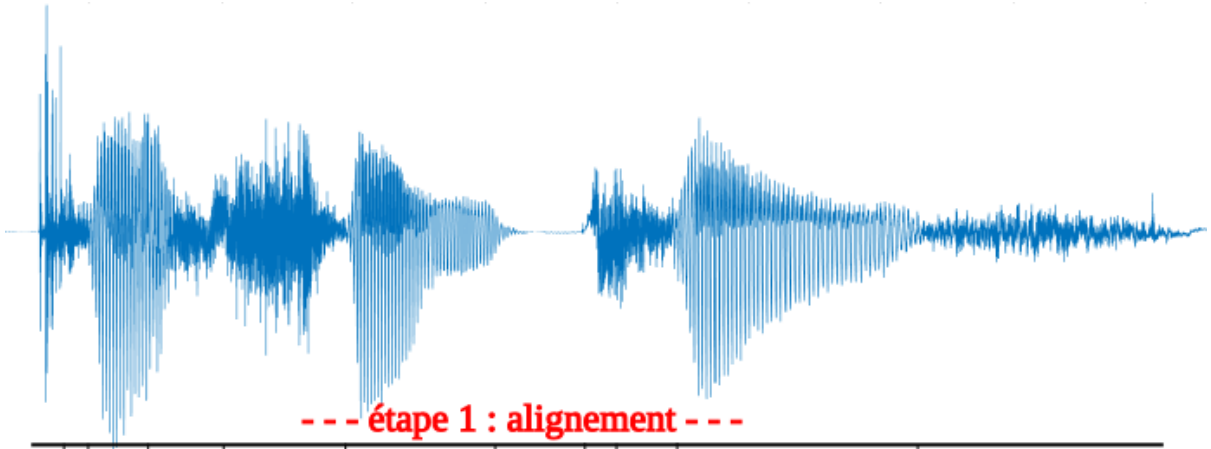
Goodness of Pronunciation – GOP



Goodness of Pronunciation – GOP



Goodness of Pronunciation – GOP



- - - étape 1 : alignement - - -

l	g	a	R	s	**o~**		p	l	e9	R
			···	-702.54	**-767.18**	···				

- - - étape 2 : reconnaissance libre - - -

l	k	a	R	s	**a~**		p	y	e2	U~
			···	-702.54	**-725.51**	···				

- - - étape 3 : calcul des scores GOP - - -

score(o~) = (-725,51+767,18) / 11 = 3,788

Exemple d'étude 1 : locuteurs ayant un paralysie faciale unilatérale

- Objectif : évaluer l'algorithme GOP sur de la parole altérée
- Méthode :

1. Enregistrement d'un corpus de phrases contenant des sons cibles (occlusives)
2. Transcription phonétique manuelle par deux experts
3. Comparaison des prédictions du GOP avec la vérité terrain

Exemple d'étude 1 : locuteurs ayant un paralysie faciale unilatérale

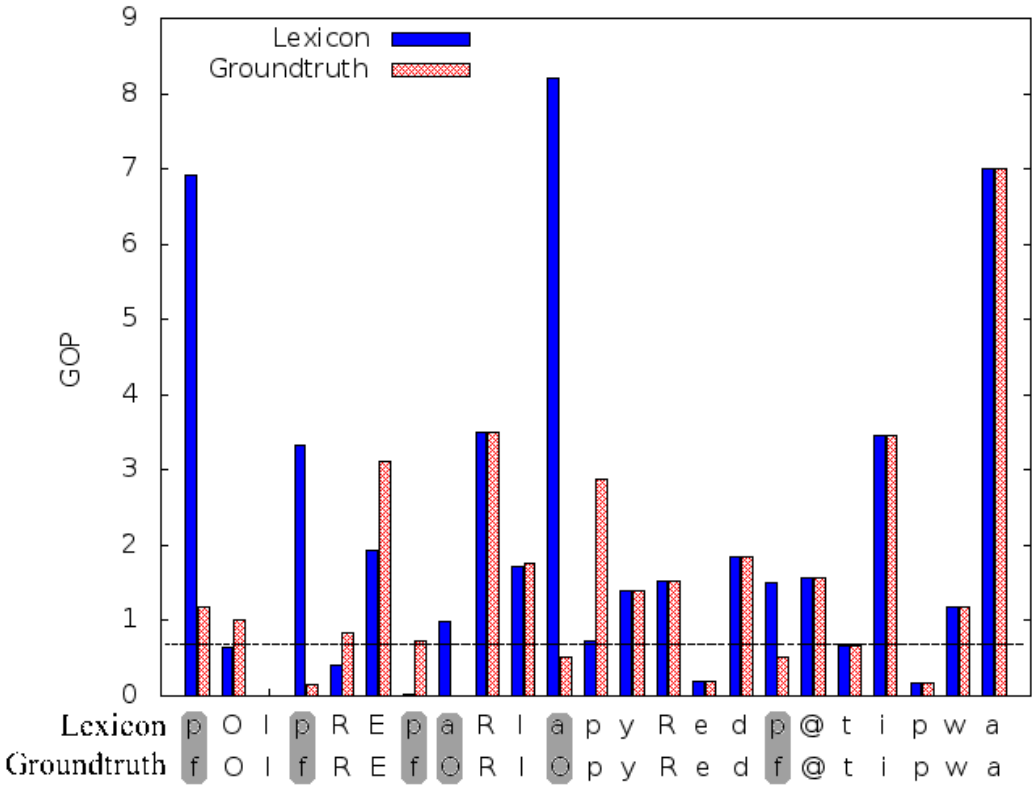
Corpus

- 17 phrases lues par 32 locuteurs
- 4 groupes de sévérité : contrôle, grades House et Brackman III, IV, V et VI.

Analyse auditive

- Difficulté de contrôler les lèvres → mauvais contrôle de l'air
- Plus grand impact sur les consonnes que sur les voyelles :
  - Occlusives /p, b/ souvent perçues comme /f, v/ (explosion difficile)
  - Tendance à produire des variantes rétroflexes de /S, d/ (compensation avec la langue vers le palais)
  - 233 réalisations sur 4000 jugées incorrectes par deux experts (6%)

Paul prépare la purée de petits pois



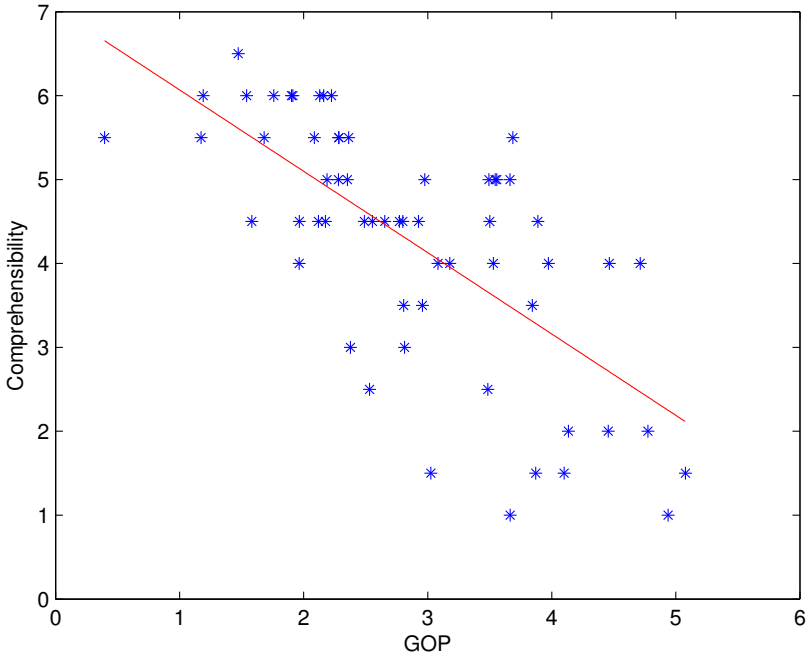
Performance du GOP : détection correcte de 85% et de 50% des prononciations jugées correctes et incorrectes resp.

Exemple d'étude 2 : prédiction de l'intelligibilité

- Objectif : prédire l'intelligibilité/compréhension de la parole altérée
- Corpus :
  - 12 locuteurs avec une pathologie anatomique et / ou neurologique resp.
  - 10 commandes orales aléatoires parmi une centaine
  - « Mettez l'ours à gauche du kangourou »
- Méthode :
  1. Scores subjectifs donnés par deux experts sur une échelle de 1 à 7
  2. Corrélation avec les scores GOP

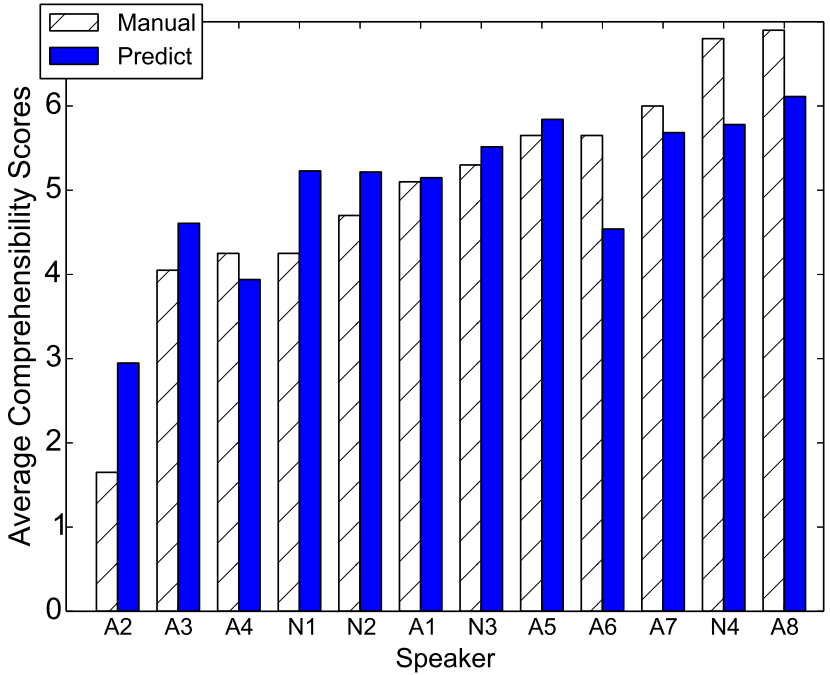
Exemple d'étude 2 : prédiction de l'intelligibilité

Corrélation(GOP, scores humains) = -0,68 ; $p < .001$



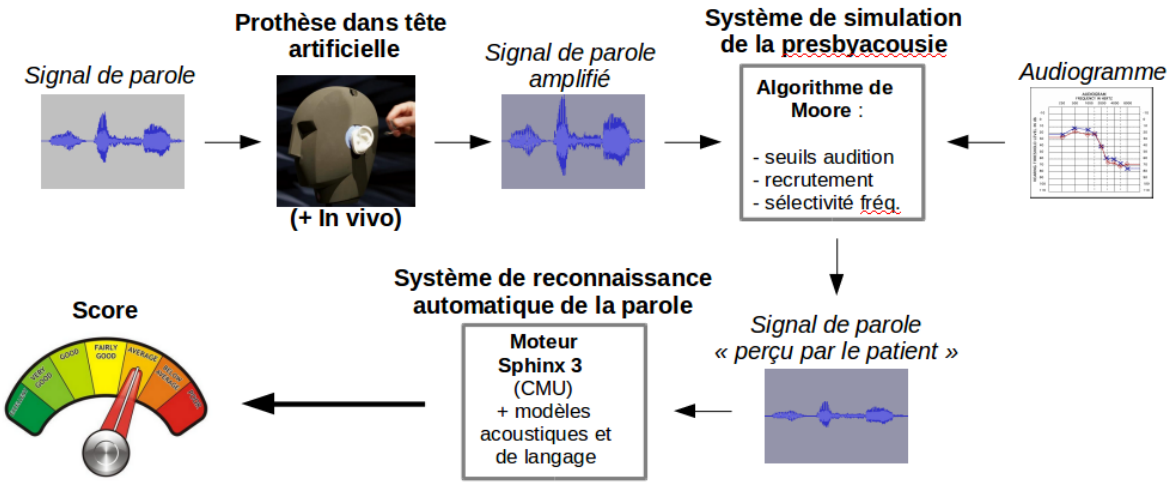
Exemple d'étude 2 : prédiction de l'intelligibilité

- Simple régression linéaire
- Validation croisée : 11 locuteurs / 1 locuteur
- Corrélation : $r = 0,74$
- Erreur relative : 20,9%



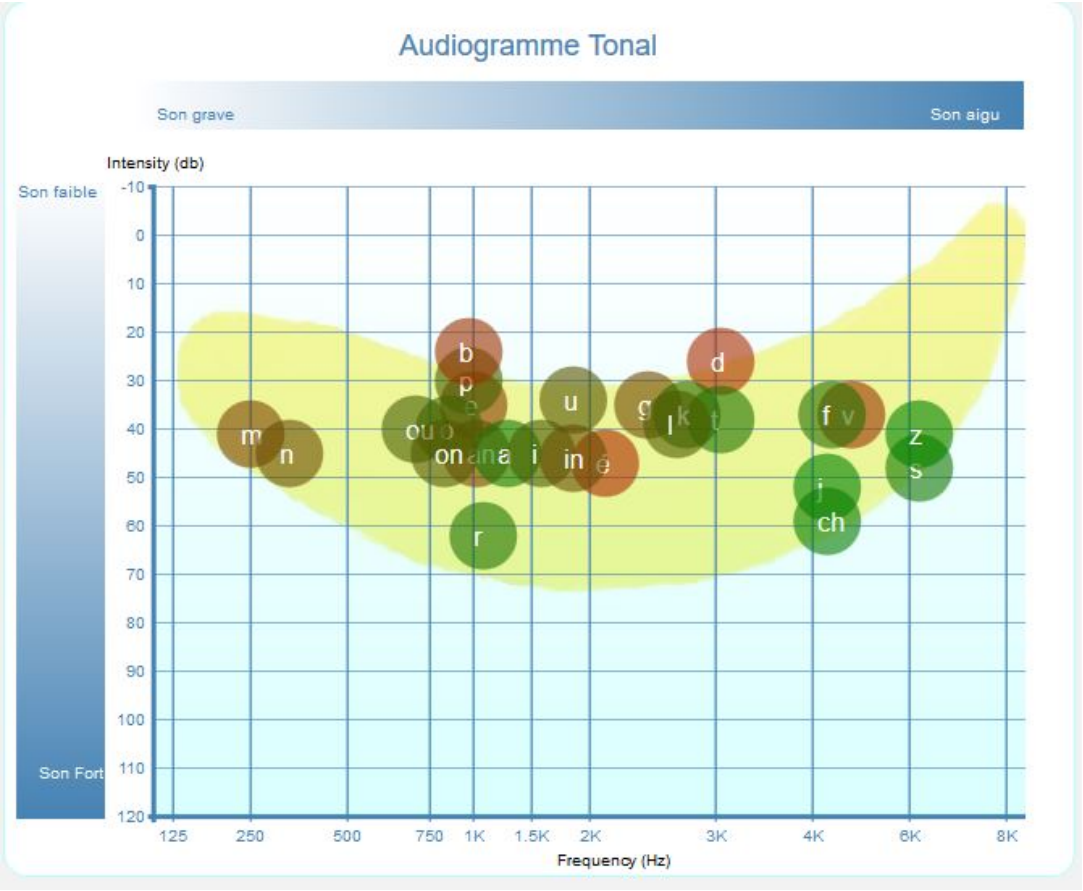
Exemple d'étude 3 : presbyacousie

- Projet ARCHEAN 2013-2015 : Mesure de la compréhension de la parole (ARCHEAN Technologies, Octogone PETRA et IRIT-SAMoVA)
- Objectif : Outil d'aide au réglage de prothèses auditives

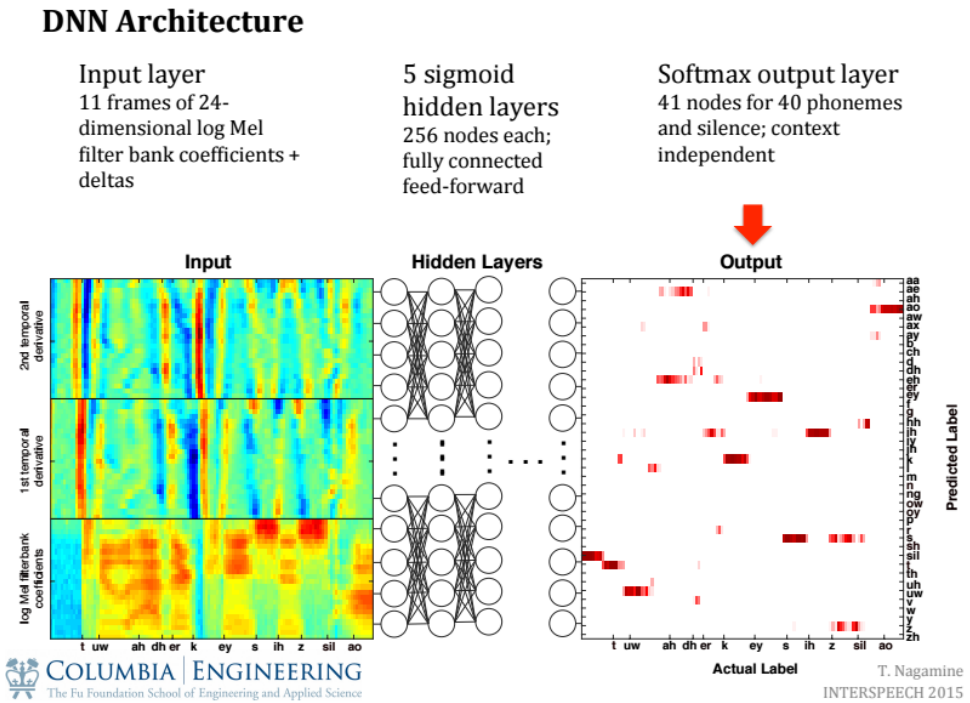


Exemple d'étude 3 : presbyacousie

Représentation : « banane » de la parole

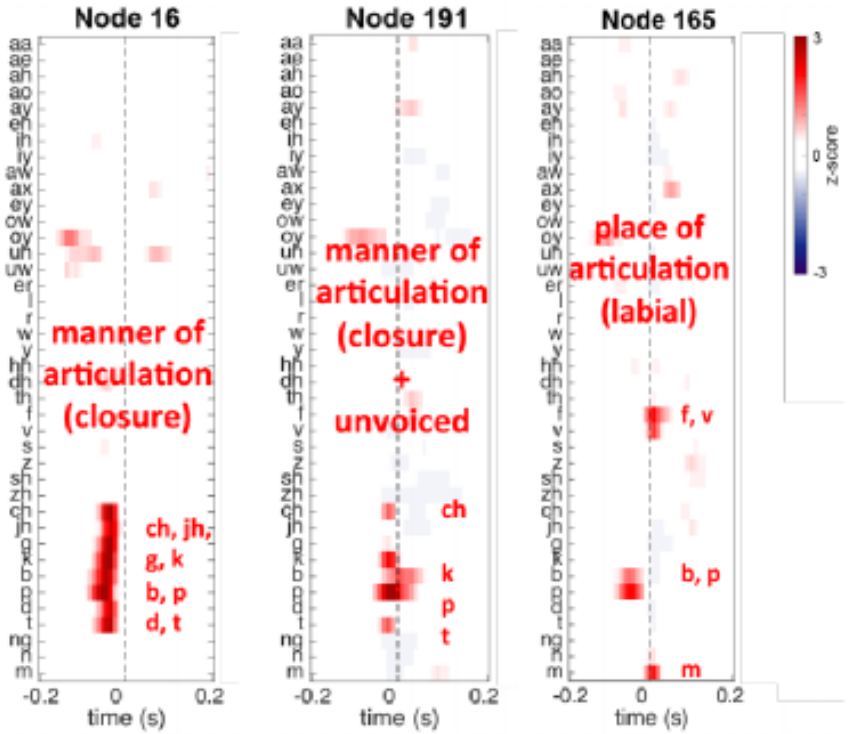


Avancées récentes



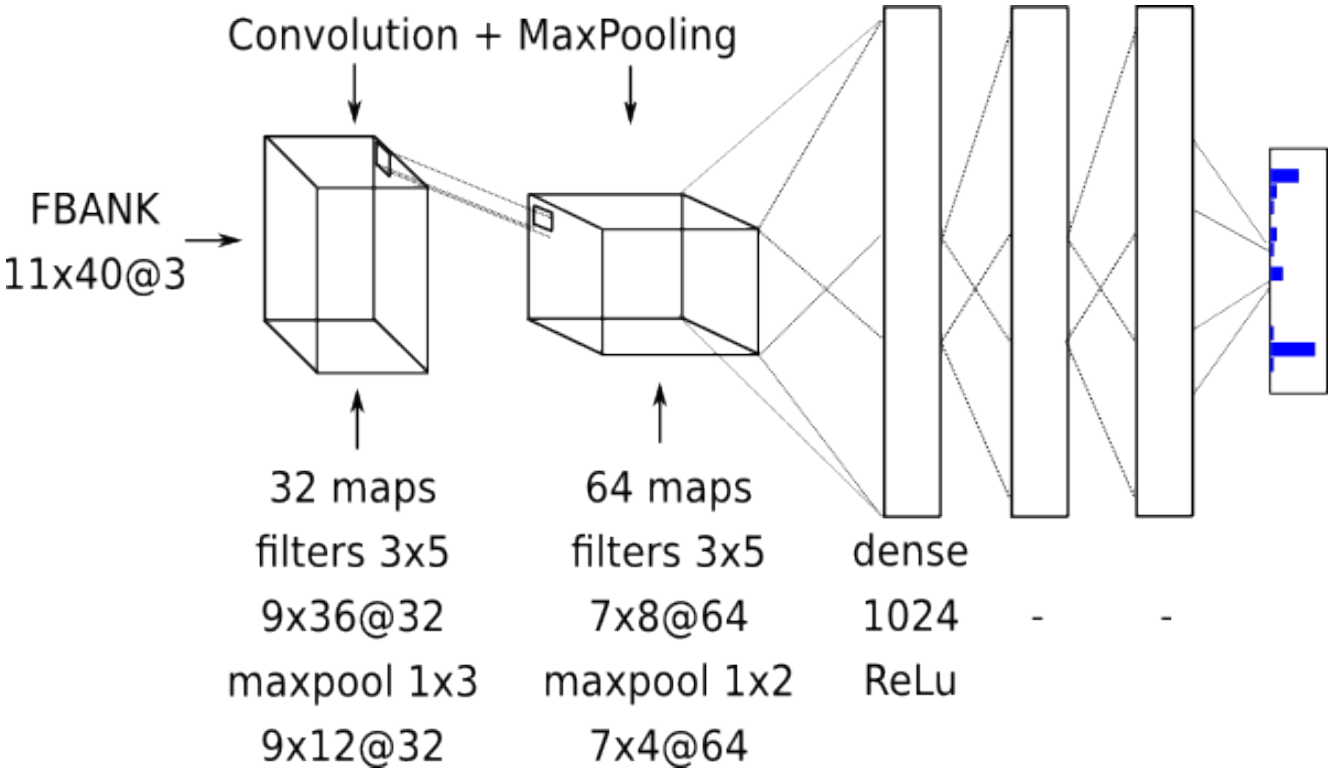
Source : Nagamine et al. Exploring How Deep Neural Networks Form Phonemic Categories. INTERSPEECH 2015

Avancées récentes

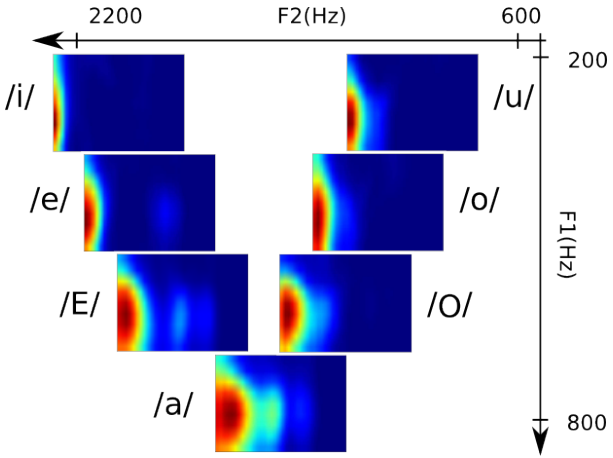


Source : Nagamine et al. Exploring How Deep Neural Networks Form Phonemic Categories. INTERSPEECH 2015

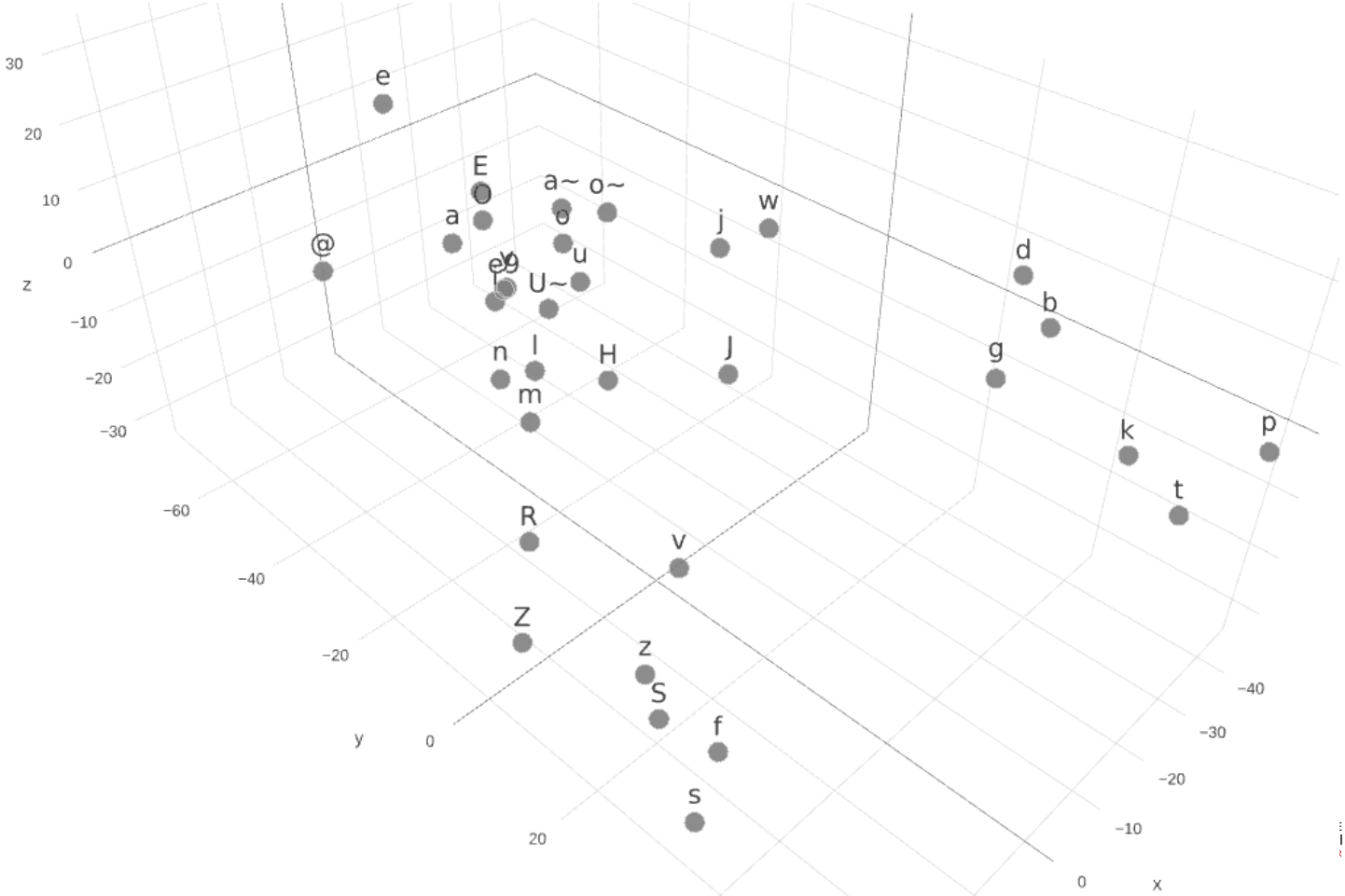
Avancées récentes



Avancées récentes



Avancées récentes



Plan

Caractérisation au niveau suprasegmental : prosodie

Analyse structurales

- Syntaxiques, lexicales, morphologiques
- Prosodiques :
  - Pauses : durée et place
  - Intonation : F0
  - Jitter, shimmer
  - Nombre de reformulations / hésitations
  - Débit articulatoire : nombre de syllabes par seconde
  - Proportion des durées vocaliques
  - ratio d'énergie harmoniques-bruit

- AuToBI (Rosenberg, 2011) : annotation automatique de contours prosodiques
- WinPitch LTL (Martin, 2005) : apprentissage de la prosodie
- VocaLab (Sicard, 2001-2016) : logiciel destiné aux orthophonistes

Caractérisation au niveau suprasegmental : prosodie

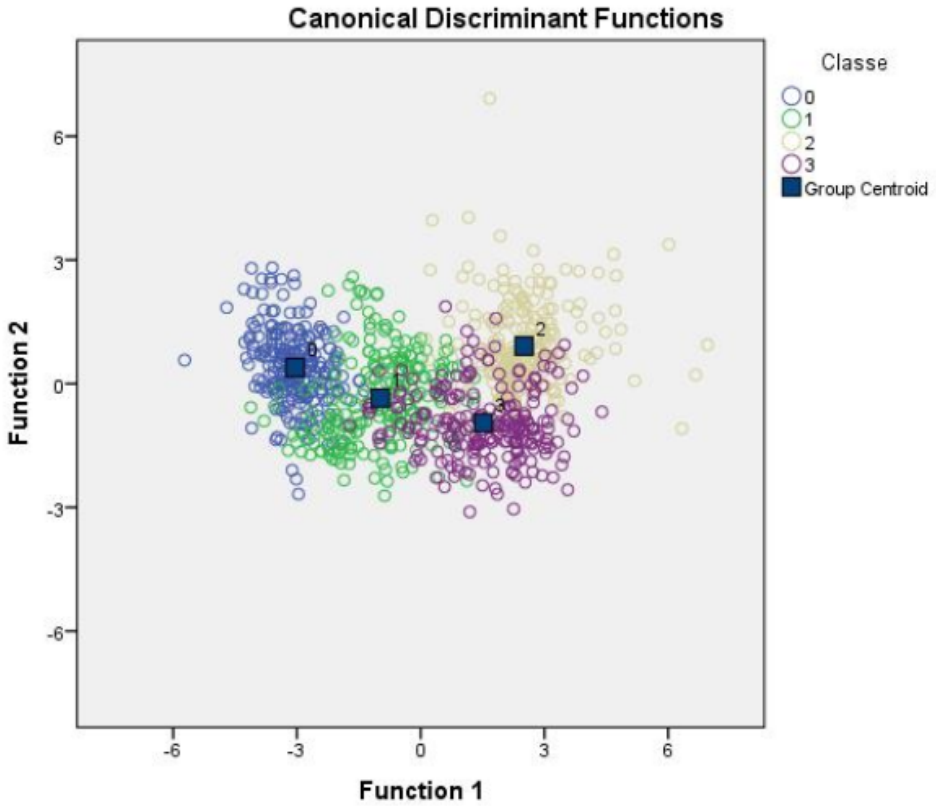
Kadi, *et al.* Automatic Assessment of the Dysarthria Severity Levels, dans Proc. LCPTS, 2015

WILK'S LAMBDA OF THE ACOUSTICS FEATURES

Feature	Wilks' lambda
Articulation rate	0.565
Number of period	0.595
Mean pitch	0.701
Voice breaks	0.835
%V	0.861
HNR	0.864
Jitter	0.925
Shimmer	0.962
Std Pitch	0.979
Std Period	0.984
NHR	0.989

Caractérisation au niveau suprasegmental : prosodie

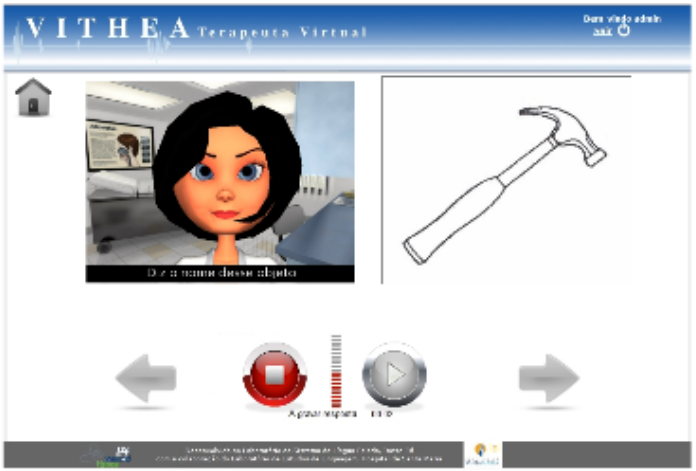
Kadi, *et al.* Automatic Assessment of the Dysarthria Severity Levels, dans Proc. LCPTS, 2015



Aide à la réalisation de tests de fluence lexicale

- Projet VITHEA : An on-line system for virtual treatment of aphasia (INESC-ID, Portugal)
- Objectif : aide à la réhabilitation de patients aphasiques
- Différents exercices de fluence lexicale : listes d'objets, d'animaux ; nommer des images, des proverbes, etc.
- Technique : détection automatique de mots clés

Aide à la réalisation de tests de fluence lexicale



Plan

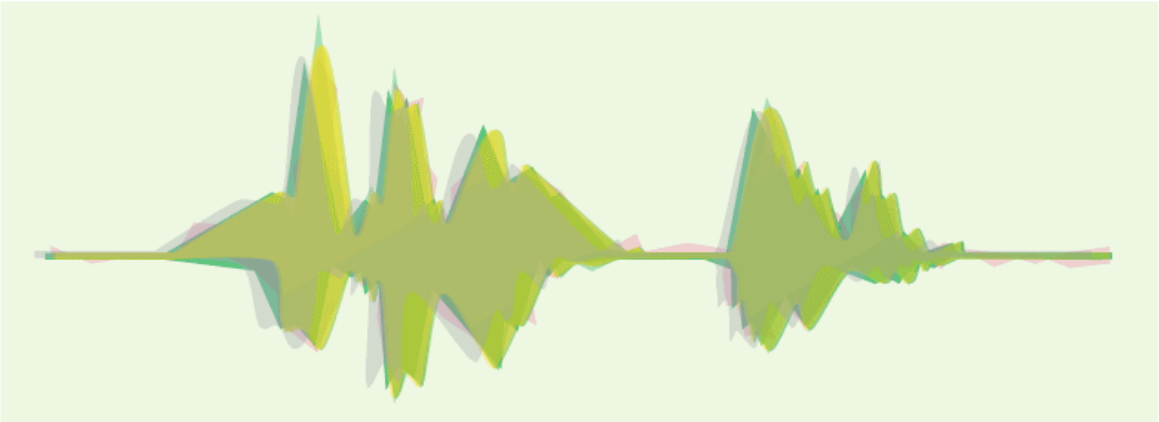
Conclusion et perspectives

- Deux principaux objectifs :
  - Améliorer la qualité de vie des locuteurs ayant une pathologie de la parole
  - Aider les praticiens en facilitant les tâches : scores et statistiques générés automatiquement

Conclusion et perspectives

- Les techniques de RAP sont aujourd'hui suffisamment matures pour fournir des indices rapides et objectifs de production et de réception de la parole
- Néanmoins encore beaucoup de travail à faire...
- Des progrès sont nécessaires en reconnaissance de parole atypique : revenir au plus prêt du signal ?
- De récents travaux suggèrent que les outils automatiques se rapprochent de la perception humaine du point de vue qualitatif aussi (ex. l'étude de systèmes de réseaux neuronaux profonds de Nagamine *et al.*, 2015), et les recherches se concentrent sur ce sujet (ex. réseau Marie Curie *Inspire*)
- Collaboration entre informaticiens, linguistes, médecins, et usagers fondamentale

Atelier TALPA



[at@lie TALPA]

Traitement Automatique de la Parole Atypique

4 juillet 2016
Paris, France

Références

Cas d'étude 1 : locuteurs ayant un paralysie faciale unilatérale

- T. Pellegrini *et al.* Automatic Assessment of Speech Capability Loss in Disordered Speech, in ACM Transactions on Accessible Computing, ACM, Special Issue on Speech and Language Processing for AT (Part 1), Vol. 6 N. 3, May 2015
- T. Pellegrini *et al.* The Goodness of Pronunciation algorithm applied to disordered speech. In Proc. *INTERSPEECH*, Singapour, p. 1463-1467, Sept. 2014

Cas d'étude 2 : prédiction de l'intelligibilité

- L. Fontan *et al.* Predicting disordered speech comprehensibility from Goodness of Pronunciation scores, in Workshop on Speech and Language Processing for Assistive Technologies (SLPAT 2015), Dresden, p. 1-5, Sept. 2015

Vielmols merci !