

Vers une représentation automatique du rythme de la parole



Jérôme Farinas – Corine Astésano



AfIA

Association française
pour l'Intelligence Artificielle



*Journée AfIA-TLH et AFCP – 11 décembre 2023 – "Extraction de
connaissances interprétables pour l'étude de la communication parlée" –
Avignon Université, Campus Hannah Arendt*

Objectifs de cette présentation

1. Détailler le cadre théorique prosodique utilisé
2. Présenter une représentation du rythme développée dans le cadre de la caractérisation prosodique de la parole pathologique
3. Analyse des perspectives :
 - a) Consolider la représentation en incluant une représentation intonative
 - b) Automatisation de la détection des niveaux prosodiques
 - c) Etudier l'évolution temporelle de la représentation
 - d) Envisager l'extraction des paramètres stables pour caractériser le rythme



Objectifs de cette présentation

1. Détailler le cadre théorique prosodique utilisé
2. Présenter une représentation du rythme développée dans le cadre de la caractérisation prosodique de la parole pathologique
3. Analyse des perspectives :
 - a) Consolider la représentation en incluant une représentation intonative
 - b) Automatisation de la détection des niveaux prosodiques
 - c) Etudier l'évolution temporelle de la représentation
 - d) Envisager l'extraction des paramètres stables pour caractériser le rythme



Matérialisation du rythme

Structuration de la parole via :

- Matérialité **phonétique**
 - Fréquence fondamentale (F0)
 - Intensité
 - Durée
- Matérialité **phonologique**
 - Structuration en **unités hiérarchiques**

Structuration prosodique

Niveaux de structuration hiérarchique :

- Le syntagme intonatif (IP, *Intonational phrase*)
-
-

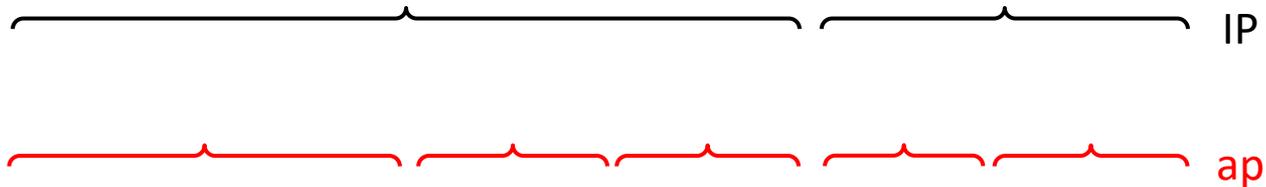


L'ordinateur portable de Gabriel est cassé. Il faudra le changer.

Structuration prosodique

Niveaux de structuration hiérarchique :

- Le syntagme intonatif (IP, *Intonational phrase*)
-
- Le syntagme accentuel (ap, *accentual phrase*) (Jun & Fougeron, 2000)

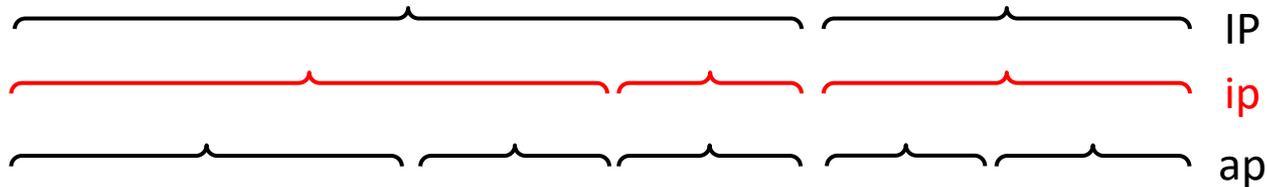


L'ordinateur portable de Gabriel est cassé. Il faudra le changer.

Structuration prosodique

Niveaux de structuration hiérarchique :

- Le syntagme intonatif (IP, *Intonational phrase*)
- Le syntagme intermédiaire (ip, *intermediate phrase*) (Michelas & D'Imperio, 2010)
- Le syntagme accentuel (ap, *accentual phrase*) (Jun & Fougeron, 2000)

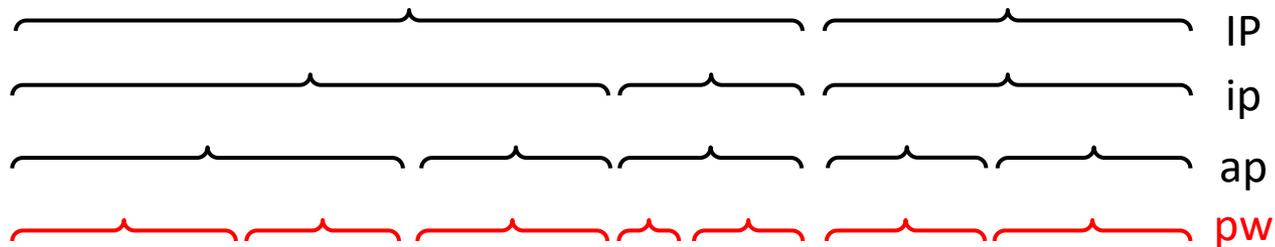


L'ordinateur portable de Gabriel est cassé. Il faudra le changer.

Structuration prosodique

Niveaux de structuration hiérarchique :

- Le syntagme intonatif (IP, *Intonational phrase*)
- Le syntagme intermédiaire (ip, *intermediate phrase*) (Michélas & D'Imperio, 2010)
- Le syntagme accentuel (ap, *accentual phrase*) (Jun & Fougeron, 2000)
- Le mot prosodique (pw, *prosodic word*) (Astésano 2019)



L'ordinateur portable de Gabriel est cassé. Il faudra le changer.

Objectifs de cette présentation

1. Détailler le cadre théorique prosodique utilisé
2. Présenter une représentation du rythme développée dans le cadre de la caractérisation prosodique de la parole pathologique
3. Analyse des perspectives :
 - a) Consolider la représentation en incluant une représentation intonative
 - b) Automatisation de la détection des niveaux prosodiques
 - c) Etudier l'évolution temporelle de la représentation
 - d) Envisager l'extraction des paramètres stables pour caractériser le rythme



Problématique de l'étude voix pathologiques

La caractérisation automatique du rythme de la parole pathologique

- On souhaite utiliser / créer des outils totalement automatiques pour caractériser le prosodie des patients
- Analyser comment se comporte le rythme des personnes atteintes de pathologies de la voix
 - Quel impact sur la fluence des patients ?
 - Compensent-ils leurs pertes d'articulation par une prosodie particulière ?



Cadre de l'étude voix patho

- **Projet ANR RUGBI 2019-2023** (*“Relevant linguistic Units to improve the intelligiBility measurement of speech production disorders”*)

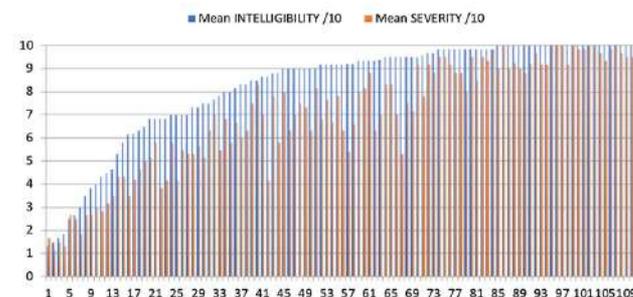
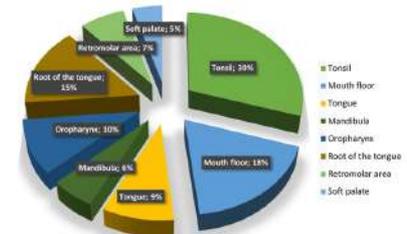
- Complémenter et expliciter la mesure du déficit d'intelligibilité de la parole
 - Cancer des Voies Aéro Digestives Supérieures (VADS)
 - Maladie de Parkinson (Park)
- Recherche automatique d'unités prosodiques pertinentes



- **Groupement d'Intérêt Scientifique PAROLOTHEQUE** depuis 2022 : accès facilité aux données cliniques et métadonnées (perceptives et automatiques)

- **Données :**

- **cancer ORL** : 87 patients et 26 témoins
annotations : sévérité, intelligibilité, prosodie...
- **Maladie de Parkinson** : 205 patients et 111 témoins
annotations : sévérité, intelligibilité...



Doctorat Robin Vaysse



THÈSE

En vue de l'obtention du
DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE
Délivré par l'Université Toulouse 3 - Paul Sabatier

Présentée et soutenue par
Robin VAYSSE

Le 21 mars 2023

**Caractérisation automatique du rythme de la parole : application
aux cancers des voies aéro-digestives supérieures et à la maladie
de Parkinson**

Ecole doctorale : **EDMITT - Ecole Doctorale Mathématiques, Informatique et
Télécommunications de Toulouse**

Spécialité : **Informatique et Télécommunications**

Unité de recherche :

IRIT : Institut de Recherche en Informatique de Toulouse

Thèse dirigée par

Jérôme FARINAS et Corine ASTESANO

Jury

M. François PELLEGRINO, Rapporteur
Mme Elisabeth DELAIS-ROUSSARIE, Rapporteur
Mme Cécile FOUGERON, Examinatrice
Mme Virginie WOISARD-BASSOLS, Examinatrice
M. Jérôme FARINAS, Directeur de thèse
Mme Corine ASTESANO, Co-directrice de thèse



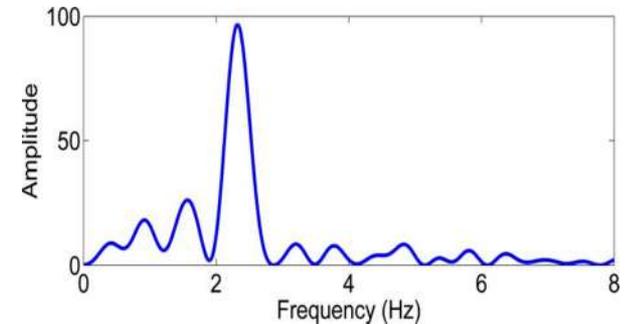
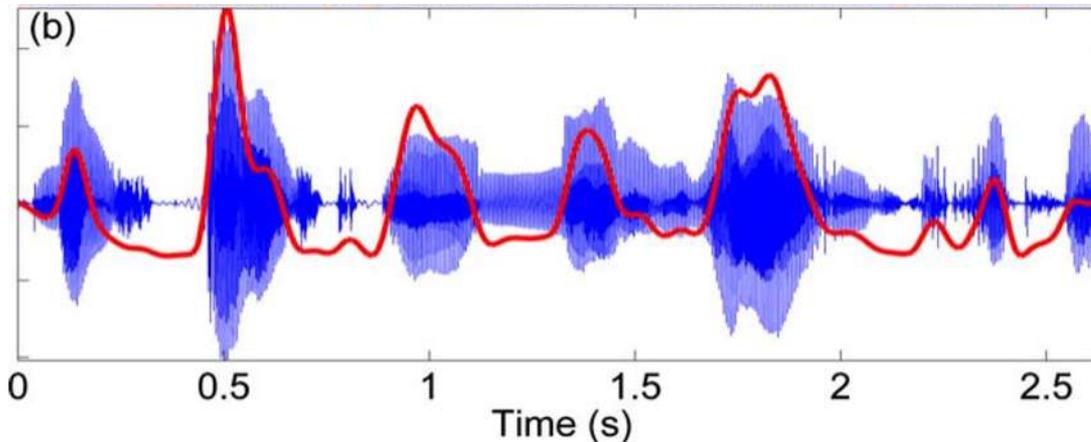
Spectre de modulation d'amplitude (1/3)

- Extraction de l'enveloppe d'amplitude du signal

(Tilsen & Johnson 2008)

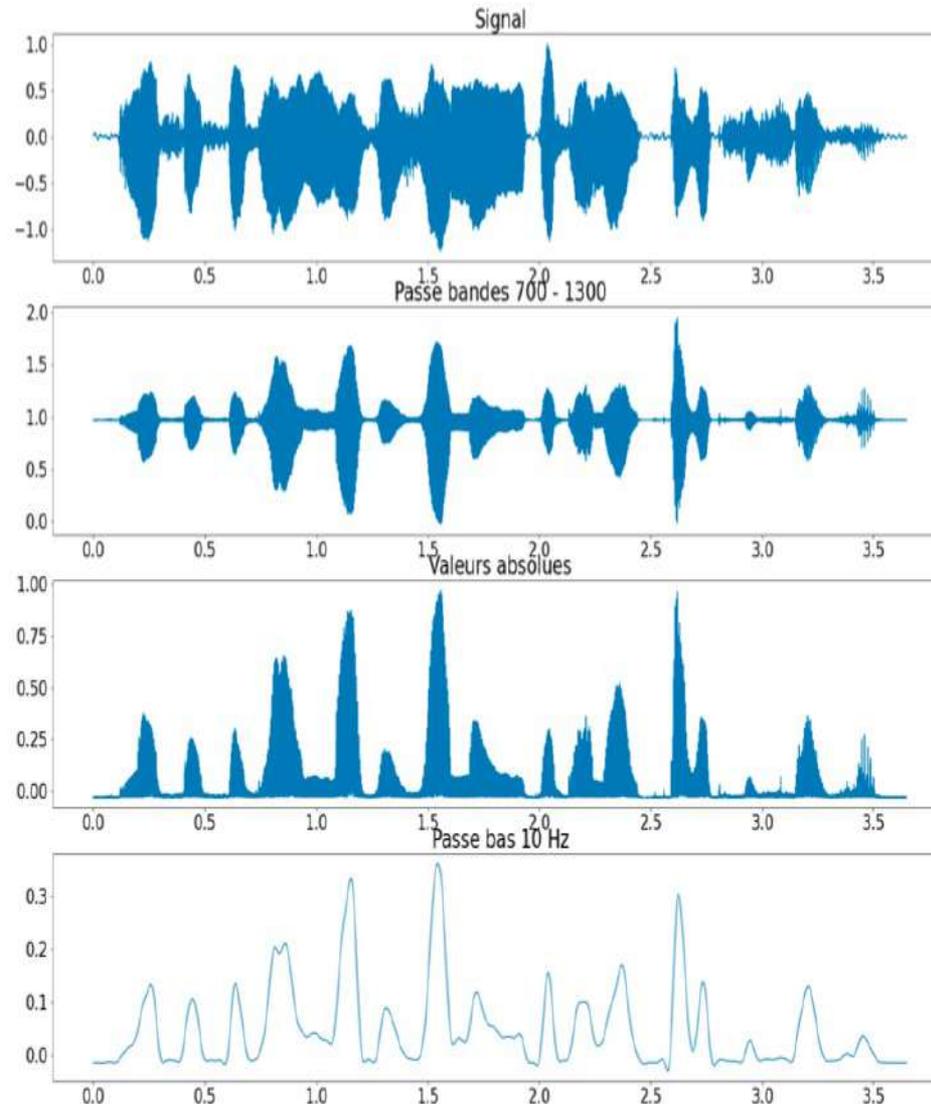
- Filtre **passé bande** 700 - 1300 Hz
- **Rectification** du signal
- Filtre **passé bas** à 10 Hz

- Calcul de la transformée de Fourier de l'enveloppe (**EMS**, *Envelope Modulations Spectrum*)

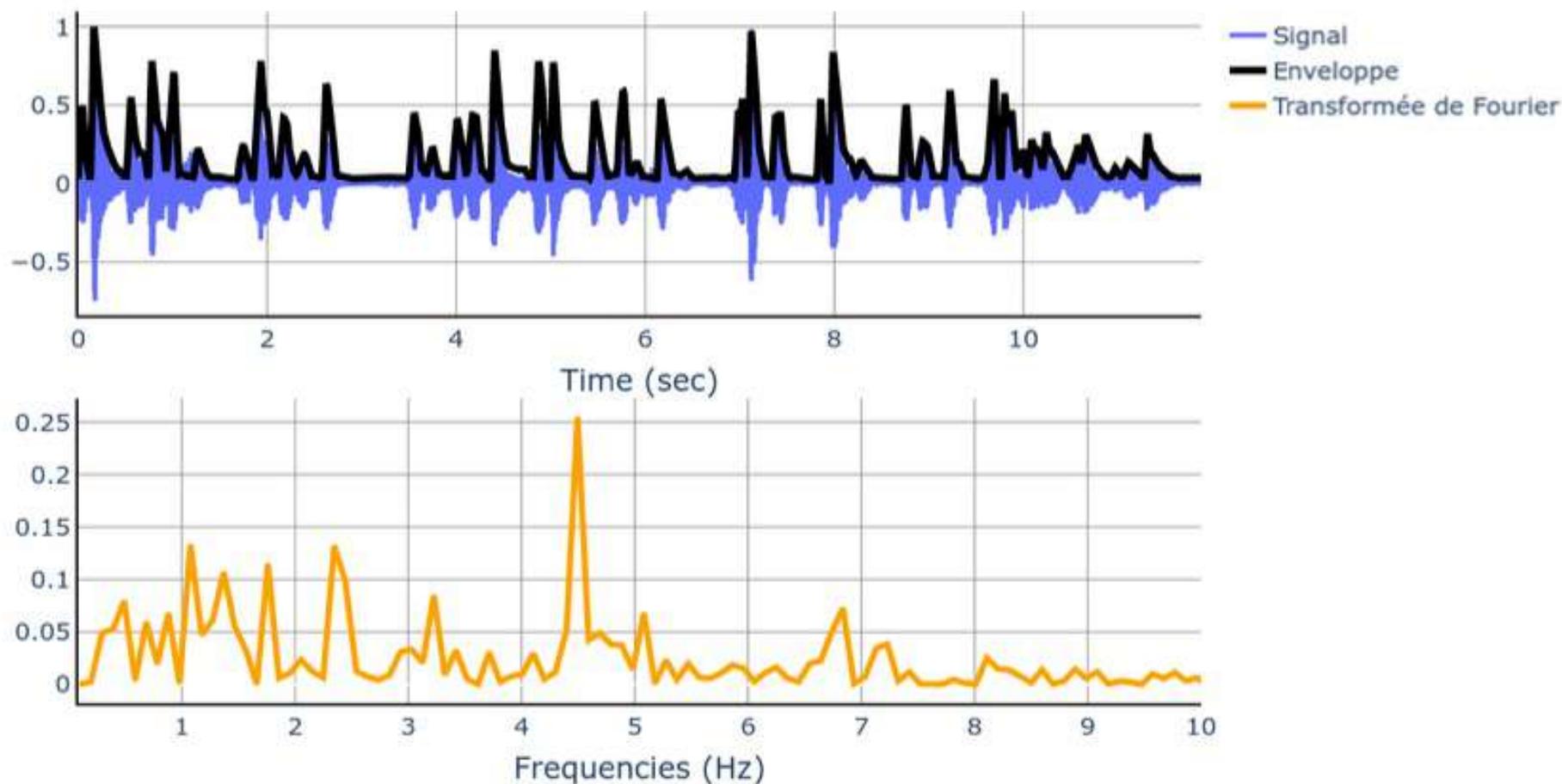


Tilsen, Sam & Johnson, Keith. (2008). Low-frequency Fourier analysis of speech rhythm. The Journal of the Acoustical Society of America. p.124. EL34-9. 10.1121/1.2947626.

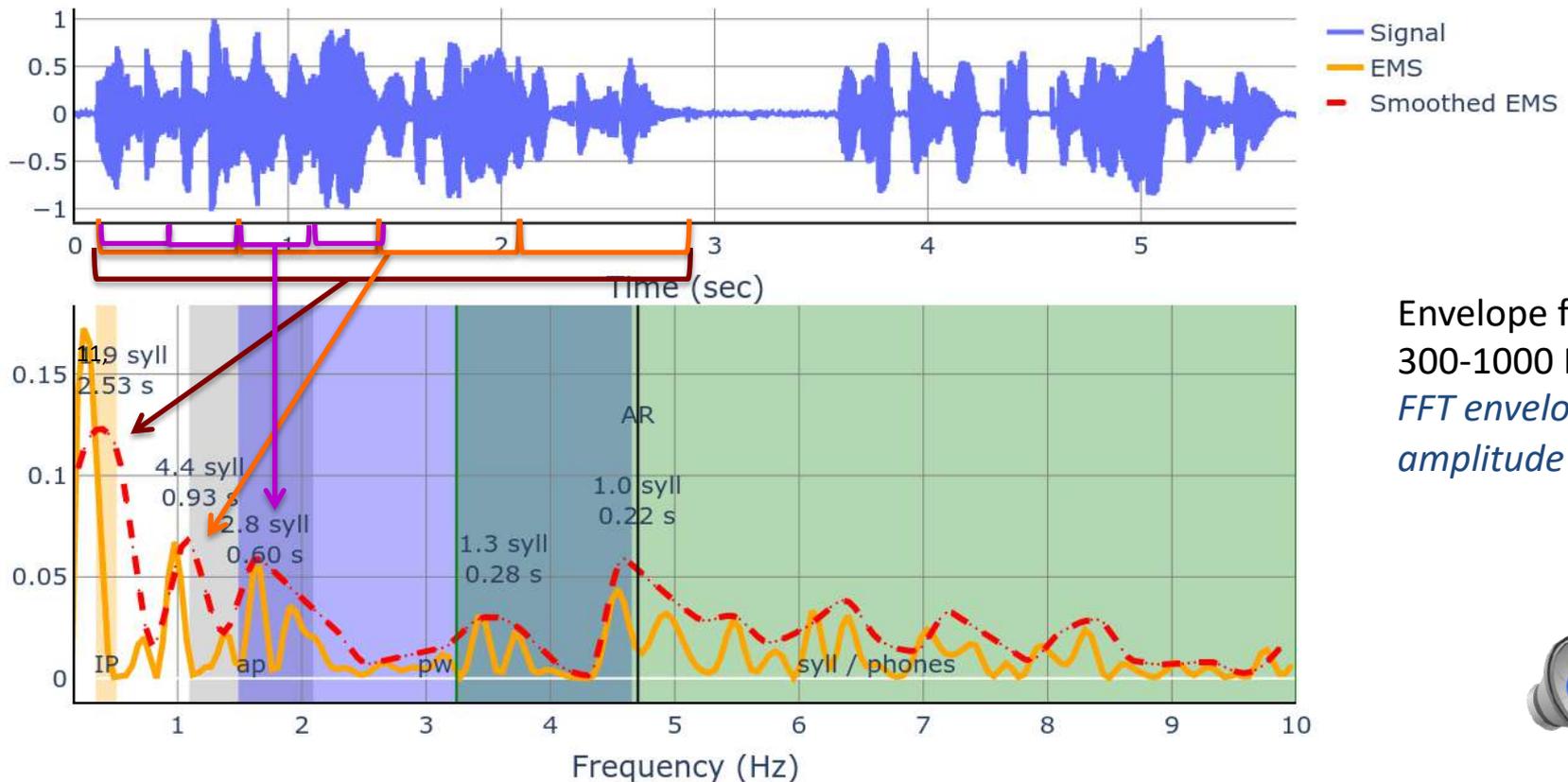
Spectre de modulation d'amplitude (2/3)



Spectre de modulation d'amplitude (3/3)



Exemple 1 - spectre de modulation d'amplitude (locuteur contrôle)

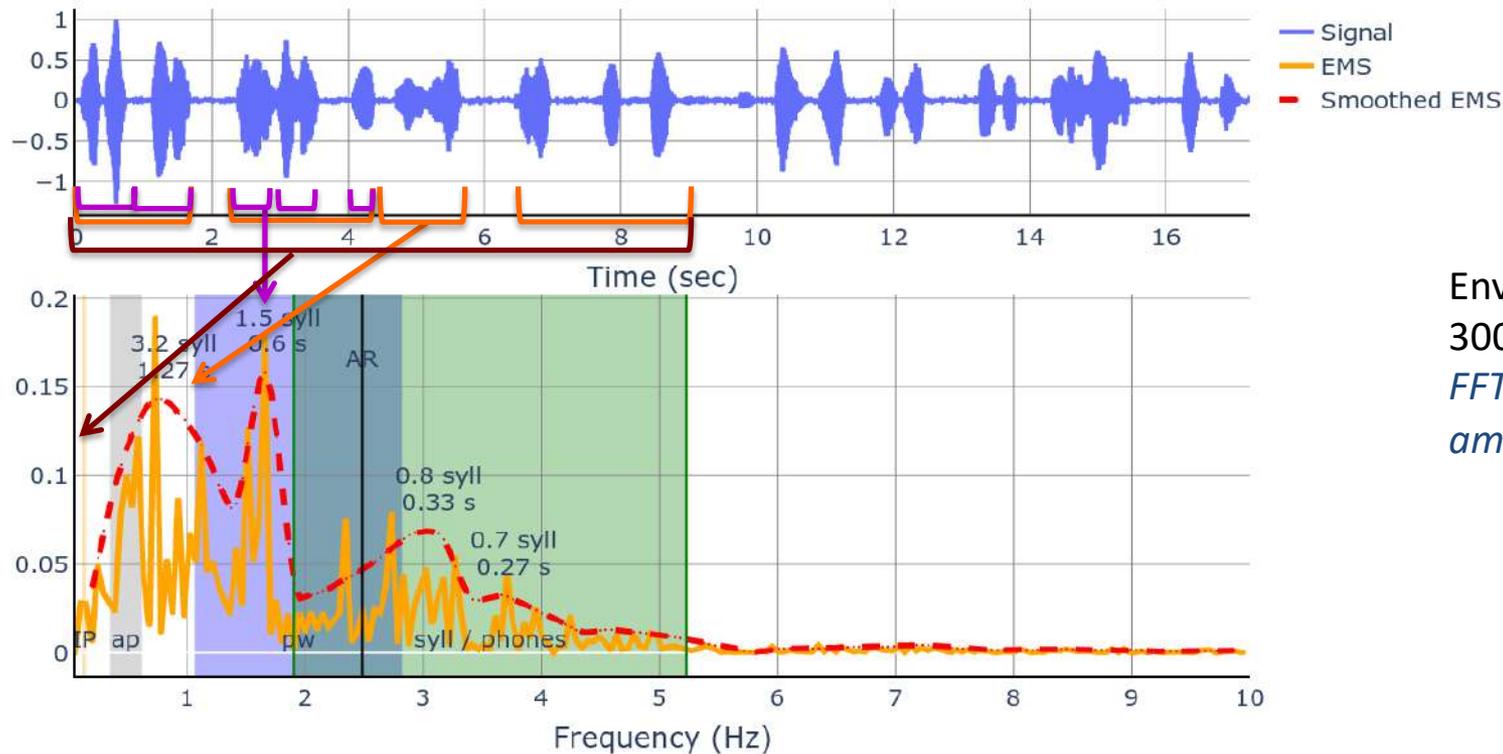


'Normal' speaking rate (4,3 syll/sec)

PW (1,66 Hz; 600 ms), AP (1,1 Hz; 900 ms) and IP (0,5 Hz; 2,5 s)



Exemple 2 - spectre de modulation d'amplitude (patient cancer)



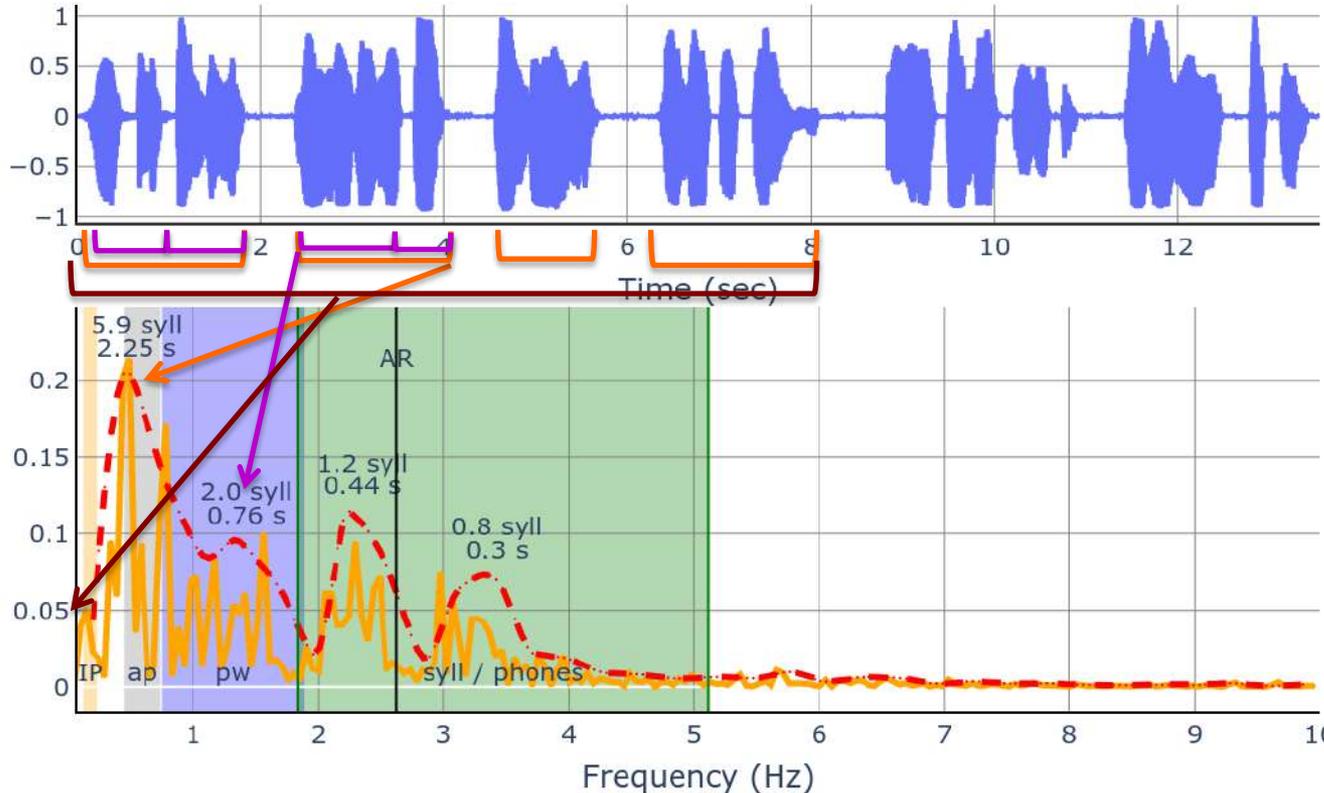
Envelope filtering :
300-1000 Hz
FFT envelope
amplitude



Haltingly speaker (word by word; 1,3 syll/sec)
PW (1,66 Hz; 600 ms), AP (0,77 Hz; 1,3 s.) and IP (0,1 Hz; 9 s.)



Exemple 3 - spectre de modulation d'amplitude (patient cancer, rythme fluide)



Envelope filtering :
300-1000 Hz
FFT envelope
amplitude



Fluent speaker with slow speaking rate (2 syll/sec)
PW (1,3 Hz; 760 ms), AP (0,44 Hz; 2,3 s.) and IP (0,13 Hz; 8 s.)



Objectifs de cette présentation

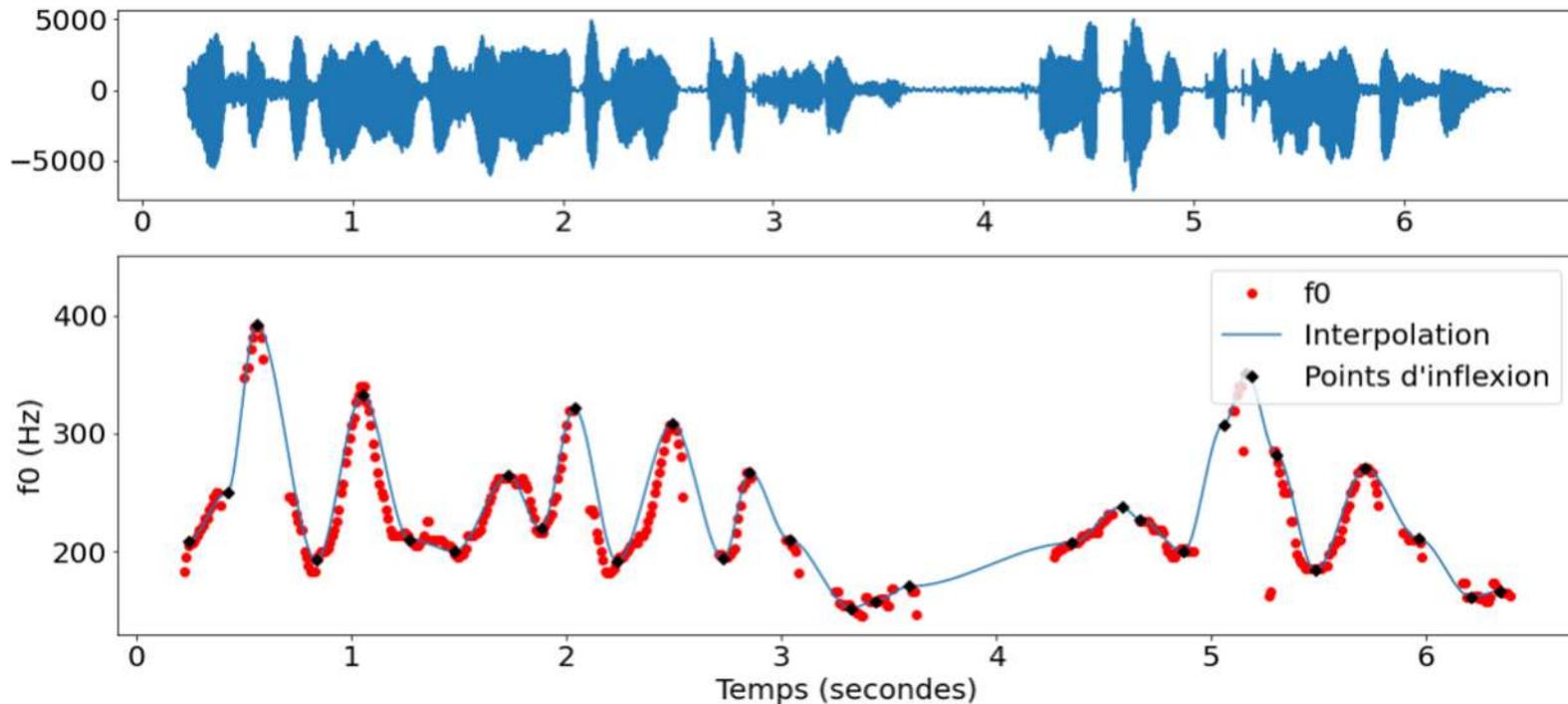
1. Détailler le cadre théorique prosodique utilisé
2. Présenter une représentation du rythme développée dans le cadre de la caractérisation prosodique de la parole pathologique
3. Analyse des perspectives :
 - a) Consolider la représentation en incluant une représentation intonative
 - b) Automatisation de la détection des niveaux prosodiques
 - c) Etudier l'évolution temporelle de la représentation
 - d) Envisager l'extraction des paramètres stables pour caractériser le rythme



Spectre de modulations de fréquence (1/2)

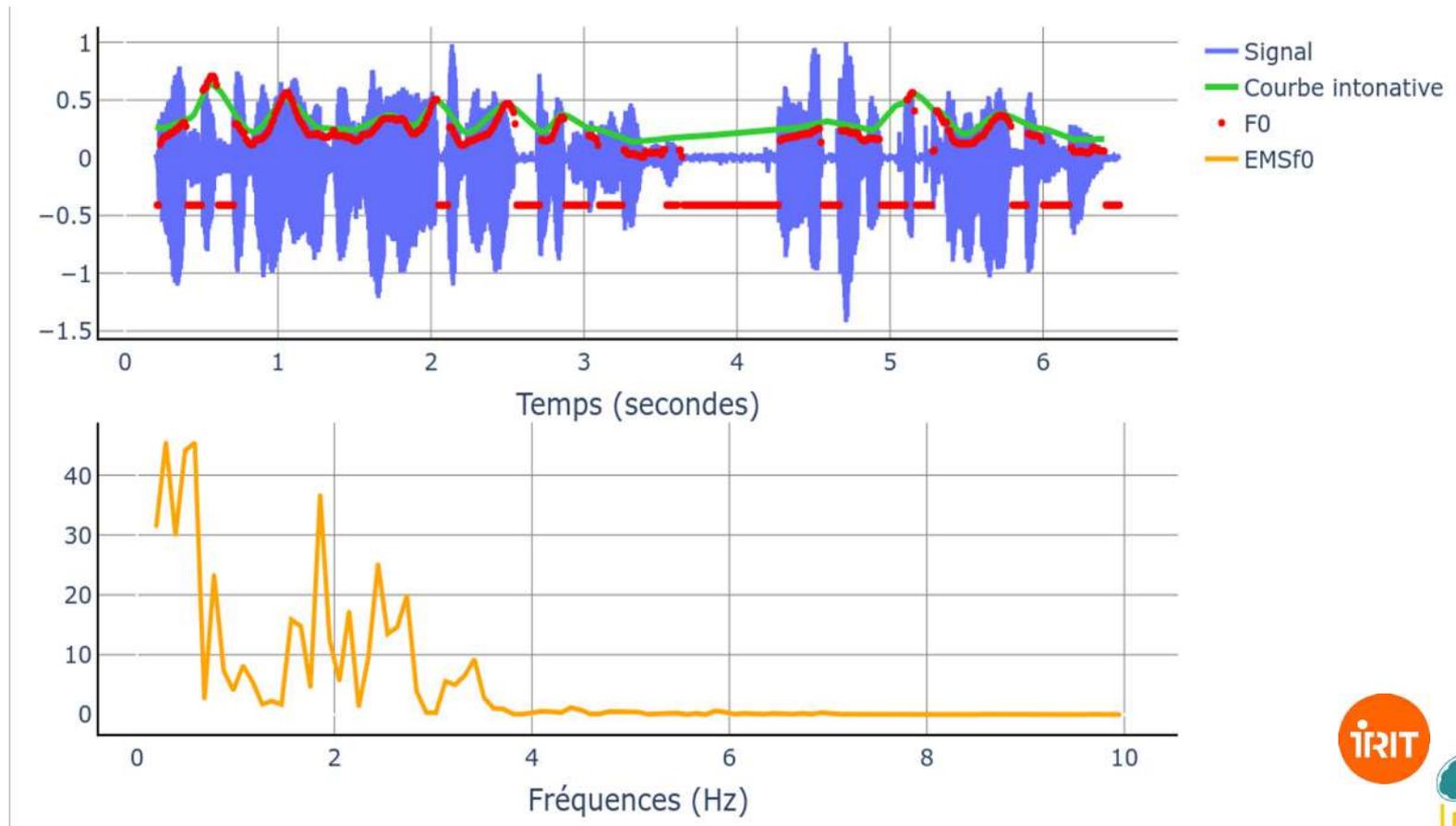
- Utilisation de la F0 pour remplacer l'enveloppe d'amplitude (Gibbon, 2021)

Utilisation de la **courbe intonative** via l'algorithme **Momel** (Hirst & Espesser, 1993)



Spectre de modulations de fréquence (2/2)

- Les pics obtenus ne représentent que les larges unités prosodiques
 - Absence d'informations sur les niveaux inférieurs à l'ip (ap, pw, syllabes, phonèmes)



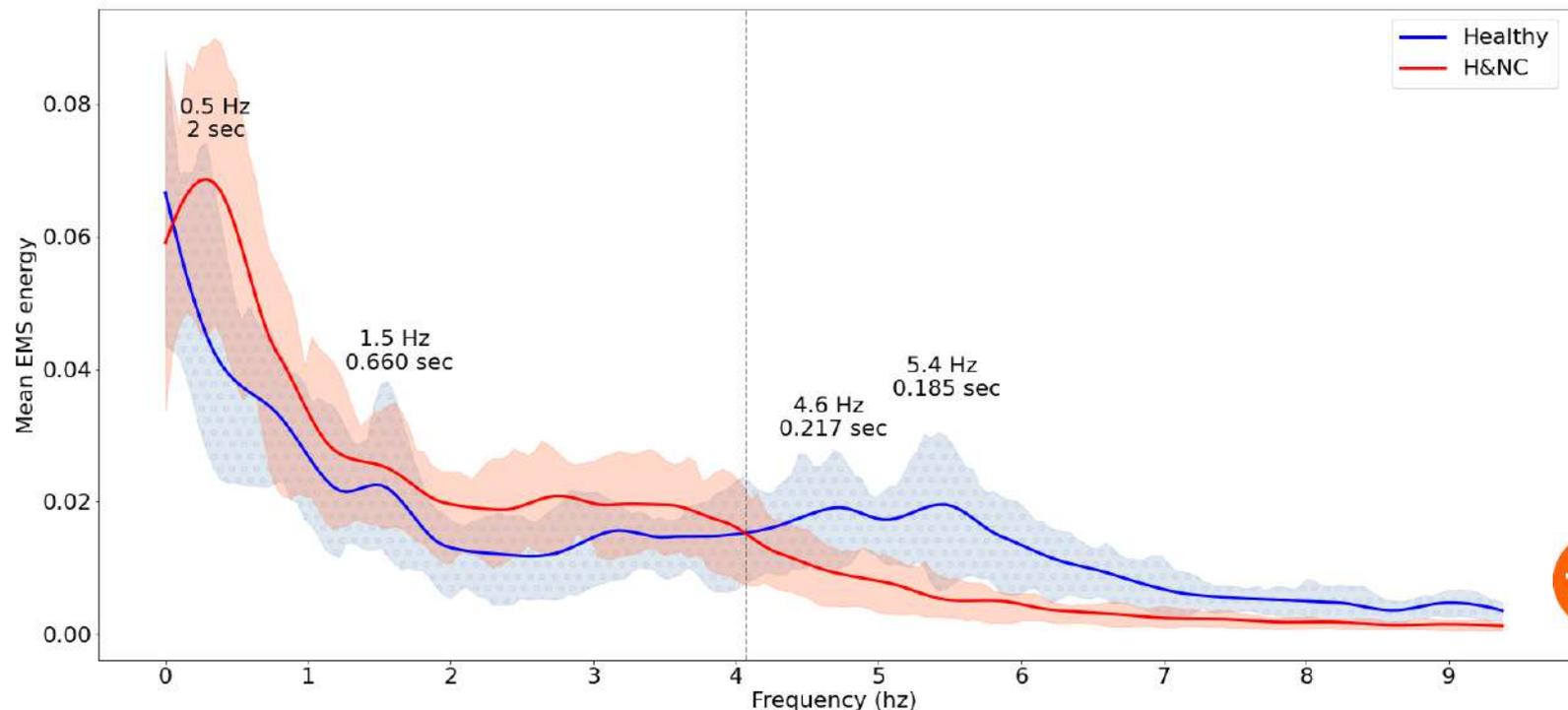
Objectifs de cette présentation

1. Détailler le cadre théorique prosodique utilisé
2. Présenter une représentation du rythme développée dans le cadre de la caractérisation prosodique de la parole pathologique
3. **Analyse des perspectives :**
 - a) Consolider la représentation en incluant une représentation intonative
 - b) **Automatisation de la détection des niveaux prosodiques**
 - c) Etudier l'évolution temporelle de la représentation
 - d) Envisager l'extraction des paramètres stables pour caractériser le rythme



Compensation rythmique pour les patients cancer

- **PW** : relativement stable autour de 650 ms
- **AP** descent en fréquence et remplace les IP “normaux”
- Perte d’information en dessous de la syllable : > 4 Hz = peu d’information sur l’articulation des phonèmes

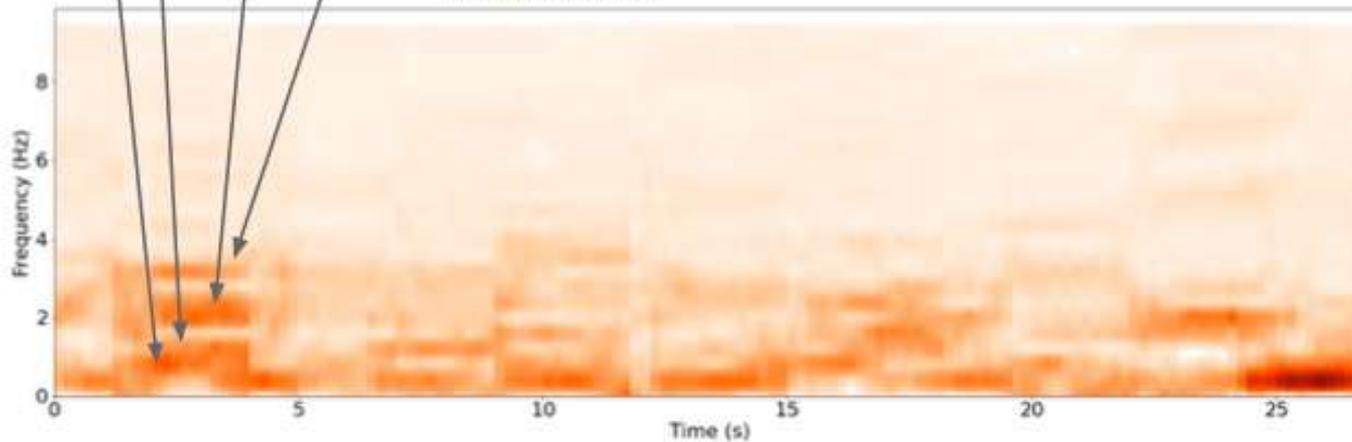
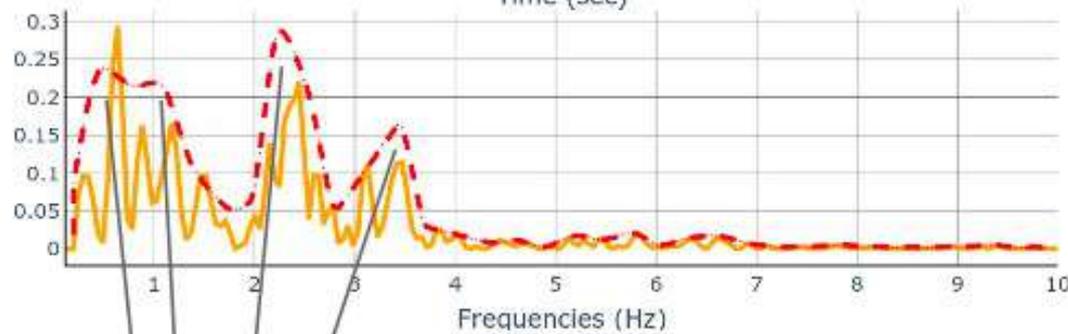
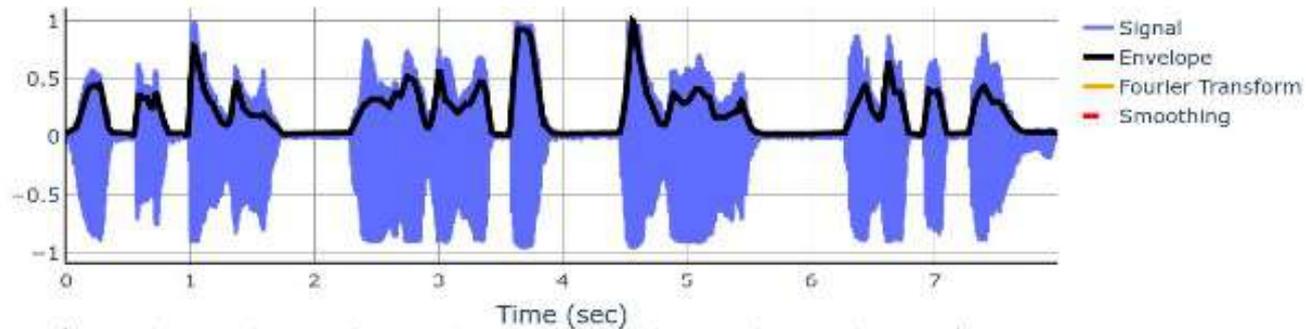


Objectifs de cette présentation

1. Détailler le cadre théorique prosodique utilisé
2. Présenter une représentation du rythme développée dans le cadre de la caractérisation prosodique de la parole pathologique
3. **Analyse des perspectives :**
 - a) Consolider la représentation en incluant une représentation intonative
 - b) Automatisation de la détection des niveaux prosodiques
 - c) **Etudier l'évolution temporelle de la représentation**
 - d) Envisager l'extraction des paramètres stables pour caractériser le rythme



Evolution temporelle de la représentation



Objectifs de cette présentation

1. Détailler le cadre théorique prosodique utilisé
2. Présenter une représentation du rythme développée dans le cadre de la caractérisation prosodique de la parole pathologique
3. **Analyse des perspectives :**
 - a) Consolider la représentation en incluant une représentation intonative
 - b) Automatisation de la détection des niveaux prosodiques
 - c) Etudier l'évolution temporelle de la représentation
 - d) **Envisager l'extraction des paramètres stables pour caractériser le rythme**



Caractérisation rythme

