



**HAL**  
open science

# Spectre moyen de la parole - Intérêt pour le bilan et la rééducation orthophonique

Etienne Sicard, Anne Menin-Sicard

► **To cite this version:**

Etienne Sicard, Anne Menin-Sicard. Spectre moyen de la parole - Intérêt pour le bilan et la rééducation orthophonique. LURCO Unadreo; INSA Toulouse. 2023. hal-04152457

**HAL Id: hal-04152457**

**<https://hal.science/hal-04152457v1>**

Submitted on 10 Jul 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Spectre moyen de la parole – Intérêt pour le bilan et la rééducation orthophonique

Etienne Sicard<sup>1,2</sup>, Anne Menin-Sicard<sup>1</sup>

(1) Laboratoire LURCO/UNADREO, équipe de recherche ERU 46

(2) INSA-GEI, 135 Avenue de Rangueil, 31077 Toulouse

Résumé : Le spectre moyen de la parole permet d'afficher la distribution de l'énergie en fonction de la fréquence, en donnant des informations intéressantes de possibles altérations, en un seul graphique et de manière automatique. Toutefois, son intérêt en orthophonie clinique n'est pas démontré, bien que différentes études scientifiques utilisent le spectre moyen pour analyser les caractéristiques de la parole pathologique. Nous avons étudié une quarantaine d'articles scientifiques sur le sujet, afin de dégager des métriques potentiellement intéressantes pour évaluer les progrès suite à la thérapie vocale. Le spectre moyen souffre d'un manque de références normatives et soulève des questions, notamment sur la diversité, le manque de maturité et de pertinence de certaines approches. Nous avons réalisé une implémentation du spectre moyen et retrouvé au travers de différents cas avant/après prise en soin les tendances observées dans la littérature scientifique. Le spectre moyen peut être considéré comme un outil complémentaire du spectrogramme apportant des informations objectives et illustratives sur l'évolution des performances du patient au cours de la thérapie. Il présente l'avantage de fournir une illustration d'une grande précision sur la phonation dans son ensemble, la voix et la parole étant souvent analysées de façon distincte.

*Mots-clés : orthophonie, voix, parole, spectre moyen, analyse objective, fondamental usuel moyen, fo, pathologies vocales, pathologies de la parole, rapport alpha, moments, voix d'hommes, voix de femmes, voix d'enfants, ERU 15, ERU 46, VOCALAB, DIADOLAB.*

### Introduction

Le spectre moyen à long terme (*Long Term Average Spectrum - LTAS*<sup>1</sup>) est utilisé depuis plus de 50 ans pour analyser les caractéristiques fréquentielles du signal de parole. Il permet d'afficher la distribution de l'énergie moyenne de la parole, sur des modalités allant de simples voyelles isolées jusqu'à des portions de discours spontané. Le spectre moyen donne notamment des informations sur la fréquence fondamentale (*fo*), la résonance du conduit vocal et le bruit de souffle (turbulences d'air en lien avec des rétrécissements), le tout en un seul graphique.

Dès les années 70, l'intérêt d'utiliser le spectre moyen pour évaluer les progrès grâce à la thérapie vocale a été mise en évidence, notamment par [Frokjaer-Jensen 1974] comme illustré figure 1, qui correspond au spectre de la parole moyennée sur 45 secondes d'une patiente présentant une paralysie récurrentielle unilatérale, avant et après prise en charge. On note des zones où l'énergie relative est significativement plus élevée après thérapie (5 dB dans la bande 1-3.5 KHz, 10 dB autour de 3 KHz), ce qui correspond à l'évolution attendue suite à la prise en charge orthophonique.

---

<sup>1</sup> D'autres dénominations existent dans la littérature scientifique telles que *Long Term Average Speech Spectrum* (LTASS)

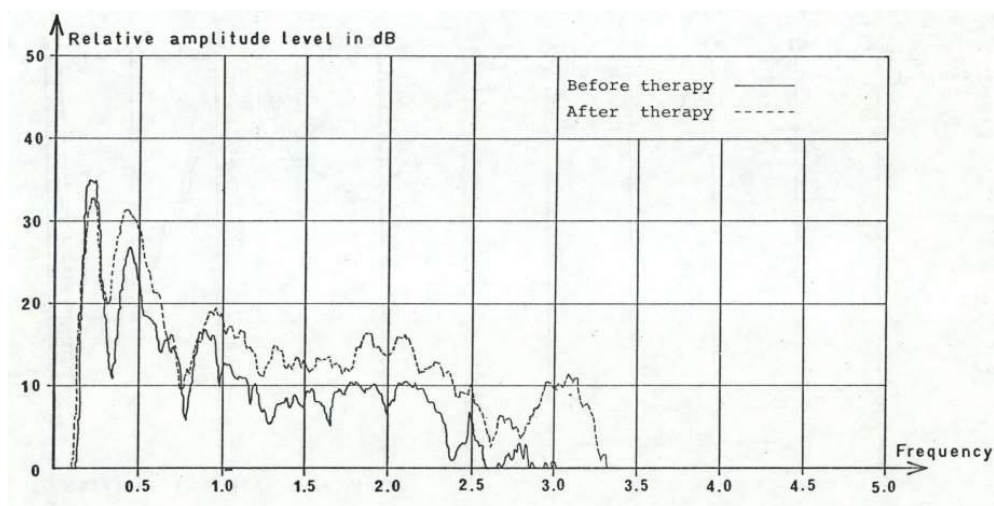


Figure 1 : Distribution spectrale du niveau d'amplitude de la parole moyennée sur 45 secondes d'une patiente présentant une paralysie récurrentielle unilatérale, avant et après prise en charge [Frokjær-Jensen 1974]

Autant Baken (2000) s'est contenté de lister dans son ouvrage « *Clinical Measurement of Speech* » quelques références traitant du spectre moyen, Kent (2000) donne plus de détails dans l'ouvrage « *Voice Quality Measurements* » en détaillant les formules des différents indices proposés, notamment en lien avec les rapports d'énergie. Une revue de littérature sur le spectre moyen long terme a été proposée par [Master 2006] avec pour objectif de recenser les applications possibles en lien avec l'analyse de la voix notamment en pratique orthophonie. Cette revue ne donne toutefois pas d'analyse comparative des indices acoustiques extraits des spectres moyens ni de valeurs de référence.

Le but de notre travail est d'étudier les caractéristiques des spectres moyens de la parole, de faire une synthèse des publications les plus intéressantes permettant de comparer les performances des patients et des témoins, afin d'extraire les indices les plus intéressants pour guider l'orthophoniste clinicien. Cette analyse est proposée au chapitre 2. Nous discutons au chapitre 3 du spectre moyen typique des hommes, femmes et enfants. Nous détaillons au chapitre 4 l'implémentation d'une sélection d'indices dans la plateforme DIADOLAB, et évaluons au travers de divers exemples de prise en charge la pertinence de ce type de métrique au chapitre 5, avant une discussion générale et une conclusion.

## Notations

Nous utilisons les notations issues d'un consensus international publié par [Titze 2015], avec notamment la description de la fréquence fondamentale sous le sigle «  $f_0$  » (fréquence d'oscillation<sup>2</sup>), tandis que  $F_1..F_n$  représentent les groupes d'harmoniques particulièrement énergétiques, appelés formants. Une illustration de ces différentes métriques est proposée à la figure 2, où l'on distingue  $f_0$ ,  $2.f_0$ , leur niveau correspondant en dB appelés  $L_1$  et  $L_2$ , ainsi que les formants  $F_1..F_3$ .

<sup>2</sup> Nous n'utilisons plus le symbole «  $F_0$  » qui induit une confusion avec la notion de formant ( $F_1..F_n$ )

HARMONIQUES	FREQUENCE (HZ)	AMPLITUDE (V)	NIVEAU (DB)
FONDAMENTAL, FREQUENCE D’OSCILLATION	$f_0$	A1	L1
2EME HARMONIQUE	$2.f_0$	A2	L2
N-IEME HARMONIQUE	$n.f_0$	$A_n$	$L_n$

FORMANTS	FREQUENCE (HZ)	NIVEAU (DB)	LARGEUR (HZ)
FORMANT 1	F1	LF1	BF1
FORMANT 2	F2	LF2	BF2
FORMANT N	$F_n$	$LF_n$	$BF_n$

Table 1 : symboles décrivant les harmoniques et les formants selon le consensus [Titze 2015]

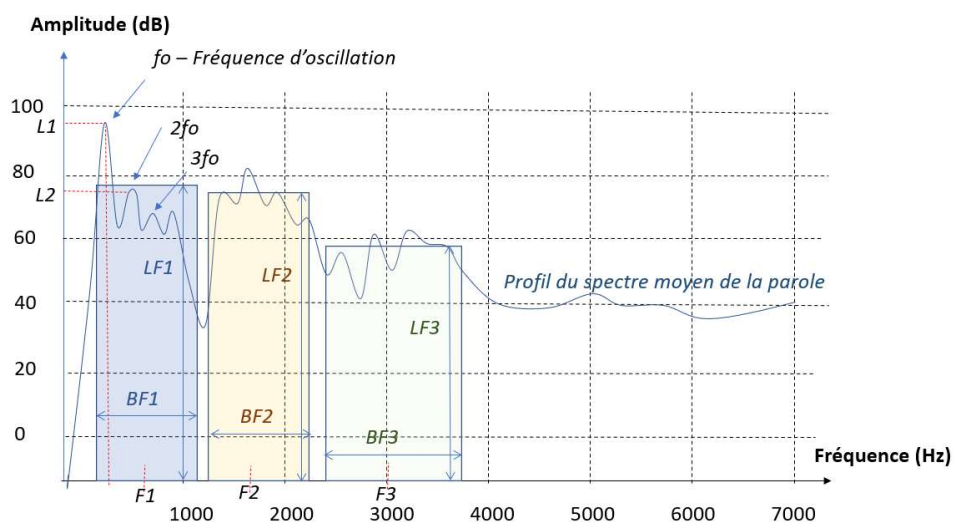


Figure 2 : Illustration des principales notations appliquées à une courbe de spectre moyen de la parole, selon le consensus [Titze 2015]

### Autres notations

Nous définissons aussi les variables suivantes, et les valeurs utilisées dans les logiciels de prise en charge orthophonique de la parole DIADOLAB [Sicard 2023b] et de la voix VOCALAB [Sicard 2023c] :

- $f_0$ , fréquence d’oscillation, moyennée sur l’ensemble de l’échantillon sonore. La valeur typique de  $f_0$  est autour de 120 Hz pour les hommes, 220 Hz pour les femmes et 250 Hz pour les enfants [Sicard 2023].
- $F_e$ , fréquence d’échantillonnage : 22050 Hz
- $f_{max}=F_e/2$  soit 11025 Hz
- $f_{min}=85$  Hz
- $FFT\_N$  : nombre de points de calculs de la FFT : 1024 points du signal sonore

La limite haute de fréquence  $f_{max}$  considérée comme pertinente pour le calcul du spectre moyen a été définie autour de 5 KHz dans les publications les plus anciennes [Frokjaer-Jensen 1975], souvent en lien avec des limitations matérielles, puis s’est élevée à 10 KHz à partir des années 1990 [Hammarberg 1986] [Mendoza 1996] [Stajner-Katusic 2012].

La limite basse  $f_{min}$  varie selon les auteurs de 0 Hz (publications les plus anciennes, mais qui était en réalité la fréquence de coupure basse du microphone, probablement autour de 100 Hz) à 250 Hz, avec au fil du temps le positionnement de cette valeur autour de la fréquence fondamentale la plus basse, c'est-à-dire 70-80 Hz comme dans [Stajner-Katusic 2012].

### Protocole

Même si l'ensemble des approches scientifiques s'accordent sur le principe d'enregistrer la parole du patient, aucun consensus n'existe vraiment sur le type d'épreuve à faire passer. On note toutefois chez beaucoup d'auteurs la lecture d'un texte, de l'ordre d'une minute, ou bien des tâches simples telles que du comptage, ou une description d'image.

Le protocole de bilan de la phonation décrit dans [Menin-Sicard 2016] inclut deux épreuves sur lesquelles le spectre moyen peut être aisément calculé : une phrase de présentation de type contrainte et semi-contrainte « Bonjour, je m'appelle..., j'ai ....ans et je viens chez l'orthophoniste parce que..... », la consigne de voix déclamée « Maître corbeau sur un arbre perché. » de La Fontaine. Le bilan de la parole décrit dans [Menin-Sicard 2019] inclut aussi la phrase de présentation.

### Pré-traitement de la voix

Dans la mesure où l'on cherche à analyser la parole, les silences et pauses sont en général exclus du calcul du spectre moyen, surtout si l'ambiance sonore au moment de l'enregistrement n'est pas neutre, et peut se traduire par des composantes spectrales parasites. En général, le calcul exclut aussi les portions de son non voisé, c'est-à-dire les consonnes fricatives ou sourdes, afin d'évaluer la qualité de la source sonore et de la résonance de la voix. En excluant du calcul l'énergie des fricatives /f,s,ch/ dont la signature spectrale est principalement à haute fréquence, on peut considérer que les composantes au-delà de 5 KHz sont dues, soit à la constriction des fricatives voisées /v,z,j/ généralement plus faible que les équivalents sourds, soit aux turbulences liées au souffle dans la voix ( serrage, fuite glottique, fuite nasale etc...). Il co-existerait donc deux sources de bruit potentielles, l'une naturelle (consonnes fricatives voisées), l'autre parasite (turbulences anormales), qui seront difficilement séparables dans le résultat final.

## Métriques objectives d'analyse du spectre moyen

De nombreuses publications ont décrit l'utilisation du spectre moyen en vue de l'évaluation de la qualité de la voix, notamment dans le contexte de la prise en soin orthophonique. Il est apparu très tôt la nécessité de définir des métriques objectives permettant de comparer les spectres, notamment au cours de la prise en charge. Certaines métriques sont considérées comme discriminantes pour séparer les groupes témoins des patients, ou pour mesurer les progrès de la thérapie. Nous sélectionnons ci-après les métriques les plus saillantes, ayant un pouvoir de séparer les classes entre elles (typiquement témoin-patient, ou avant/après prise en charge).

Les métriques sélectionnées sont listées avec leur notation et valeur de référence approximatives, extraites des figures ou des tableaux de données publiées. Lorsque les résultats portent sur différentes paramétrages comme dans [Dromey 2003], nous sélectionnons les données les plus proches des valeurs de référence précisées au chapitre précédent.

Il n'existe pas à proprement parler de métrique ou d'indice à caractère normatif, un point déjà relevé dans la revue de [Master 2006]. Concernant les paramètres tels que les plages de fréquence considérées ou la fréquence d'échantillonnage, le type de microphone ou de protocole, on trouve des variations importantes entre les études, ce qui gêne la comparaison des résultats.

Nous traitons ci-après deux classes de paramètres les plus fréquents dans la littérature scientifique : d'une part les rapports d'énergie, d'autre part les moments statistiques.

### Rapports d'énergie

Le rapport d'énergie le plus ancien, appelé rapport  $\alpha$ , a été formalisé par [Frokjaer-Jensen 1975] et correspond à un rapport entre l'intensité du signal au-dessus de 1 KHz et l'intensité en dessous de 1 KHz. Ce rapport est défini par l'équation 1.

$$\alpha = \frac{\text{Intensité}>1K}{\text{Intensité}<1KHz} \quad (\text{Eq. 1})$$

En convertissant l'intensité en dB, le rapport  $\alpha$  devient une différence entre le niveau moyen du spectre  $> 1$  KHz et le niveau moyen  $<1$  KHz. En général, pour la parole normale, il y a moins d'énergie au-dessus de 1 KHz qu'au-dessous, d'où une valeur du rapport  $\alpha$  négative en dB (autour de -10 dB), ce qui correspond en linéaire à une valeur de 0.3 selon l'équation 1. La fréquence maximale de 5 KHz fixée par [Frokjaer-Jensen 1975] était probablement initialement liée à la fréquence maximale du système d'analyse. Il existe de nombreuses publications qui exploitent le rapport alpha, y compris des publications assez récentes. Nous donnons certains résultats publiés à la table 2. Concernant le rapport alpha, certaines données sont en linéaire, d'autres en dB.

Le rapport alpha est corrélé à la pente spectrale : plus la pente spectrale est raide, plus l'intensité des harmoniques hautes est faible par rapport aux harmoniques basses, ce qui correspond en général à une voix pauvre en harmoniques. Une fermeture glottique incomplète, un rétrécissement lié à un serrage sus glottique produit un flux d'air turbulent et irrégulier, réduisant l'énergie des harmoniques dans la bande 1-5 KHz [Frokjaer -Jensen 1974] ce qui se traduit par un rapport alpha quelques dB en dessous de celui des voix normales.

REFERENCE	N	TYPE DE PAROLE/PATHOLOGIE	PARAMETRE	VALEUR DE REFERENCE
[FROKJAER-JENSEN 1976]	10 HC, 50 patients	Paralysie récurrentielle unilatérale (PRU)	$\alpha$	HC : -10 dB PRU : -15 dB
[KITZING 1986]	10 HC (8F, 2H)	Impliquée, douce, soufflée, éraillée	$\alpha$	I : 0.29 D : I+0.002 S : I+0.007 E : I-0.007
[LOFQVIST 1986]	10 HC (F)	Matin et soir	$\alpha$	Matin : 0.55 Soir : 0.45
[HARTL 2003]		Paralysie récurrentielle unilatérale (PRU)		
[LEINO 2008]	14 Good (H) 22 Interm. (H) 14 Poor (H)	Lecture de texte, étudiants	$\alpha$	G : -11.3 I : -10.5 P : -13.3
[DA SILVA 2011]	30 jeunes (F) 30 âgées (F)	Comparaison femmes 27 ans et 70 ans sur un texte de 250 mots	$\alpha$	-14.3 dB -17.7 dB
[ANTONETTI 2020]	53 HC (25F, 28H) 49 Dys. (29F, 20H)	Dysphonie	$\alpha$	HC : -23.7 dB Dys : -24.1 dB
[SOUMYA 2022]	30 HC (H) 30 Dys (H)	Age 70-75 ans	L1-L2	HC : 11.3 dB Dys : 17.6 dB

Table 2 : Sélection de publications sur le spectre long terme utilisant les rapports d'énergie alpha et L1-L2

### Discussion sur la métrique $\alpha$

Dans l’étude de [Kitzing 1986], un même passage de discours continu a été produit par 10 thérapeutes dans 4 modalités de voix différentes : voix normale, voix soufflée, voix serrée et voix douce. Des différences relativement faibles ont été observées concernant le rapport alpha, malgré une grande diversité de production. Parmi les résultats les plus significatifs, l’intensité LF1 du format F1 était 5 dB plus forte en voix serrée, et 3 dB plus faible en voix douce.

L’analyse de 10 voix de cliniciennes par [Lofqvist 1986] le matin avant le travail puis en fin d’après-midi après usage intensif de la voix a montré une nette différence entre les deux enregistrements, et une réduction globale du paramètre alpha (0.5 à 0.4), ce qui correspond à une perte d’énergie dans les aigus, que l’on peut associer à la fatigue vocale. Les auteurs soulignent que les spectres moyens sont d’une utilité limitée pour le dépistage, du fait de leur trop grande variabilité, mais peuvent être utiles pour évaluer les progrès en lien avec une prise en charge thérapeutique.

En revanche, les travaux récents de [Antonetti 2020] n’ont pas montré de différence significative concernant le rapport alpha entre une cinquantaine de témoins et de patients dysphoniques, ce qui incite à considérer cette métrique avec prudence. Des différences aussi faibles que 3 à 6 % ont été observées par [Kitzing, 1986] alors que les différences perceptives de qualité vocale entre les échantillons étaient très marquées.

### A propos de la limite fréquentielle

La justification de la valeur de 1 KHz retenue par [Frokjaer-Jensen 1975] fait suite à l’examen des graphes cumulés des spectres moyens de 50 patients, significativement moins riches en énergie que les sujets normaux au-delà de cette limite. On peut rapprocher la valeur de 1 KHz de la limite entre les formants F1 et F2 pour les hommes et pour les femmes (Table 3). Dans cette hypothèse, pour les enfants, cette valeur serait trop basse. Le calcul de la valeur alpha que nous avons implémenté dans DIADOLAB fait intervenir non pas une valeur fixe, mais la limite F1/F2 afin d’équilibrer les rapports d’énergie en fonction du genre.

De manière similaire, la limite rigide de 5 KHz, probablement liée aux limitations matérielles de l’expérimentation historique de [Frokjaer-Jensen 1976], mériterait d’être rendue dépendante du genre par souci de cohérence. Une proposition serait de suivre la limite des formants F4/F5, comme reporté à la table 3, qui peut aussi être considérée comme la limite haute des harmoniques principales de la parole, incluant le « formant du chanteur », donc de la parole « utile ».

<i>Genre</i>	<i>Formant</i>	<i>Fréquence basse</i>	<i>Fréquence haute</i>	<i>Limite alpha<sup>3</sup></i>	<i>F_limite<sup>4</sup></i>
<i>Homme</i>	F1	245	980	980	4000
	F2	980	1960		
<i>Femme</i>	F1	275	1100	1100	4500
	F2	1100	2200		
<i>Enfant (6 ans)</i>	F1	345	1380	1380	5500
	F2	1380	2760		

Table 3 : Plage des formants F1 et F2 et limite alpha utilisée dans DIADOLAB [Sicard 2023] selon le genre

<sup>3</sup> Limite implémentée dans DIADOLAB, qui évolue selon le genre du patient

<sup>4</sup> Limite implémentée dans DIADOLAB, qui évolue selon le genre du patient et peut être manuellement ajustée



L'étude de [Watts 2011] considère d'ailleurs le rapport entre l'énergie en dessous de 4 KHz et celle au-dessus. Au-delà de cette limite  $f_{limite}$ , se retrouvent principalement l'énergie des fricatives et le bruit de souffle antériorisé (ex : /j/ /z/ et /v/). Il apparaît déjà prudent, dans l'éventualité de la mise au point d'un outil d'analyse du spectre moyen, de permettre au clinicien d'adapter cette frontière en fonction des caractéristiques propres au patient, plutôt que de se cantonner à « 5 KHz ».

#### Variante de la métrique alfa

Les différences entre l'amplitude L1 de la première harmonique  $f_0$  et l'amplitude L2 de la seconde harmonique de fréquence  $2.f_0$  sont une alternative à la métrique alpha, et se concentrent sur la partie très basse fréquence du spectre moyen. En général, L1 est plus énergétique que L2, donc L1-L2 est positif, de l'ordre de 10 dB. Une forte corrélation a été observée entre L1-L2 et le degré de sévérité de la dysphonie évaluée avec le GRBAS [Hirano 1981] par [Soumya 2022]. La valeur de L1-L2 est de l'ordre de 11 dB pour les témoins et tend à augmenter avec la sévérité de la dysphonie. La même métrique L1-L2 s'avère pertinente selon [Narasimhan 2017] pour séparer un groupe de 40 témoins sans atteinte ORL avec un autre groupe de 40 patients souffrant d'une voix enrouée suite à des nodules, polypes ou œdèmes des cordes vocales.

D'autres rapports d'énergie ont été publiés, mais sont très peu réemployés par d'autres chercheurs. L'étude de [Hartl 2003] compare l'énergie de quatre bandes de fréquences B0 (0 à 400 Hz), B1 (400–2 000 Hz), B2 (2–5 KHz) et B3 (5–10 KHz) et constate une différence nettement moindre entre la bande basse fréquence B0 et les bandes supérieures B1, B2, B3 dans le cas de paralysie unilatérale des cordes vocales. Une autre variante étudiée par [Antonetti 2020] consiste à s'intéresser au rapport entre LF1 (niveau en dB du formant 1) et L1 (niveau en dB de  $f_0$ ), comme l'avait étudié, entre autres [Kitzing 1986]. On peut constater qu'il existe une grande diversité d'approches concernant les rapports d'énergie, un peu à l'image des indicateurs d'altération de la phonation décrits dans [Sicard 2013].

#### Moments

Les moments sont des métriques relativement simples qui caractérisent la forme d'une distribution aléatoire de points. On suppose une distribution de valeurs du spectre moyen sous forme de probabilité  $p(k)$ , dont la somme vaut 1. Le rang  $k$  est relié à la fréquence de manière linéaire, avec un pas fréquentiel noté  $fk$  (eq. 7).

Le 1<sup>er</sup> moment du LTAS, codé M1, est le centre de gravité de la distribution du spectre (Centre of Gravity – COG), exprimé ici en Hz. La formulation de M1 est donnée par l'équation 2, selon [Forrest 1988]. En appliquant cette formule, la valeur de M1 se situe en général entre 2 et 5 KHz.

Le 2<sup>ème</sup> moment M2 est corrélé à la variance et s'exprime en  $\text{Hz}^2$ . Sa racine carrée est l'écart-type spectral (noté  $s$  ou  $stdev$  ou  $m2$ ) par rapport au centre de gravité et s'exprime en Hz. M2 est corrélé à l'étalement de la distribution fréquentielle du LTAS. L'équation correspondante est donnée par l'équation 3, et donne des valeurs de M2 entre 1 et 3 KHz. M2 varie considérablement selon les publications reportées Table 3. Une illustration des moments M1 et M2 est proposée figure 3.



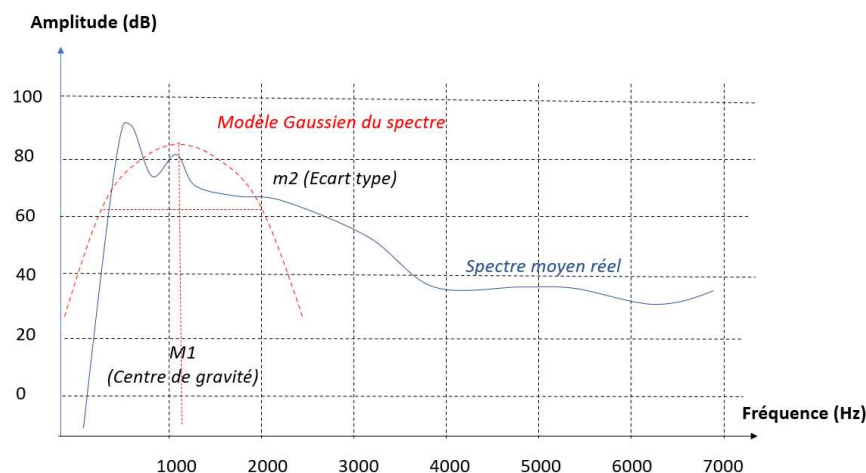


Figure 3 : illustration de l’approximation d’un spectre moyen par une gaussienne de centre de gravité M1 et écart type M2

Le troisième moment spectral M3 est corrélé à l’asymétrie spectrale et s’exprime en Hz<sup>3</sup>. La fonction EXCEL™ correspondante est COEFFICIENT.ASYMETRIE (SKEWNESS en version Anglaise) qui implémente une version modifiée de l’équation 3, sur le principe de l’équation 5 [Doane 2011]. Elle permet d’obtenir un coefficient sans dimension d’une valeur relativement peu élevée. A noter que les formulations de l’asymétrie varient d’un logiciel à l’autre, comme noté par [Doane 2011], ce qui incite à la plus grande prudence dans les comparaisons. Un certain consensus existe toutefois pour considérer qu’une asymétrie positive signifie que la partie haute fréquence de distribution du LTAS est plus étalée que la partie basse fréquence, ce qui correspond au cas du LTAS normal.

Le quatrième moment spectral M4 permet d’évaluer l’aplatissement spectral. Il s’exprime en Hz<sup>4</sup>. En pratique, on utilise une formulation proche de l’équation 6, appelée coefficient de Kurtosis [Forrest 1988], sans dimension et de valeur raisonnable. Un coefficient de kurtosis positif indique une distribution proche d’un pic, tandis qu’un kurtosis négatif indique une distribution relativement plate. La fonction EXCEL™ correspondante est KURTOSIS, qui implémente une version approchant l’équation 6.

$$M1 = \sum_k f_k \cdot p(k) \quad (\text{Eq. 2})$$

$$Mn = \sum_k (f_k - M1)^n \cdot p(k) \quad (\text{Eq. 3})$$

$$s = m_2 = \sqrt{M2} \quad (\text{Eq. 4})$$

$$m_3 = \frac{M3}{M2^{3/2}} \quad (\text{Eq. 5})$$

$$m_4 = \frac{M4}{M2^2} - 3 \quad (\text{Eq. 6})$$

$$f_k = F_e k / N \quad (\text{Eq. 7})$$

Avec N=nombre de points de la FFT, Fe=fréquence d’échantillonnage (Hz). Les moments ont une dimension Hz<sup>n</sup>, s=écart-type (Hz), les coefficients m3 et m4 sont sans dimension. f<sub>k</sub> est le pas fréquentiel du spectre calculé par la FFT (Hz)<sup>5</sup>.

<sup>5</sup> Dans VOCALAB et DIADOLAB, N=1024, Fe=22050 Hz, f<sub>k</sub>=21,45 Hz.

REFERENCE	N	TYPE DE PAROLE/PATHOLOGIE	PARAMETRE (SIGNIFICATIF)	NOTATION	VALEUR DE REFERENCE (M1 EN HZ)
[DROMEY 2003]	10 HC 10 MP	Monologue - Maladie de Parkinson (MP)	Moments (***)	M1 (HC)	281
				M1 (MP)	204
				M2 (HC)	482
				M2 (MP)	312
				M3 (HC)	12
				M3 (MP)	22
				M4 (HC)	267
				M4 (MP)	838
[TANNER 2005]	96 Dys.	Dysphonie	Moments (***)	M2 (Pre)	1570
				M2 (Post)	1070
				M3 (Pre)	4
				M3 (Post)	5
[HOUTZ 2010]	10H 49F 19H 22F	Dysphonie par tension musculaire - MTD	Moments (**)	M1 (MTD)	2230 (H), 2896 (F)
				M1 (ADSD)	2390 (H), 2220 (F)
		Dysphonie Spasmodique (ADSD)		M2 (MTD)	2572 (H), 2669 (F)
				M2 (ADSD)	2575 (H), 2324 (F)
[TJADEN 2010]	15 HC 10 PD 14 MS	Lecture de texte – HC vs Maladie de Parkinson vs Sclérose multiple	Moments (n.s)	M1 (HC)	3100
				M1 (PD)	2700
				M1 (MS)	3500
				M2 (HC)	2300
				M2 (PD)	2400
				M2 (MS)	2250
[SVIHLIK 2022]	23 HC (19H, 4F) 23 MP (19H-4F)	Maladie de Parkinson (MP)	Moments (***)	M1 (HC)	330
				M1 (MP)	280
				M2 (HC)	480
				M2 (PD)	330
				M3 (HC)	13
				M3 (PD)	17
				M4 (HC)	260
				M4 (MP)	500

Table 4 : Paramètres discriminants d'analyse du spectre moyen long terme. MP=Maladie de Parkinson, PRU= Paralyse récurrentielle unilatérale. HC=Healthy Control. MS=Sclérose Multiple, ADSD=Dysphonie Spasmodique, MTD=Diphonie par tension musculaire. N.s=Non significatif.

Concernant  $f_{min}$ , on se positionne en général autour de 80 Hz, pour éviter notamment le bruit à basse fréquence (50-60 Hz) tout en incluant le  $f_0$  des voix d'homme les plus basses. Concernant  $f_{max}$ , celle-ci est liée à la bande passante des microphones et la fréquence d'échantillonnage. Une valeur commune est 11025 Hz, soit la moitié de la fréquence d'échantillonnage  $F_e=22050$  Hz. La cohabitation de différents valeurs de bornes fréquentielles dans les publications altère la pertinence des comparaisons, la colonne « Valeur de référence » de la table 4 est donc à considérer avec prudence. En outre, certaines publications précisent des unités qui parfois souffrent de quelques incohérences. Dans nos propres évaluations, les coefficients associés aux moments M3 et M4 sont sans dimension (Eq. 5 et 6).

### Discordance des valeurs des moments

Si l'on analyse les valeurs des moments de la table 4, on constate des discordances impressionnantes au niveau des valeurs, notamment de M1 et M2. Certaines publications situent M1 autour de 200-400 Hz [Dromey 2003] [Svihlik 2022], d'autres comme [Tjaden 2010] ou [Tanner 2005] à des fréquences dix fois plus élevées. Différentes hypothèses peuvent être émises pour expliquer ces discordances : données initiales en linéaire ou en dB plage de fréquence en log ou en linéaire, ou encore calcul de M1 et M2 non pas sur la totalité de la distribution statistique mais dans une zone restreinte.

Ces discordances intriguent et compromettent toute comparaison directe. Plus ambigu encore est l'idée qu'un moment d'ordre 1 soit considéré comme meilleur s'il est plus élevé, lorsqu'il inclut des données au-delà de 5 KHz que de nombreux auteurs considèrent comme du bruit. Il nous semble donc scientifiquement indéfendable de faire porter le calcul des moments au-delà de  $f_{limite}$  si l'on cherche une métrique liée à la « qualité » de la voix. Une approche acceptable concernant la plage de calcul de M1 et M2 consiste à s'intéresser à la distribution dans la zone F1-F2, soit 500-3000 Hz comme illustré dans la représentation du triangle vocalique hommes et femmes de la figure 4, extraite de [Sicard 2020].

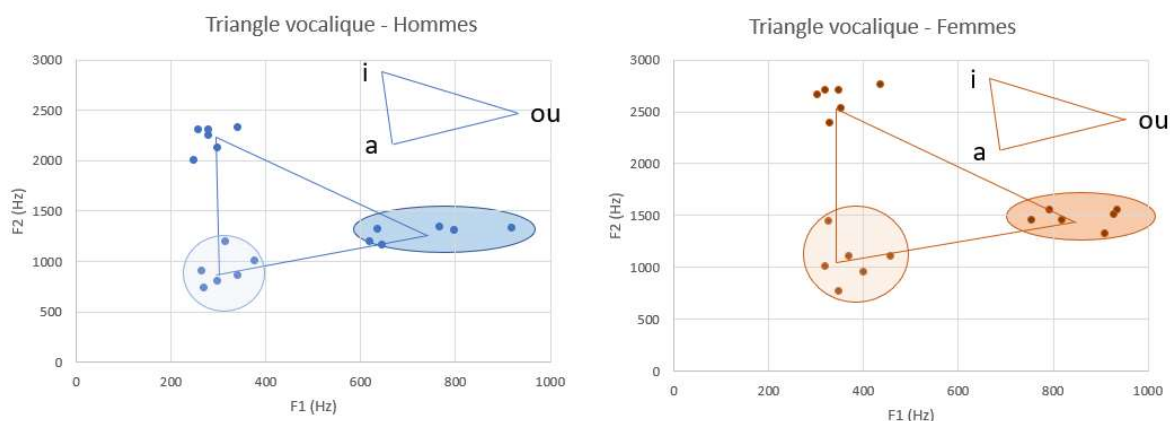


Figure 4 : Valeurs de F1 et F2 pour /a,i,ou/ selon diverses publications – Hommes à gauche, Femmes à droite gauche [Sicard 2020]

Ceci nous permet de se focaliser sur les zones de reconnaissance des voyelles, en espérant corrélérer les moments avec l'intelligibilité. En revanche, tout le contenu haute fréquence, qui est déterminant pour la discrimination des fricatives sourdes /f,s,ch/ et sonores /v,z,j/ est exclu du calcul.

### Proéminence harmonique

La proéminence harmonique est une indication de richesse harmonique de la parole. Appliquée sur une voyelle /a:/ tenue, elle rend compte du nombre d'harmoniques présente dans le spectre, avec la particularité de discriminer l'énergie des composantes harmoniques de celle du bruit. L'indicateur le plus connu est le *Cespral Peak Prominence* (CPP) qui s'élève avec la richesse harmonique [Murton 2017].

L'implémentation dans VOCALAB et DIADOLAB de l'indicateur de pauvreté harmonique fonctionne sur le même principe, mais dans un sens inverse : il s'élève avec la pauvreté harmonique, c'est-à-dire quand le niveau du bruit s'élève suffisamment pour s'approcher de celui des harmoniques. L'indicateur n'est étalonné que sur le /a:/ tenu, la norme étant une valeur inférieure à 1 [Sicard 2013].

## Qualité du spectre moyen - Théorie du papillon

Nous devons tout d'abord définir ce que l'on considère comme un spectre moyen de qualité, et comment un spectre moyen, altéré, et pathologique peut être distingué. Deux notions essentielles sont associées au spectre moyen : l'une positive qui est la richesse harmonique, qui se traduit par la présence d'énergie significatives dans la zone des formants F1, F2 et F3, l'autre négative qui est liée au bruit se traduisant par l'augmentation des composantes hautes fréquences (F4, F5, etc.). Il existe donc une fréquence  $f_{limite}$  (Table 3) au-dessous de laquelle une diminution de l'énergie sera considérée négativement (voix pauvre, peu résonante, sourde, hypophonique...) et au-dessus de laquelle une augmentation de l'énergie sera considérée comme un marqueur de mauvaise qualité de voix (souffle, constriction, serrage). La superposition du spectre moyen normal et pathologique prend alors l'aspect d'un papillon (Figure 5).

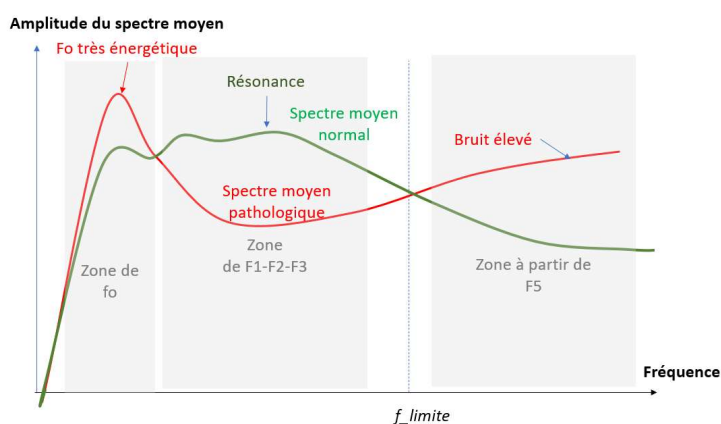


Figure 5 : Qualité du spectre moyen : fo peu marqué, résonance importante et faible bruit à haute fréquence. La superposition entre un spectre normal et pathologique prend la forme d'un papillon

Cette idée d'une fréquence limite transparait notamment dans [Lofqvist 1986] ainsi que dans [Tanner 2005] qui considèrent qu'un haut niveau d'énergie au-delà de 5 KHz peut être associée à une voix soufflée et hypo-fonctionnelle. De même, [Stajner-Katusic 2012] constate un spectre moyen plus faible à partir de 4 KHz, en postopératoire par rapport à l'état préopératoire de patients atteints de polypes, du fait de l'amélioration de la qualité de la voix. Nous retrouvons aussi la valeur de 4 KHz dans [Watts 2011] pour séparer les composantes basses et hautes fréquences et étudier leur rapport d'énergie. L'étude sur la voix d'enfants de [Sergeant 2008] met en évidence quant à elle une fréquence pivot autour de 6 KHz. La synthèse de ces indices porte à définir la fréquence  $f_{limite}$  dans la zone F4, comme illustré figure 5.

Un exemple spectaculaire de « rééducation papillon » est reporté figure 6. Les courbes correspondent au spectre moyen calculé sur la fable « maître corbeau », avant et après prise en soin, pour le cas AS033 de la base de données [ERU15], aussi décrit dans la compilation d'étude de cas [Menin-Sicard 2021b].

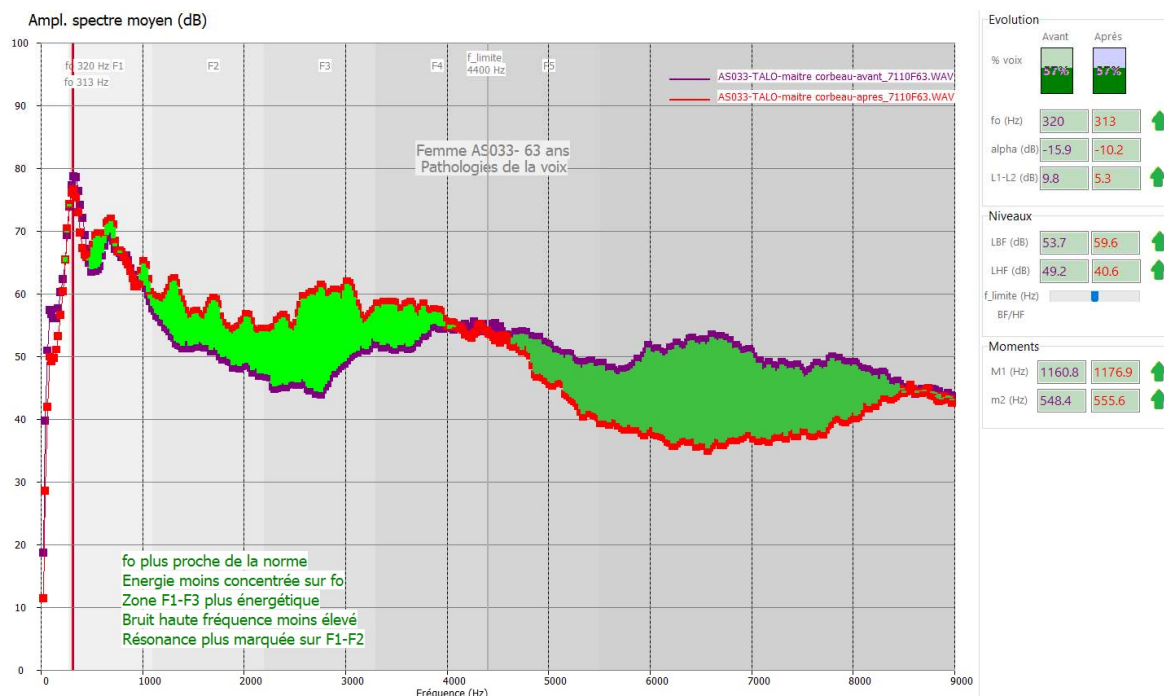


Figure 6 : Exemple de « rééducation papillon » cas de dysphonie AS033 [ERU15][Menin-Sicard 2021b]

Avant prise en charge, la patiente éprouve de grandes difficultés à parler, avec un souffle considérable dans la voix, très peu de résonance et une énergie concentrée sur fo. Le spectre suit la courbe en violet, qui comporte un bord d'aile gauche plutôt bas (peu d'harmoniques) et un bord d'aile droite plutôt haut (beaucoup de bruit). Après prise en charge, sur la même modalité, l'énergie en dessous de la fréquence limite a considérablement augmenté (courbe en rouge, bord d'aile gauche haut) et celle au-delà de la fréquence limite, plutôt associée au bruit (F5 et au-delà) a considérablement régressé (bord d'aile droite bas).

La fréquence limite par défaut (fin du formant F4, début du formant F5, soit 4500 Hz pour les femmes d'après la table 3), correspond très bien au croisement des courbes avant/après du cas AS033.

### Durée d'enregistrement

La notion de « long terme » est toute relative : il s'agit en fait de disposer d'un enregistrement de durée suffisamment longue pour qu'un échantillon représentatif des phonèmes et structures syntaxiques de la parole du patient soit intégré dans le calcul, et que les paramètres extraits des calculs se stabilisent au cours du temps. Une durée entre 15 et 20 secondes est en général considérée comme suffisante. Elle dépend cependant de la précision exigée et du type de paramètre analysé. L'étude de [Peters 2017] montre que certains paramètres ne nécessitent que 5 secondes pour se stabiliser, d'autres plus de 10 secondes.

Notre étude concerne ici la parole pathologique, il est donc important de définir une tâche d'enregistrement du patient qui soit assez courte pour limiter le temps de passation et la fatigue du patient, mais suffisamment longue pour être fiable et objective, afin de permettre des comparaisons avant/après prise en charge clinique avec une précision raisonnable.

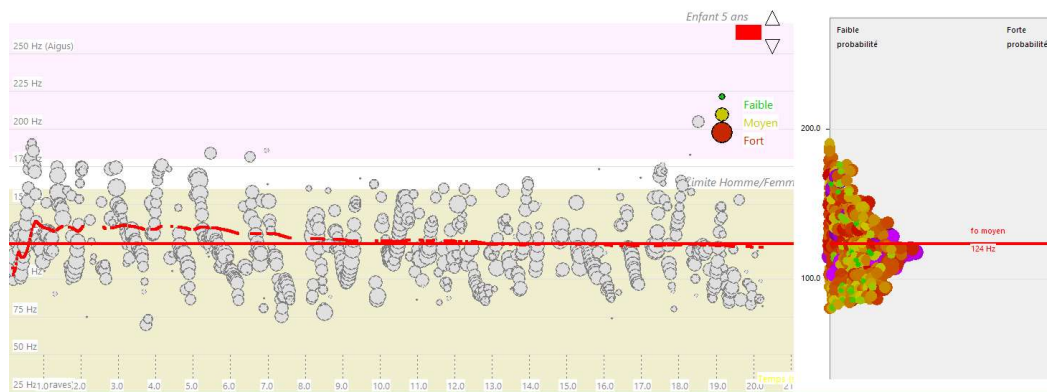


Figure 7 : Variation de la valeur moyenne de  $f_0$  au cours du temps (trait rouge pointillé) montrant la convergence vers une valeur stable après 10 secondes (Homme, 62 ans).

Dans le cas du discours spontané d'un homme de 62 ans, la valeur de  $f_0$  (trait rouge pointillé, figure 7) n'évolue quasiment plus après environ 10 secondes, le  $f_0$  moyen s'établit autour de 125 Hz (trait rouge continu de la figure 7).

### Synthèse sur les métriques

Un niveau de preuve relativement élevé est associé à la métrique L1-L2 (niveau de  $f_0$  par rapport à  $2 \cdot f_0$ ) et sa corrélation avec la sévérité de la dysphonie. Plus L1 est élevé par rapport à L2, plus la voix est altérée car l'énergie se concentre sur  $f_0$  et les harmoniques qui suivent sont atténuées, ce qui compromet notamment l'intelligibilité. La métrique alpha (Energie >1 KHz par rapport à <1 KHz) semble aussi relativement pertinente : plus elle est négative, moins bonne est la qualité de la voix.

Selon la théorie dite « du papillon », nous pouvons conduire deux raisonnements opposés concernant l'interprétation de la variation du spectre moyen au cours de la thérapie :

- La thérapie visera une augmentation de l'énergie en-dessous d'une fréquence limite  $f_{limite}$  (4-5 KHz), car elle caractérise une meilleure résonance, un meilleur rendement et une meilleure intelligibilité,
- La thérapie visera une réduction de l'énergie au-dessous de  $f_{limite}$ , car elle caractérise la présence de bruit de souffle ou de respirations fréquentes qui nuisent à la résonance, au rendement et à l'intelligibilité.

Les séances consistent à faire émerger des composantes situées dans les zones de sensibilité maximales de l'oreille afin de produire une sensation d'intensité à moindre effort. Ce travail a pour effet de masquer les bruits faible amplitude mais aussi de diminuer la quantité de bruit que ce soit en basse fréquence qu'en haute fréquence, par rapport au niveau des harmoniques.

Concernant les moments M1 et M2, la parole pathologique a tendance à réduire de façon significative la valeur de M1 et M2, et ce pour l'ensemble des publications qui comparent les spectres moyens des patients avec ceux des témoins. Les valeurs absolues de ces moments sont en revanche difficilement comparables entre elles, avec des écarts considérables selon les publications.

Nous pouvons aussi analyser la valeur du  $f_0$  moyen, alerter quand celle-ci s'écarte de la norme pour l'âge et le genre du patient. Nous pouvons aussi souligner les progrès lorsque le  $f_0$  moyen se rapproche des valeurs attendues grâce à la thérapie. Enfin, dans la mesure où le calcul du spectre moyen ne porte que sur les segments voisés, nous pouvons indiquer quel pourcentage de la parole a été inclus dans le calcul, ce qui est une métrique pertinente en lien avec la fluence. Plus le pourcentage est faible, plus il y a de pauses ou de bruit dans la parole.

## Le Spectre Moyen Universel

Le spectre moyen à long terme a été analysé par [Bryne 1994] pour 12 langues : anglais (plusieurs dialectes), suédois, danois, allemand, français (canadien), japonais, cantonais, mandarin, russe, gallois, cingalais et vietnamien. Le LTAS était similaire pour toutes les langues bien qu’il y ait certaines différences, d’amplitude plutôt faible. Les auteurs ont émis l’hypothèse d’un spectre moyen « universel » et donné des profils types pour hommes et femmes. La forme de ce spectre est toutefois différente des spectres regroupés dans la revue bibliographique de [Master 2006], ce qui laisse penser qu’il y a là encore une grande variabilité d’approches qui rendent le spectre dit « universel » difficilement exploitable en se basant uniquement sur la littérature scientifique.

Nous ne sommes pas nécessairement convaincus de l’intérêt de disposer d’un spectre moyen de référence, vu les critiques récurrentes qui lui ont été faites au fil des publications, justifiées par le côté unique de chaque physiologie et le danger de comparer sa parole avec une norme, aussi « universelle » soit-elle. Cependant, il nous a paru intéressant de développer une capacité à cumuler des spectres moyens de parole d’individus d’un même genre, afin de tenter d’identifier une courbe « typique ».

Nous avons donc regroupé 140 échantillons de parole d’enfants, 80 échantillons de parole de femme et 80 échantillons de parole d’homme, de durées de l’ordre de 10 secondes en moyenne, sur lesquels nous avons fait porter les calculs de spectre moyen cumulé. Les modalités d’enregistrement sont assez variées, incluant du texte lu, de la description d’images, de voix conversationnelle, de répétition de logatomes, etc. Les enregistrements sont de qualité, sans pollution sonore significative et proviennent de différentes sources. Concernant les enfants, nous utilisons en particulier le corpus de référence d’enfants de 5 ans [Remacle 2020] et les enregistrements effectués dans le cadre des mémoires de [Grenier 2020] et [Bardelang 2020] couvrant les tranches d’âges 6 à 8 ans principalement. Pour les hommes et femmes, nous disposons notamment du groupe témoin de [Papazian 2019].

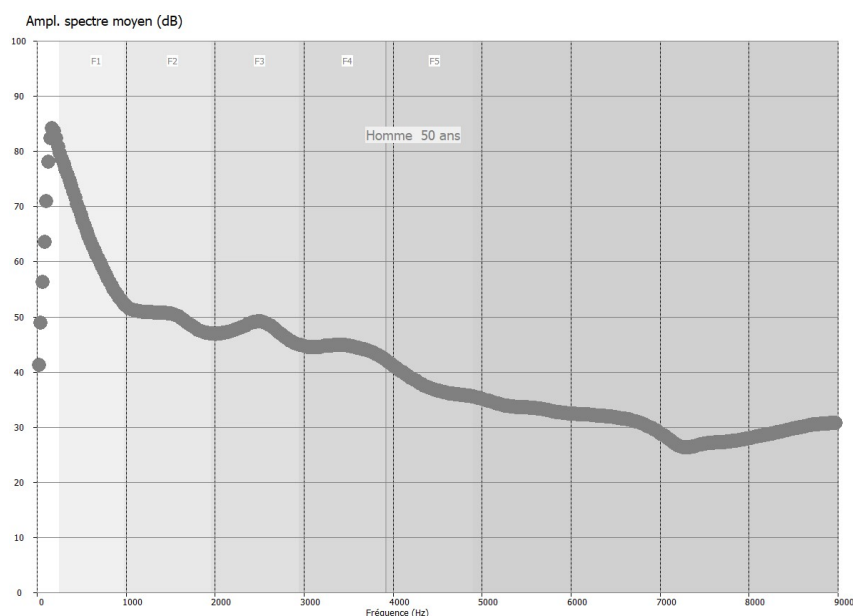


Figure 8 : Spectre moyen de la parole cumulé de 60 hommes



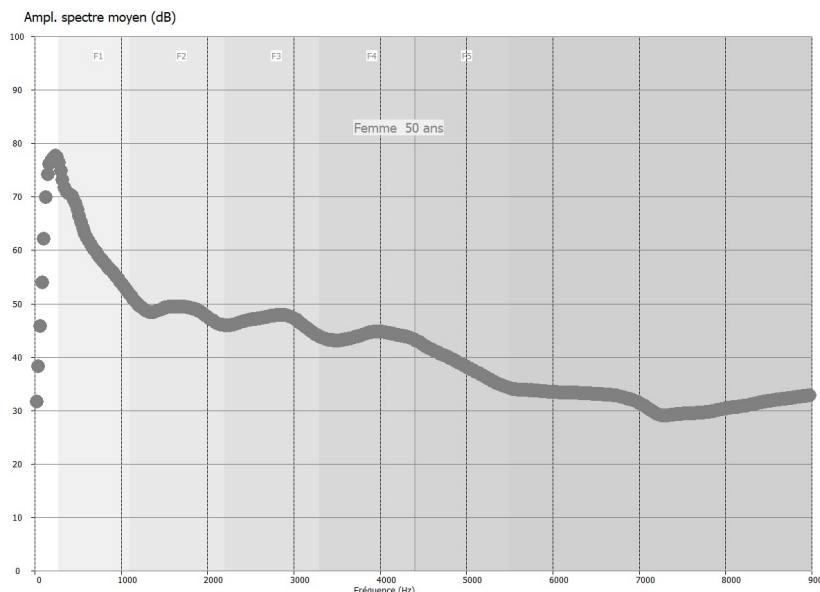


Figure 9 : Spectre moyen de la parole cumulé de 60 femmes

Les spectres moyens cumulés des hommes, femmes et enfants sont reportés dans les figures 8,9 et 10. On peut constater sur le spectre moyen est légèrement proéminent dans les zones correspondant aux formants, ce qui est plutôt rassurant dans la mesure où la notion de formants est basée sur l’hypothèse de l’existence de zones de résonance privilégiées, liés, en caricaturant un peu à la « résonance des tubes » [Sicard 2020]. Les valeurs de référence sont bien positionnées pour hommes et femmes. Concernant les enfants, nous avons mixé tous les âges disponibles de 4 à 12 ans, ce qui est sans doute la raison pour laquelle les zones de résonance ne sont pas facilement identifiables : une séparation par tranches d’âge de 2-3 ans maximum aurait probablement permis de mieux identifier les formants, dont les fréquences décroissent avec l’âge pour atteindre des valeurs comparables aux femmes à l’approche de la puberté [Audibert 2018] [Lee 1999] [Remacle 2017].

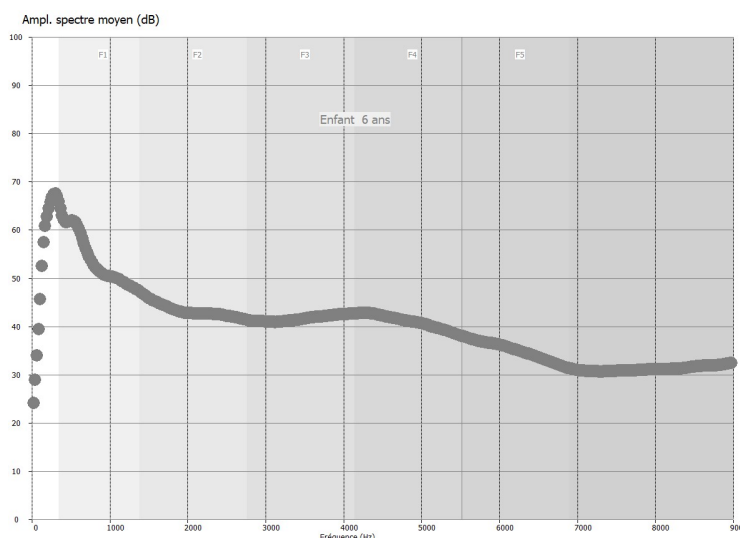


Figure 10 : Spectre moyen de la parole cumulée de 80 enfants

## Différence Homme/Femme

La différence majeure entre les spectres moyens des hommes et des femmes est liée à la différence de fréquence fondamentale  $f_0$ . Le spectre moyen de la parole d'un homme aura un pic prédominant autour de 120 Hz, celui d'une femme seulement autour de 220 Hz. Si l'on s'intéresse aux plus hautes fréquences, l'étude de [Cox 1988] révèle de légères différences dans la bande 800-1000 Hz et au-dessus de 5 KHz, avec un spectre plus énergétique (+3 à 5 dB) pour les femmes. Celle de [Mendoza 1996] portant sur 31 femmes et 24 hommes souligne un spectre moyen plus élevé dans la bande 3-5 KHz pour les femmes que pour les hommes, en particulier autour de 3 KHz et 4,1 KHz. L'inclinaison spectrale est globalement plus faible pour les voix féminines. Le niveau plus élevé serait lié au bruit en lien avec des reprises inspiratoires plus fréquentes pour les femmes que pour les hommes, ainsi qu'une voix plus soufflée [Klatt 1990], bien que de grandes disparités existent entre les sujets d'un même genre. Ces tendances se retrouvent à peu près dans l'illustration de la figure 11 qui superpose les spectres hommes et femmes et montre un pic  $f_0$  plus haut en dB et plus bas en fréquence pour les hommes et une courbe qui passe légèrement en dessous de celle des femmes au-delà de 4 KHz.

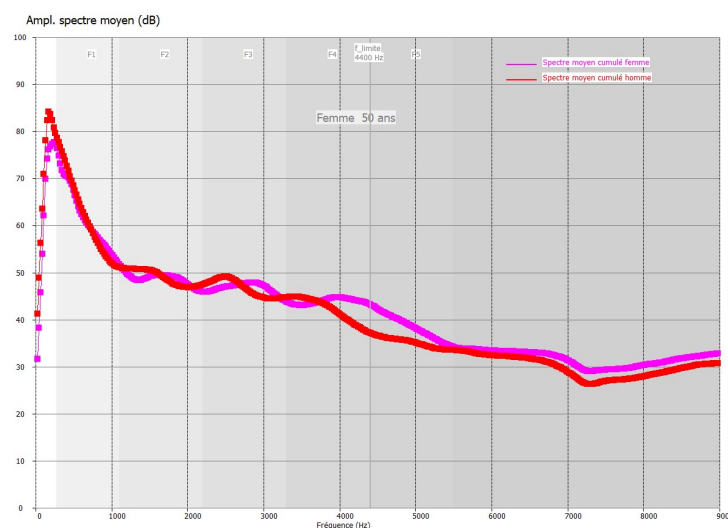


Figure 11 : Spectres moyens homme et femmes montrant certaines différences en termes de fréquence de résonance des formants et de différence d'énergie au-delà de F4

## Spectres moyens des enfants

Dans une étude portant sur trois cent vingt enfants âgés de 4 à 11 ans, [Sergeant 2008] a mis en évidence une augmentation de 5 dB de l'énergie spectrale dans toutes les fréquences inférieures à 6 KHz, ainsi qu'une réduction concomitante de l'énergie au-dessus de 6 KHz avec l'âge. Le développement des compétences vocales et l'accroissement des capacités pulmonaires semblent s'accompagner d'une meilleure efficacité laryngée. On retrouve donc « l'effet papillon », mais à une fréquence  $f_{limite}$  plus élevée que pour les femmes, ce qui n'est pas surprenant dans la mesure où  $f_0$  et  $F_n$  sont aussi significativement plus hauts en fréquence. La comparaison sur nos propres corpus montre des tendances similaires. Le spectre moyen d'un enfant très jeune est pauvre dans la zone F1..F3, bruyé à haute fréquence, et se rapproche graduellement du modèle femme en augmentant en âge. Cette tendance est illustrée à la figure 12 en cumulant les courbes de 3 à 12 ans. La courbe principale concerne les enfants de 6 ans. Elle est liée à la croissance et volume des cavités pharyngées, nasales et buccales.

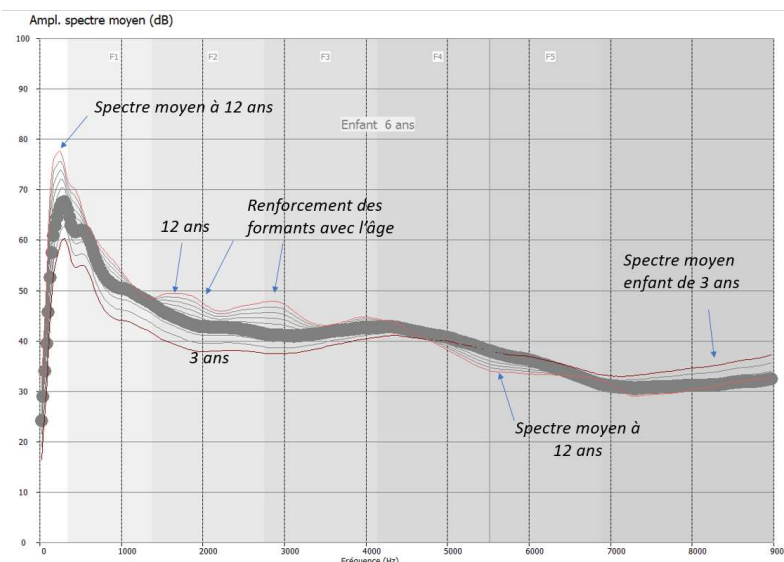


Figure 12 : Effet de l'âge sur le spectre moyen par extrapolation

## Voix âgée

On peut supposer que dans le cadre du processus de vieillissement normal, la voix a tendance à devenir moins forte, à perdre en résonance, être plus soufflée, avec un débit ralenti, ce qui peut nuire à l'intelligibilité. Linville et Rens (2001) ont été parmi les premiers à étudier les modifications de résonance suite au processus de vieillissement. Leurs analyses portant sur 80 locuteurs ont montré une diminution de la fréquence des formants, ainsi que de leur niveau d'énergie un peu plus faible, notamment autour de 3 KHz chez les femmes âgées.

L'étude de [Da Silva 2011] a en particulier montré une réduction significative (-3 dB) de tout le spectre de 500 Hz à 5 KHz accompagné d'une réduction similaire du rapport alpha en comparant les femmes âgées de 70 ans en moyenne par rapport au groupe âgé de 27 ans, ce qui est interprété comme un ralentissement du processus d'accolement des cordes vocales avec l'âge. De son côté, [Martinez-Nicolas 2022] a montré que les patients avec vieillissement naturel de la voix (sénescence non pathologique - NPS), avaient un spectre moyen autour de 76 dB, qui se réduisait légèrement avec l'âge, de l'ordre de 2 dB environ entre la tranche 60-70 ans et > 85 ans. L'étude portait sur 188 témoins.

Les seuils normal/altéré et altéré/pathologique doivent être ajustés au-delà de 60 ans, comme cela a été fait notamment pour les indicateurs d'altération du /a:/ tenu [Barbera 2016], de la diadococinésie [Sicard 2020b], ou du temps phonatoire [Sicard 2022c].

## Implémentation du spectre moyen dans DIADOLAB

DIADOLAB est un logiciel d'évaluation et de rééducation orthophonique de la parole développé par A. Menin-Sicard, Orthophoniste, et E. Sicard, Professeur à l'INSA Toulouse [Sicard 2023b]. Ce programme interactif permet d'évaluer la parole du patient selon deux macro-compétences intelligibilité et fluence, puis, au moyen d'outils interactifs, de travailler la conscience articulatoire, les praxies coordonnées, l'articulation, la phonologie, la diadococinésie et la prosodie. DIADOLAB est très utile à l'orthophoniste qui accompagne des patients atteints des pathologies telles que le déficit auditif, les troubles de l'articulation et de la parole, les dysarthries neurologiques, apraxies de la parole, la dyspraxie verbale, la dyslexie phonologique, le bégaiement, le bredouillement ou encore les troubles envahissants du développement et l'illettrisme.

Le spectre moyen existe depuis quelques années dans DIADOLAB en tant qu’outil de recherche, afin d’étudier les caractéristiques de la parole pour les patients ayant été traités pour un cancer ORL [Sicard 2017], puis pour l’analyse des oppositions de voyelles orales/nasales [Sicard 2022]. Il a notamment permis de comparer les spectres cumulés de témoins et de patients, ou encore d’identifier les zones de formants pour chaque voyelle orale et nasale, mettant en évidence les différences intergenre et les formants ayant le fort pouvoir séparateur. Jusqu’ici cependant, l’idée de mettre cet outil à disposition des cliniciens n’avait pas été évoquée.

### Indices spectrales

Les portions de silence sont ignorées afin de ne pas faire porter le calcul sur un signal qui ne contient pas de parole. Nous considérons donc essentiellement les portions voisées, ce qui exclut du spectre moyen les consonnes fricatives, les plosives, les reprises inspiratoires, qui peuvent être fréquentes dans le cas de voix pathologiques [Svilhic 2022]. En revanche, nous évaluons le pourcentage de portions voisées par rapport au signal total, ce qui est utile pour détecter des disfluences, les spasmes, allongements, blocages pouvant émailler les corpus des patients atteints de bégaiement développemental ou neurologiques. Les paramètres listés table 4 sont automatiquement évalués sur de la parole segmentée ou non.

VARIABLE	UNITE	DESCRIPTION	COMMENTAIRE
<b>VOISEMENT</b>	%		Particulièrement pertinent dans la parole disfluente
<b>FO</b>	Hz	Valeur moyenne du fondamental	Comparé à une norme fonction de l’âge et du genre
<b>L1-L2</b>	dB	Différence L1 de fo et L2 de 2.fo	Corrélié à la qualité de la voix et l’intelligibilité
<b>ALPHA</b>	dB	Différence entre le niveau d’énergie > 1KHz et < 1KHz	Corrélié à la qualité de la voix et l’intelligibilité
<b>LBF</b>	dB	Energie < 5 KHz	Zone de résonance F1..F3
<b>LHF</b>	dB	Energie > 5 KHz	Zone de bruit (>=F5)
<b>M1</b>	Hz	Moment d’ordre 1 dans la bande F1-F2	Espérance de la distribution du spectre
<b>M2</b>	Hz <sup>2</sup>	Moment d’ordre 2 dans la bande F1-F2	m2=écart-type (Hz), racine de M2

Table 4 : Indices acoustiques implémentées dans DIADOLAB pour l’analyse du spectre moyen

Les métriques implémentées dans DIADOLAB sont listées à la Table 4. Les données produites se basent sur des paramètres de configuration suivants :

- Le signal de parole, échantillonné à 22050 Hz est analysé en sections de 1024 points par Transformée de Fourier Rapide (*Fast Fourier Transform – FFT*), soit des portions de 47 ms de durée environ.
- Un fenêtrage de type Blackman est appliqué au signal pour limiter le bruit numérique.
- Les portions se chevauchent légèrement pour éviter une perte d’information, une approche assez classique pour la construction des spectrogrammes et spectres moyens.
- Le calcul des moments porte spécifiquement sur la zone F1-F2 où se situe principalement la discrimination des voyelles, en suivant l’argumentaire donné au paragraphe sur les métriques.

## Recommandations

Les spectres moyens ne sont comparables que dans des situations d'enregistrement et un protocole identiques. Nous recommandons en particulier de contrôler la distance bouche-micro à l'aide d'une règle ou d'un guide langue permettant de positionner la bouche du patient à environ 5-10 cm du micro. Nous recommandons de veiller à toujours placer le microphone dans l'axe de la bouche. Le choix du microphone a aussi son importance : un micro unidirectionnel (super-cardioïde comme le MB85 BETA, SM58) isole la voix du patient de l'environnement sonore, ce qui évite d'inclure des bruits parasites dans le spectre moyen.

Concernant les aspects protocoles, nous conseillons de donner la même consigne, qui peut être par exemple de demander au patient de se présenter sous la forme « Bonjour, je m'appelle ..., je viens chez l'orthophoniste parce que ... ». Un enregistrement de 10 à 15 secondes est suffisant pour faire converger les calculs vers des valeurs stables.

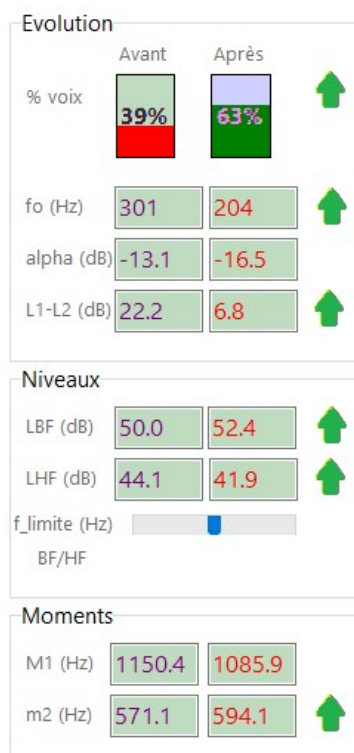


Figure 13 : Evolution des paramètres du spectre moyen, cas AS009. La flèche verte indique un progrès par rapport à la tendance attendue

## Aide à l'interprétation

Les métriques implémentées sont hétérogènes, non normalisées, d'unités différentes et varient dans un sens ou dans l'autre en fonction de l'altération de la parole. Afin d'aider le clinicien à l'interprétation des résultats, nous faisons apparaître une flèche verte en cas d'évolution favorable d'un indicateur, c'est-à-dire se rapprochant de la norme attendue. Par exemple, le % de voisement est perçu comme positif s'il augmente, ce qui est le cas de la figure 13. Le rapport L1-L2, au contraire est perçu comme positif s'il s'abaisse, ce qui est aussi le cas dans cet exemple. Certains indicateurs comme le *fo* moyen sont comparés à une valeur de référence en fonction du genre et de l'âge. La flèche verte indique un *fo* plus conforme à une femme de 64 ans après prise en charge [Sicard 2023].

Pour simplifier un peu plus l'exploitation des mesures avant/après prise en charge, des commentaires sont rajoutés au spectre moyen qui interprète en mots simples les progrès issus des mesures objectives. Le tableau suivant détaille le lien entre l'évolution des indices et les commentaires associés.

INDICE	FAVORABLE SI	MESSAGE SI EVOLUTION FAVORABLE
<b>% VOISEMENT</b>	% augmente	Pourcentage de parole voisée plus élevé
<b>FO</b>	Se rapproche de la norme pour le genre et l’âge	<i>f</i> <sub>0</sub> plus proche de la norme
<b>ALPHA L1-L2</b>	Augmente Diminue	Rapport sup/inf 1KHz plus favorable Energie moins concentrée sur fo
<b>LBF</b>	Augmente	Zone F1-F3 plus énergétique
<b>LHF</b>	Diminue	Bruit haute fréquence moins élevé
<b>M1, M2</b>	Augmente	Résonance plus marquée sur F1-F2

Table 5 : Aide à l’interprétation du spectre moyen

### Spectre moyen en bilan



Le spectre moyen est accessible dans DIADOLAB depuis le module *Phonologie*. Le spectre moyen a été conçu pour permettre d’observer deux courbes, typiquement avant puis après prise en soin, afin d’analyser l’évolution des performances. Nous donnons à la figure 14 le spectre comparatif avant/après prise en charge de la voix conversationnelle sur la base d’un protocole de type « *Bonjour, je m’appelle..j’ai ....âge et je viens chez l’orthophoniste parce que ....* » où le patient doit expliquer sa gêne. Le cas présenté concerne une femme de 64 ans avec dysphonie dysfonctionnelle (Classification 8100), correspondant au cas AS009 de la base de données de voix du LURCO [ERU15 2018]. Le profil en violet représente le spectre avant prise en charge, le profil en rouge après prise en charge.

On note plusieurs points positifs :

- Un pourcentage de parole voisée significativement plus élevé : pathologique avant, normal après
- Le fondamental de la parole *f*<sub>0</sub> plus proche de la norme
- Une énergie moins concentrée sur *f*<sub>0</sub> (L1-L2 réduit)
- Une zone F1-F3 plus énergétique : augmentation du spectre à gauche de *f*<sub>limite</sub>
- Un bruit haute fréquence moins élevé : réduction du spectre à droite de *f*<sub>limite</sub>

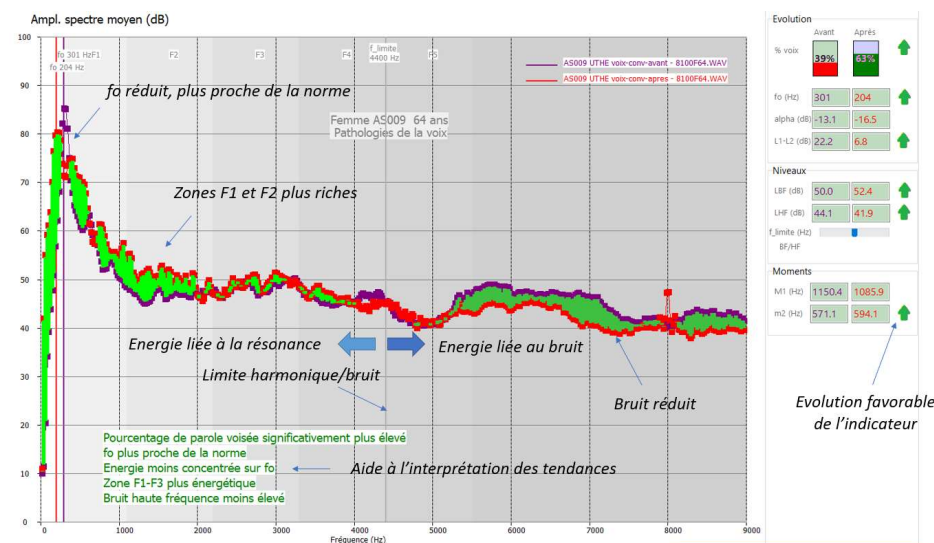


Figure 14 : Comparaison du spectre moyen avant et après prise en soin, cas AS009 de la base de données de voix du LURCO [ERU15 2018]



## Spectre moyen en rééducation

Une possibilité offerte aux orthophonistes est de produire un spectre instantané pour l'analyse en temps réel de la parole. Cette fonctionnalité existe déjà dans le spectrogramme [Sicard 2021] ou dans le module hauteur tonale, et ces outils de feedback en temps réel s'avèrent très utiles pour motiver et guider le patient vers l'amélioration de la voix et de la parole. Un exemple d'utilisation du spectre moyen suite à une acquisition en temps réel de parole spontanée est reporté figure 15. Le patient est enregistré sans effort de résonance (en haut à gauche). Puis il améliore ses performances en recherchant la résonance, ce qui se traduit par une augmentation significative de l'énergie dans les zones des formants F1-F3, et une réduction de l'énergie à partir de F5 qui est assimilée à du bruit.

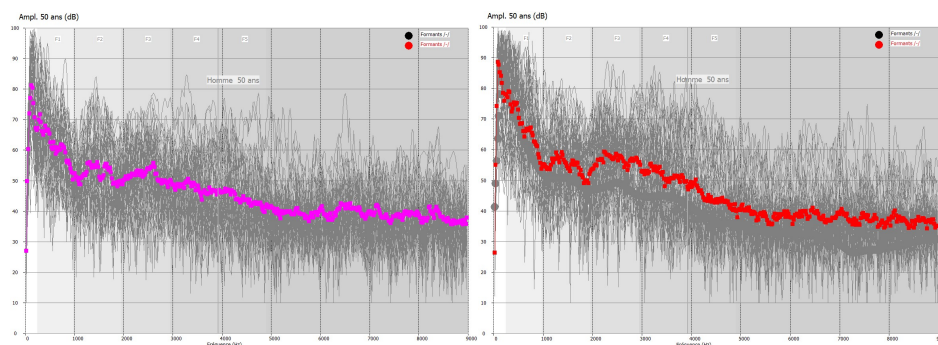


Figure 15 : spectre moyen obtenu en temps réel, sans effort (à gauche) et avec effort de résonance (à droite)

## Parole pathologique

Nous reportons dans les paragraphes suivants des observations publiées dans la littérature scientifique concernant diverses pathologies. Nous illustrons les propos par nos propres études de cas avant/après prise en charge permettant de positionner les performances du patient par rapport aux résultats publiés.

### Autisme

Selon [Bonneh 2011] qui a analysé la parole de 83 enfants dont 41 avec troubles du spectre autistique (TSA) et 42 témoins, il existe très peu de différence entre les spectres moyens à l'exception de la zone autour de la fréquence fondamentale  $f_0$ , autour de 250 Hz en moyenne chez les enfants. Le groupe contrôle avait un spectre moyen cumulé faisant apparaître un pic assez net dans la zone  $f_0$ , tandis que le groupe avec TSA montrait un pic beaucoup moins marqué, lié à la variation souvent inappropriée de  $f_0$  dans le contour prosodique des patients, amenant à une distribution statistique de  $f_0$  plus étalée. Selon la revue de littérature de [McCann 2003] concernant l'analyse de la prosodie pour la parole de patients avec TSA, il existe des contradictions importantes dans les résultats obtenus selon les études, ne permettant pas de dégager une tendance générale ni des caractéristiques propres à la dysprosodie. Si l'on se concentre sur les basses fréquences, le spectre moyen donne donc une indication de la variabilité de  $f_0$ , sans toutefois pouvoir clairement indiquer s'il s'agit d'une parole dysprosodique.

### Bégaiement

Le cas d'une femme de 22 ans avec un bégaiement sévère avec spasmes laryngés, associé à des troubles expressifs et un dysfonctionnement vélaire illustre la possible utilisation du spectre moyen dans les troubles de la fluence. La comparaison du spectre moyen avant et après prise en charge est reportée figure 16. L'évolution la plus marquante est le pourcentage de parole voisée qui passe de 30% à 50%. Le fondamental s'élève aussi, s'éloignant de la transition homme/femme. La phonation prolongée ayant permis de limiter les spasmes, blocages et améliorer la coordination pneumophonique.



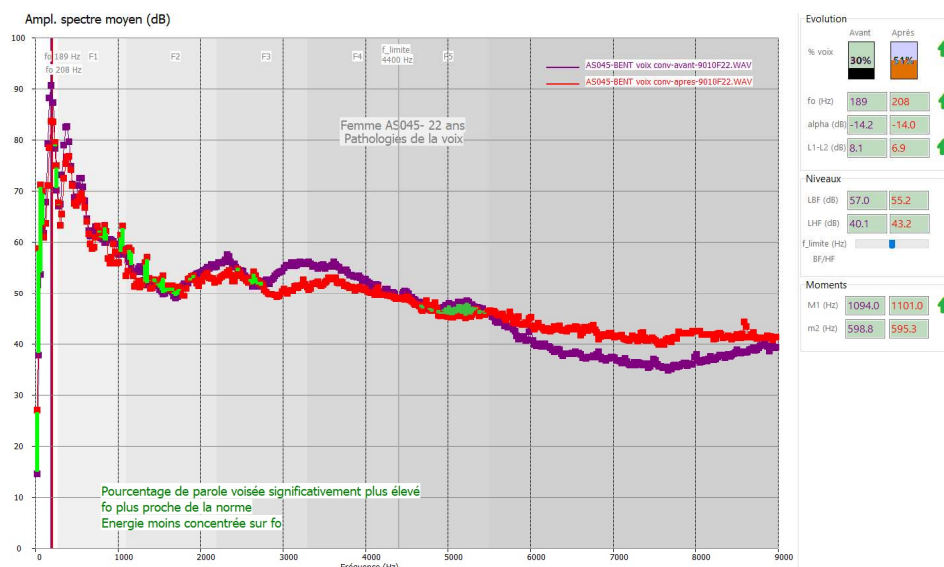
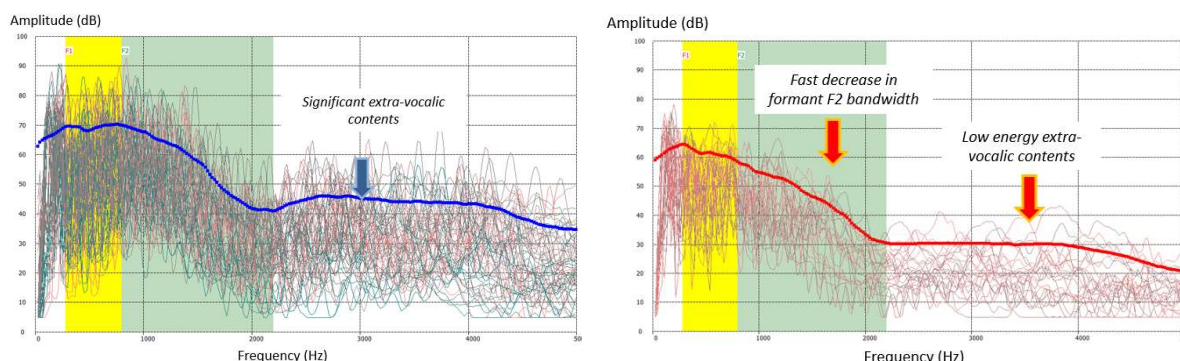


Figure 16 : Comparaison des spectres moyens avant/après prise en charge, cas AS045 [ERU15]

### Cancer ORL

Dans le cadre du projet C2SI, nous avons pu comparer le spectre moyen de 23 patients ayant été pris en charge pour un cancer ORL et 60 témoins [Sicard 2017]. La moyenne cumulée du spectre a permis de mettre en évidence un appauvrissement significatif de la zone F2-F4, ce qui a une influence sur l'intelligibilité de la parole. Si l'on observe plus spécifiquement la zone F3 (2-4 KHz), une réduction de 15 dB des composants haute fréquence est observée pour les patients par rapport aux témoins, ce qui est considérable (Figure 17).



(a) 60 cas de voix normales

(b) 23 cas de voix après traitement d'un cancer ORL

Figure 17: Spectre moyen de la parole de 60 témoins (a) et de 23 patients ayant été traités pour un cancer ORL (b) mettant en évidence une décroissance rapide du spectre moyen et un faible contenu extra-vocalique chez les patients [Sicard 2017]

### Dysarthrie

L'étude de [Svihlik 2022] a montré que les moments spectraux permettent de différencier 23 patients atteints de Maladie de Parkinson et 23 témoins appariés en genre et en âge sur une tâche de lecture et un monologue. Le centre de gravité M1 était nettement plus haut chez les témoins que chez les patients, l'écart-type spectral m2 était aussi significativement plus haut. En revanche, l'asymétrie m3 et l'aplatissement m4 étaient plus bas pour les témoins que pour les cas MP.

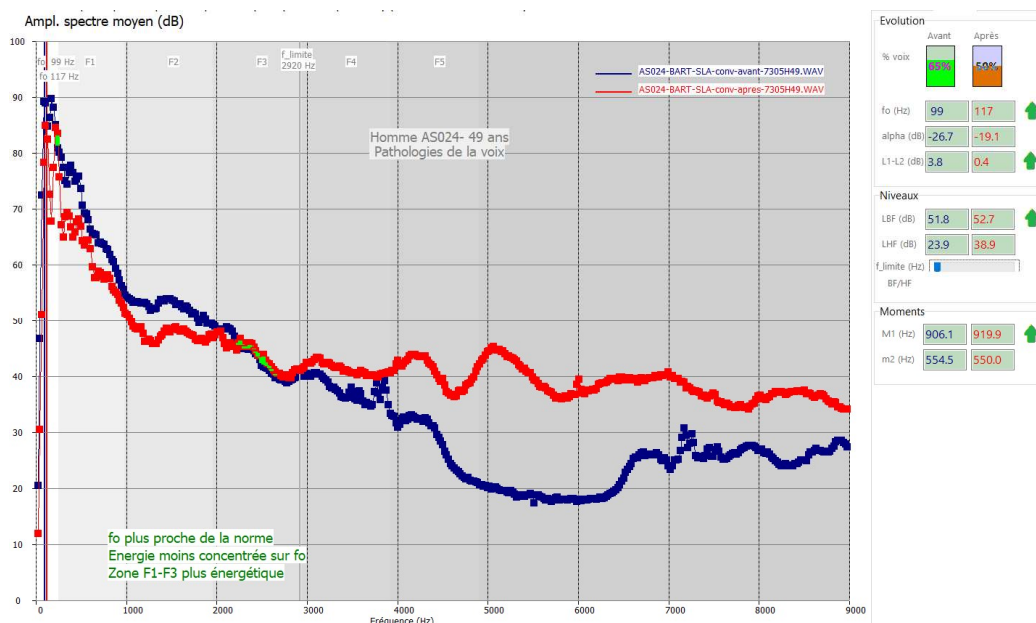


Figure 18 : Evolution de la parole spontanée du cas AS024 atteint de SLA [ERU15]

L'étude de [Tjaden 2010] a examiné la relation entre les mesures du spectre moyen à long terme (LTAS) pour les locuteurs atteints de la maladie de Parkinson (MP) et de la sclérose en plaques (MS) avec l'échelle d'estimation de la sévérité de la dysarthrie perçue de la parole. Des locuteurs témoins sains ont également été étudiés. Les résultats n'ont pas indiqué de différence entre les groupes MP et MS, mais les auteurs suggèrent que le spectre moyen peut s'avérer utile pour évaluer les changements liés au traitement ou à la progression de la maladie.

Nous donnons à la figure 18 le cas du patient AS024 atteint de sclérose latérale amyotrophique (SLA). La dégradation rapide de l'état clinique et l'augmentation de la sévérité de la dysarthrie sont observées entre les deux enregistrements espacés de deux ans, malgré une prise en charge régulière. Le patient fait des efforts considérables pour garder une certaine intelligibilité, mais on observe l'inverse de ce qui serait attendu : le % de voisement passe de normal à altéré, l'énergie dans la zone F1-F3 est en régression, et la zone au-delà de F5 est en très nette augmentation, que l'on peut attribuer à de la constriction et du bruit parasite qui nuisent à l'intelligibilité. A noter que le curseur de fréquence limite a également baissé, ce qui peut être corrélé au degré de sévérité de la maladie.

### Dyspraxie verbale

L'analyse de la parole de patients dyspraxiques montre différentes caractéristiques :

- Débit de parole lent avec allongement des transitions [Kent 1983],
- Variation restreinte de l'intensité des syllabes, mouvements lents et imprécis des articulateurs,
- Mauvaise synchronisation sur des enchainements de consonnes ou processus de parole coordonnés [Sicard 2022d],
- Difficultés d'initiation des mouvements [Auzou 2000] souvent caractérisées par des faux départs, imprécisions, erreurs et disfluences sur des séquences sonores complexes.

L'utilité du spectre moyen dans le cas spécifique de la dyspraxie verbale est abordée dans la revue de littérature de Terband (2019) sous l'angle des moments et des rapports d'énergie, qui permettent de différencier les patients et témoins.

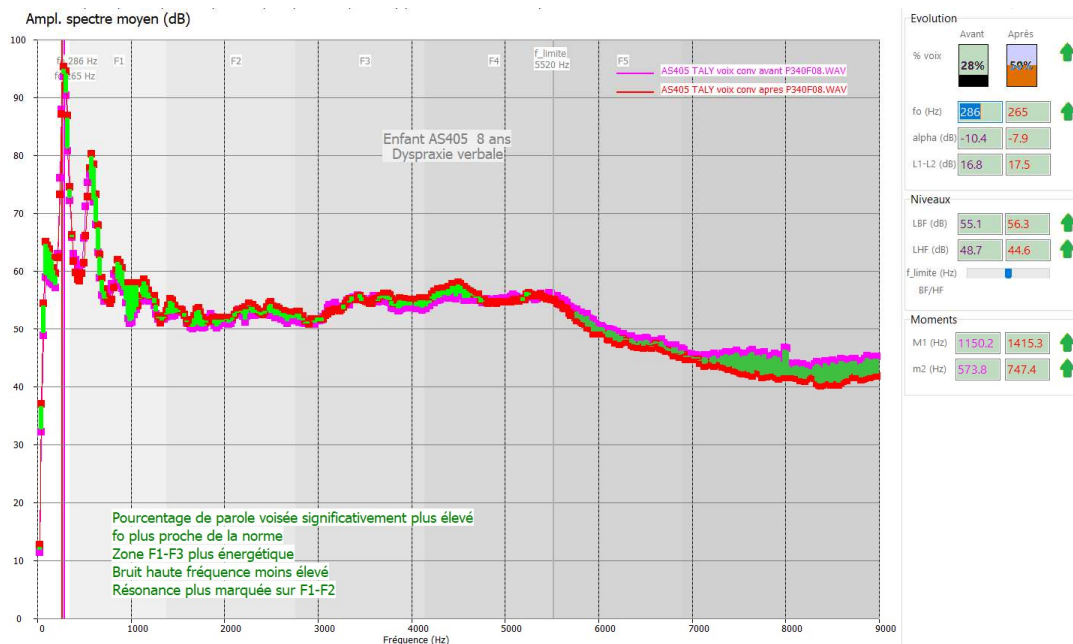


Figure 19 : Evolution de la parole spontanée du cas de dyspraxie verbale AS405

Le cas AS405 d'une fille de 8 ans avec dyspraxie verbale montre, avant prise en charge, un débit particulièrement lent avec de nombreuses pauses et reprises inspiratoires. Après prise en charge à raison de 2 séances en moyenne par semaine, la patiente a nettement progressé avec un taux de voisement passant de 28 à 50%, un *fo* plus adapté, une zone F1..F3 enrichie et une diminution du bruit haute fréquence (Figure 19).

## Dysphonies

L'étude de [Tanner 2005] a permis de comparer des échantillons de voix avant et après traitement de 93 patients atteints de dysphonie fonctionnelle de sévérité diverses à l'aide des spectres moyens. Deux améliorations dans la forme du spectre moyen ont été relevés grâce à la thérapie vocale : tout d'abord une augmentation de l'ampleur des basses fréquences, notamment autour de *fo*, et d'autre part une réduction d'énergie à partir de 5 KHz associée à une réduction du souffle dans la voix et une augmentation du rendement.

Une évolution similaire à la tendance décrite par [Tanner 2005] peut être observée sur le cas de dysphonie AS009 de la base de données de voix en ligne du LURCO [ERU15], décrit dans [Menin-Sicard 2021b] : des progrès apparaissent en vert en dessous de 5 KHz, signe d'une augmentation de l'énergie de la parole, tandis que l'énergie diminue au-delà de 5 KHz, qui apparaît en vert aussi car elle est assimilée à la réduction du bruit. Cette image illustre assez nettement la théorie du papillon exposée plus haut.

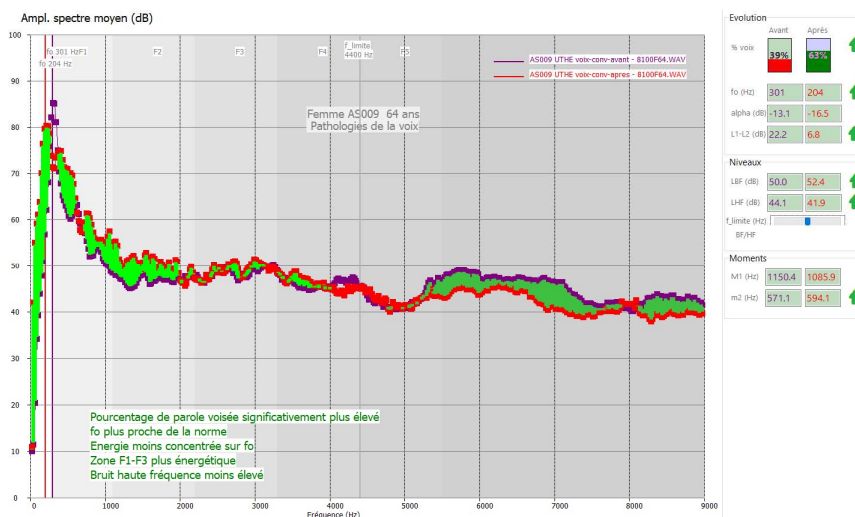


Figure 19 : Evolution de la parole spontanée du cas de dysphonie AS009 [ERU15][Menin-Sicard 2021b]

Les performances vocales d'un groupe de 30 participants masculins âgés (Moyenne 73 ans) n'ayant aucun antécédent de troubles de la voix ont été comparées par [Soumya 2022] à 30 participants masculins de même âge moyen, avec des symptômes vocaux liés à la dysphonie. Une corrélation élevée a été observée entre L1-L2 et le degré de sévérité de la dysphonie évaluée avec le GRBAS [Hirano 1981]. Le rapport est de l'ordre de 11 dB pour les voix normales ou peu altérées, et s'abaisse autour de 17 dB pour le groupe de dysphoniques.

Dans l'exemple du cas AS009 de la figure 19, l'énergie est beaucoup moins concentrée sur *fo* après rééducation car la différence entre les pics L1 et L2 est passé de 22 dB (dysphonie marquée, L1-L2 pathologique) à 7 dB (parole normale, L1-L2 normal). Une autre étude faite par [Antonetti 2020] portant sur 49 voix dysphoniques et 53 témoins suggère que ce type de métrique fournit des informations pertinentes sur la fuite d'air à travers la glotte fermée, aidant à l'analyse objective de l'efficacité phonatoire. Dans le cadre du diagnostic différentiel, Houtz (2010) a comparé le spectre moyen de 51 sujets avec dysphonie spasmodique et 41 sujets avec dysphonie par tension musculaire. Aucune différence spectrale significative n'a été identifiée entre les patients des deux groupes.

### Fentes palatines

Les caractéristiques spectrales de 40 enfants (20 filles et 20 garçons) présentant une fente palatine ont été comparées par [Yang 2018] avec celles d'enfants normaux appariés en nombre, genre et âge, sur la base de la lecture de phrases d'une durée de 60 secondes environ. Aucune différence significative n'a été observée entre le groupe témoin et le groupe clinique jusqu'à 2500 Hz. En revanche, des différences importantes ont été observées au-delà de 2500 Hz, notamment chez les garçons, avec un spectre moyen significativement plus bas pour les patients par rapport aux enfants normaux. Ces résultats sont assez surprenants car de nombreux auteurs ont pu identifier des formants spécifiques à l'hypernasalisation qui auraient dû influencer sur la forme du spectre moyen.

Nous avons cherché dans [Sicard 2022] à identifier à l'aide du spectre moyen les différences entre les formants des voyelles orales et nasales, en étudiant les oppositions /au-on/, /a-an/ et /è-in/. Nous avons fait porter un calcul de formants sur le spectre moyen, afin de dessiner les nuages de points des formants F1, F2 et F3 selon le genre et le type d'opposition. Concernant les voyelles /a-an/ enfants, on note un écart significatif notamment concernant F2, en fréquence mais aussi en amplitude, ainsi qu'un écart assez significatif pour F1 et F3 (Figure 21). Nous n'avons cependant pas encore conduit d'étude sur la parole spontanée à l'aide du spectre moyen.

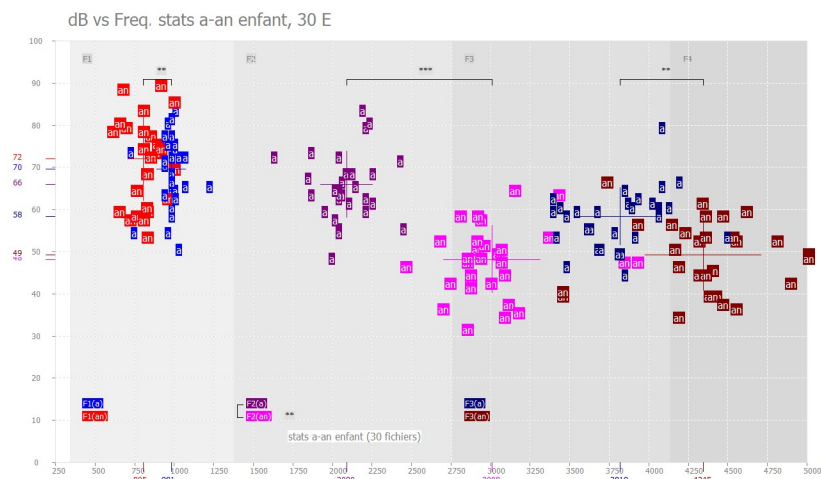


Figure 21 : Formants F1, F2 et F3 des oppositions /a-an/ enfants normo-phoniques (NP)

### Paralysie unilatérale des cordes vocales

L'étude de [Hartl 2003] a porté sur l'analyse comparative de différents paramètres acoustiques et aérodynamiques objectifs entre un groupe de 8 patients atteints de paralysie unilatérale des cordes vocales et un groupe témoin apparié. Le spectre moyen à long terme a été calculé pour 40 s de texte. L'énergie haute fréquence était toujours significativement plus élevée par rapport à la basse fréquence pour le groupe patients, probablement en lien avec du bruit lié à un mauvais accolement des cordes vocales.

Nous illustrons ces propos avec le cas AS033 [ERU15] concernant une femme de 61 ans, secrétaire de profession, avec une atteinte du nerf laryngé supérieur, associé à une désynchronisation des CV, dont la prise en charge orthophonique est décrite dans [Menin-Sicard 2021b]. On note des progrès spectaculaires, notamment dans la zone F1.F2 avec un renforcement de l'ordre de 10 dB, une valeur très au-dessus des progressions habituellement constatées ou publiées. De manière symétrique, le bruit très perceptible avant prise en charge provoque dans le spectre moyen un dôme entre 6 et 8 KHz, et se retrouve très nettement réduit après rééducation (-10dB).

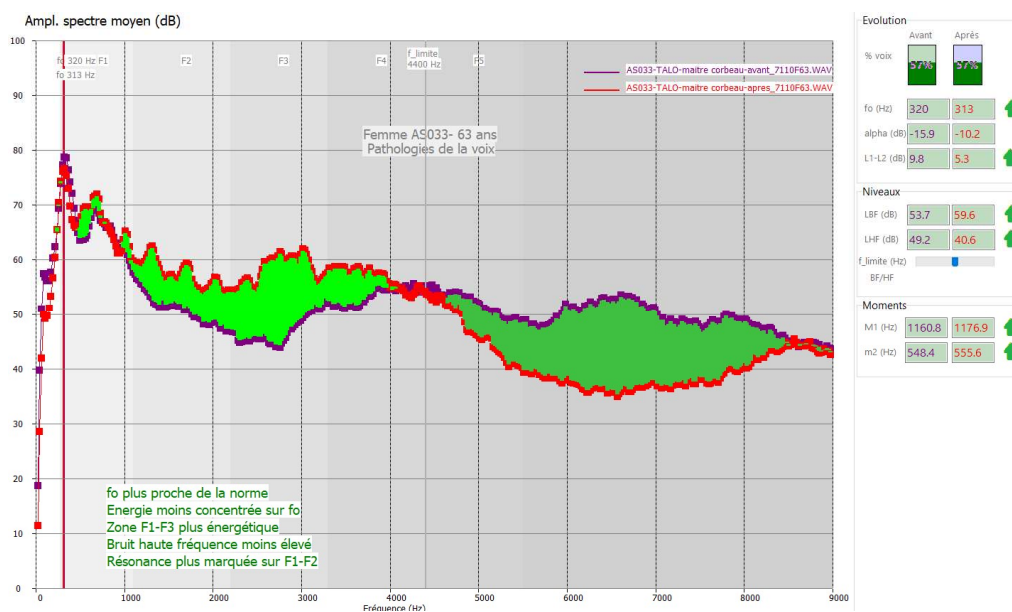


Figure 22 : Evolution de la parole spontanée du cas de dysphonie AS033 [ERU15][Menin-Sicard 2021b]



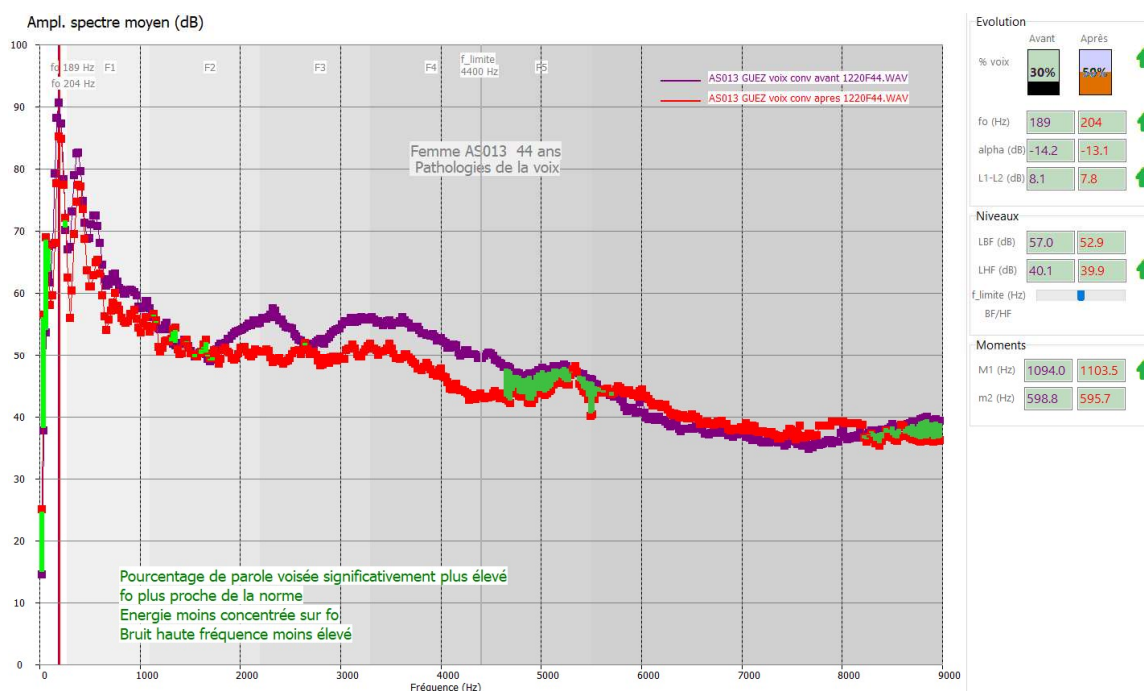


Figure 23 : Evolution de la parole spontanée du cas d'œdème de Reinke AS013 [ERU15]

## Polypes & œdèmes

La comparaison par [Stajner-Katusic 2012] des courbes de spectres moyens préopératoires et postopératoires de 20 patients (10 hommes et 10 femmes) atteints de polypes et ayant une voix soufflée montre que la fréquence fondamentale  $f_0$  postopératoire est plus conforme, que les fréquences jusqu'à 3 KHz sont plus intenses (+5 dB) et que les composantes hautes fréquences, à partir de 4 KHz, sont plus faibles (-3 dB) par rapport aux état préopératoire.

Le cas AS013 [ERU15] concerne une femme de 44 ans avec un œdème empêchant l'accolement glottique. La prise en charge permet d'augmenter significativement le pourcentage de parole voisée, éloigne  $f_0$  de la zone de transition Homme/Femme et s'accompagne d'une légère réduction du bruit haute fréquence (Figure 23). Un autre cas AS036 correspondant à une parésie récurrentielle bilatérale du nerf laryngé, décrit dans [Menin-Sicard 2021b], montre aussi une augmentation de l'énergie dans F1-F3, et un plus grand pourcentage de sons voisés.

## Voix enrouée - Nodules

Les différences spectrales entre 40 témoins sans atteinte ORL et 40 patients souffrant d'une voix enrouée suite à des nodules, polypes ou œdèmes des cordes vocales ont été étudiées par [Narasimhan 2017]. Les valeurs des amplitudes  $L_n$  des harmoniques et celles du pic maximum des formants  $L_{Fn}$  ne figurent malheureusement pas dans la publication. Les auteurs indiquent que toutes les différences étudiées (dont L1-L2) se sont révélées statistiquement significatives.

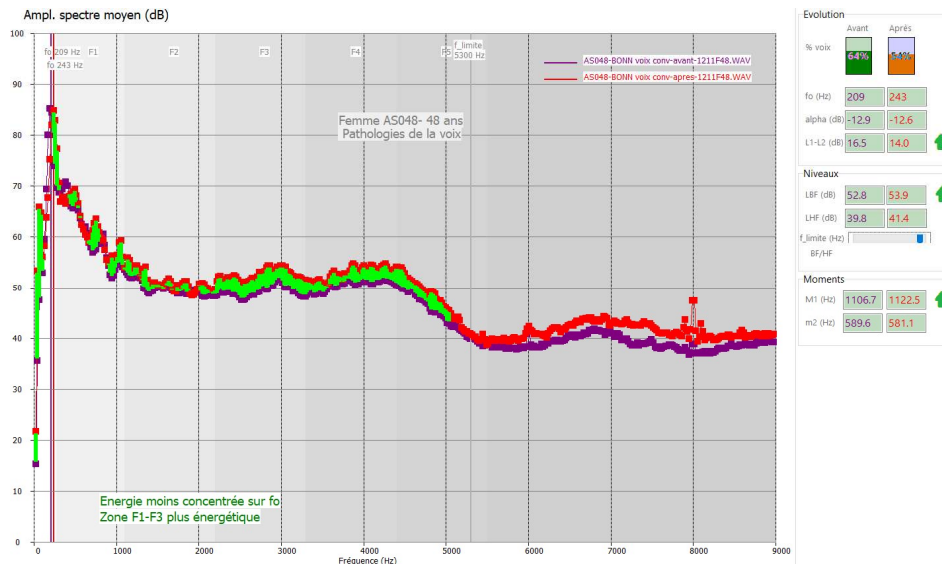


Figure 24 : Evolution de la parole spontanée du cas de nodule AS048 [ERU15][Menin-Sicard 2016]

Le cas AS048 décrit dans le chapitre 4 de l'ouvrage [Menin-Sicard 2016] concerne un nodule des cordes vocales, avec une voix enrouée à des moments globalement peu fréquents. On constate après prise en charge une réduction de la distance L1-L2, ainsi qu'une remontée globale de l'énergie F1-F4 qui améliore l'intelligibilité (Figure 24).

## Discussions

Contrairement aux notations qui ont fait l'objet d'un consensus [Titze 2015], aucune procédure standardisée n'existe à notre connaissance pour la mesure du spectre moyen. L'absence de métrique normative, de corpus de référence et de guide de bonne pratique limite la pertinence des comparaisons des résultats entre les différentes études.

Selon l'algorithme utilisé, les valeurs extraites ne sont pas les mêmes. Certains écarts constatés dans la littérature sont impressionnants, comme M1 et M2 du tableau 3. Si l'on peut espérer une cohérence entre les outils sur la valeur de la fréquence fondamentale  $f_0$  et son niveau L1, l'extraction des paramètres des formants tels que la largeur BFn et le niveau LFn est aussi sensible à l'algorithme utilisé. De grandes disparités de valeurs peuvent aussi être observées dans le cas de voix pathologiques.

Les études portant sur les caractéristiques de la parole des enfants, notamment dans les phases d'acquisition des phonèmes et des processus de parole, sont peu nombreuses. Il en est aussi de même pour la parole âgée.

Si le fichier n'est pas segmenté, le spectre moyen est calculé sur toutes les portions sonores significatives. Les fricatives sourdes ne sont donc pas incluses dans le calcul. Or la qualité et la tonicité de la constriction ont un impact important sur l'intelligibilité. Cette mesure ne prend pas en compte les contrastes d'amplitude des plosives liée à l'efficacité des articulateurs (lèvres/voile/langue) qui assurent la mise en pression nécessaire à l'explosion. Cette explosion interrompt naturellement le flux d'air provoquant un silence. Le contraste silence/voisement donne un contraste en lien direct avec le degré de contamination.

Si le fichier est segmenté, le spectre moyen est établi sur les segments de voyelles et de consonnes voisées. Des différences non négligeables peuvent être observées entre le spectre moyen avec ou sans segmentation, car certains bruits d'énergie significative, avec un  $f_0$  plausible, peuvent être inclus dans le calcul sans support de segmentation.



PARAMETRE		SPECTROGRAMME	SPECTRE MOYEN	COMMENTAIRE
<b>EVOLUTION AU COURS DU TEMPS</b>		Très précise	Impossible	Le spectre moyen ne permet pas de situer les évènements dans le temps
<b>PRECISION DE MESURE D'INTENSITE</b>		Moyenne	Elevée	Le spectrogramme se base sur un jeu de couleurs
<b>EVALUATION DES FORMANTS</b>		Possible	Facile	Le spectrogramme à bande étroite montre les harmoniques mieux que les formants
<b>EVALUATION DES RAPPORTS D'ENERGIE</b>		Approximative	Très précise	Les rapports HB/BF sont difficiles à évaluer sur le spectre
<b>EVOLUTION AVANT/APRES</b>		Possible en interprétant le jeu de couleurs	Evaluation précise en dB	
<b>METRIQUES DISPONIBLES</b>		Taux de constriction, évaluation des disfluences	de % de voisement Fo moyen Rapport alpha Rapport L1-L2 Rapport HF/BF	Le spectre moyen donne une vision globale de la qualité de la parole

Table 6 : comparaison entre les performances du spectrogramme et du spectre moyen

Il est donc recommandé de comparer des enregistrements sonores entre eux, soit segmentés selon la même stratégie [Menin-Sicard 2021], soit non segmentés selon la même modalité de parole, afin d'augmenter la pertinence des observations différentielles. En cas de changement de microphone, de carte son ou de modification de réglages au cours de la prise en charge orthophonique, il devient délicat de comparer les spectres moyens.

Les caractéristiques comparatives du spectrogramme et du spectre moyen sont regroupées à la table 6. Le spectre moyen est une méthode d'analyse acoustique simple de la parole continue, qui ne nécessite pas de segmentation ni d'intervention spécifique de la part du clinicien. Des indices acoustiques relativement pertinentes sont automatiquement calculés et démontrent une certaine sensibilité à l'altération de la parole, notamment pour les dysphonies, les dysarthries et de manière plus générale, toute atteinte des cordes vocales. Nous le considérons comme un outil complémentaire des mesures objectives sur le /a:/ tenu et la sirène usuellement proposé dans les bilans de voix, ou des répétitions de syllabes proposées dans les bilans de parole. Le spectre moyen peut être calculé sur des phrases types, lecture de texte ou voix conversationnelle sans difficulté particulière. Cette idée de complémentarité est aussi défendue par [Maryn 2010] qui suggère de combiner plusieurs types d'échantillons de voix pour affiner l'analyse acoustique et la précision des mesures.

Calculer un spectre moyen et non plusieurs spectres instantanés comme dans le spectrogramme fait perdre la dimension temporelle, mais « résume » au travers d'une simple courbe beaucoup d'informations révélatrices du placement, de la qualité de la voix, et de l'intelligibilité de la parole. L'implémentation du spectre moyen dans DIADOLAB donne des résultats intéressants, bien corrélés aux résultats obtenus à l'aide de métriques usuelles telles que les indicateurs d'altération du /a:/, le phonétogramme ou l'étendue vocale. Cette approche complémentaire et efficace pourrait jouer un rôle dans le bilan de voix et de la parole en pratique clinique dans les années à venir.

## Conclusion

Nous avons étudié les caractéristiques des spectres moyens de la parole, en analysant une quarantaine de publications présentant des analyses de patients et de témoins, qui ont permis d'extraire les indices les plus intéressants pour guider l'orthophoniste clinicien dans l'utilisation de cette représentation, simple et complémentaire du spectrogramme. Nous avons mis en évidence une grande diversité de métriques applicables à la caractérisation des spectres moyens, ainsi qu'une absence de repères normatifs solides applicables selon le genre et l'âge des patients, ce qui limite la portée et la pertinence des résultats obtenus.

Une implémentation du spectre moyen a été proposée dans la plateforme DIADOLAB, en justifiant les choix effectués d'après les données les plus probantes, avec une illustration de l'évolution des performances avant/après prise en soin, concernant notamment des pathologies telles que le bégaiement, le cancer ORL, la dysarthrie, les dysphonies, les fentes palatines, les paralysies des cordes vocales, ainsi que les nodules, polypes et œdèmes. Des données convergentes, concordantes et complémentaires des indices acoustiques extraites des voyelles tenues, des sirènes ou de la parole segmentée ont été obtenues. Bien que la preuve de l'utilité du spectre moyen en clinique orthophonique n'ait pas pu être formellement apportée, nous pensons que cet outil est le « chaînon manquant » de l'évaluation de la voix et de la parole, qu'il peut être un médiateur utile vis-à-vis du patient pour mesurer l'évolution des progrès en thérapie et qu'il peut contribuer à donner une vision pertinente du rendement de la voix ainsi que de l'intelligibilité de la parole.

## Remerciements

Nous tenons à remercier les orthophonistes contributrices à l'ERU 15 « voix pathologique » et de l'ERU 46 « parole pathologique » du laboratoire LURCO ainsi que les orthophonistes qui ont accepté d'enregistrer des témoins et des patients, ou de co-encadrer des mémoires en lien avec la voix, la parole et leurs pathologies.

Nous remercions aussi l'ensemble des orthophonistes, étudiants et chercheurs qui nous ont accordé leur confiance en utilisant nos méthodologies, approches et outils dans le cadre de leur pratique clinique ou de leurs travaux scientifiques.

## A propos des auteurs



**Etienne SICARD** est professeur à l'INSA Toulouse en électronique, informatique et traitement du signal. Il est directeur de recherches de l'équipe ERU 46 sur la parole pathologique au Laboratoire LURCO. Il a été chercheur associé à l'IRIT dans le cadre du projet ANR

Voice4PD sur le diagnostic différentiel des Syndromes Parkinsoniens. Etienne SICARD est co-auteur des logiciels VOCALAB et DIADOLAB pour le bilan et la rééducation orthophonique de la voix et la parole. Il a co-écrit une vingtaine d'ouvrages et plus de 140 publications scientifiques notamment dans le domaine de l'analyse de la voix et parole pathologique. Il a été nommé en 2006 *Distinguished Lecturer* de la société savante IEEE.

Site : <https://cv.hal.science/etienne-sicard>

Email : [etienne.sicard@insa-toulouse.fr](mailto:etienne.sicard@insa-toulouse.fr)



**Anne MENIN-SICARD** est orthophoniste et titulaire d'un master recherche en Sciences du langage de l'université de Grenoble. Elle a exercé en libéral à Toulouse de 1994 à 2015. Anne MENIN-SICARD est co-auteur des logiciels VOCALAB et DIADOLAB.

Dans le cadre de AMS Formation, elle organise et anime des formations aux orthophonistes sur le bilan et la prise en charge de la voix et parole pathologique. Elle a fondé en 2019 Opop, pour la formation en ligne personnalisée à destination des orthophonistes. Chercheuse associée à l'ERU 46 du LURCO, elle développe des outils de mesures objectives et de rééducation de la parole en lien avec le logiciel DIADOLAB. Elle est auteur d'un ouvrage sur l'évaluation objective de la voix ainsi que de nombreuses publications dans le domaine de la voix et de la parole pathologique.

## Références

- [Antonetti 2020] Antonetti, A. E. D. S., (2020). Relationship of cepstral peak prominence-smoothed and long-term average spectrum with auditory-perceptual analysis. *Applied Sciences*, 10(23), 8598.
- [Audibert 2018] Audibert, N., & Falk, S. (2018, June). Vowel space and f0 characteristics of infant-directed singing and speech. In *Proceedings of the 19th International Conference on Speech Prosody* (pp. 153-157)
- [Auzou 2000] Auzou, P., (2000). Voice onset time in aphasia, apraxia of speech and dysarthria: a review. *Clinical linguistics & phonetics*, 14(2), 131-150.
- [Baken 2000] Baken, R. J., & Orlikoff, R. F. (2000). *Clinical measurement of speech and voice*. Cengage Learning.
- [Barbera 2016] Barbera, O., & Frenay, C. (2016). Acquisition d'une population adulte tout venant pour la caractérisation de l'évolution de la voix au cours du vieillissement par le logiciel VOCALAB: questionnement des limites entre pathologie et normalité. *Mémoire d'orthophonie*, Ecole d'orthophonie de Lyon.
- [Bardelang 2020] Bardelang T. (2020), L'apport d'une phrase supplémentaire au protocole de DIADOLAB3. *Mémoire d'orthophonie*. Université de Lyon. Encadré par A. Menin-Sicard.
- [Bonneh 2011] Bonneh, Y. S., (2011). Abnormal speech spectrum and increased pitch variability in young autistic children. *Frontiers in human neuroscience*, 4, 237.
- [Byrne 1994] Byrne, D., (1994). An international comparison of long-term average speech spectra. *The journal of the acoustical society of America*, 96(4), 2108-2120.
- [Carr 1964] Carr, P. B., & Trill, D. (1964). Long-Term Larynx-Excitation Spectra. *The Journal of the Acoustical Society of America*, 36(11), 2033-2040.
- [Cox 1988] Cox, R. M., & Moore, J. N. (1988). Composite speech spectrum for hearing aid gain prescriptions. *Journal of Speech, Language, and Hearing Research*, 31(1), 102-107
- [Da Silva 2011] Da Silva, P. T., (2011). Acoustic and long-term average spectrum measures to detect vocal aging in women. *Journal of voice*, 25(4), 411-419.
- [Doane 2011] Doane, D. P., & Seward, L. E. (2011). Measuring skewness: a forgotten statistic?. *Journal of statistics education*, 19(2).
- [Dromey 2003] Dromey, C. (2003). Spectral measures and perceptual ratings of hypokinetic dysarthria. *Journal of Medical Speech-Language Pathology*, 11, 85-94
- [ERU15 2018] Base de données de voix en ligne (100 fiches) sur <https://www.unadreo.org/lurco/les-erus-en-detail/eru-15-nouveaux-outils-danalyse-de-la-qualite-de-la-voix/eru-15-projet-base-de-donnees-de-voix/>
- [Folstein 1992] Folstein, M. F. (1992). A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr res*, 12, 189-198.

- [Forrest 1988] Forrest, K., (1988). Statistical analysis of word-initial voiceless obstruents: preliminary data. *The Journal of the Acoustical Society of America*, 84(1), 115-123.
- [Frokjaer -Jensen 1974] Frokjaer-Jensen, B., & Prytz, S. (1974). Evaluation of speech disorders by means of long-time-average-spectra. *Annual Report of the Institute of Phonetics University of Copenhagen*, 8, 227-237.
- [Frokjaer-Jensen 1975] Frokjaer-Jensen B., & Prytz, S. (1976). Registration of voice quality. *Annual Report of the Institute of Phonetics University of Copenhagen*, 9, 237-251.
- [Grenier 2020] Grenier, C. (2020). Etude comparative de tests phonologiques existants et du protocole d'analyse objective de la parole avec Diadolab: intérêt et limites d'un test court et des mesures objectives d'intelligibilité et de fluence. Mémoire de Master d'orthophonie. Université de Toulouse. Co-encadrement: S. Michel, E. Sicard.
- [Hammarberg 1986] Hammarberg, B., (1986). Acoustic and perceptual analysis of vocal dysfunction. *Journal of phonetics*, 14(3-4), 533-547.
- [Hanson 1999] Hanson, H. M., & Chuang, E. S. (1999). Glottal characteristics of male speakers: Acoustic correlates and comparison with female data. *The Journal of the Acoustical Society of America*, 106(2), 1064-1077.
- [Hartl 2003] Hartl, D. M., (2003). Objective acoustic and aerodynamic measures of breathiness in paralytic dysphonia. *European Archives of Oto-rhino-laryngology*, 260, 175-182.
- [Hirano 1981] Hirano M. (1981) *Clinical Examination of Voice*. New York: Springer-Verlag; 1981.
- [Houtz 2010] Houtz, D. R., (2010). Differential diagnosis of muscle tension dysphonia and adductor spasmodic dysphonia using spectral moments of the long-term average spectrum. *The laryngoscope*, 120(4), 749-757.
- [Kacha 2020] Kacha, A., (2020). Principal component analysis of the spectrogram of the speech signal: Interpretation and application to dysarthric speech. *Computer Speech & Language*, 59, 114-122.
- [Kent 2000] Kent, R. D., & Rosenbek, J. C. (1983). Acoustic patterns of apraxia of speech. *Journal of Speech, Language, and Hearing Research*, 26(2), 231-249.
- [Kent 2000] Kent, R. D., & Ball, M. J. (Eds.). (2000). *Voice quality measurement*. Singular.
- [Kitzing 1986] Kitzing, P. (1986). LTAS criteria pertinent to the measurement of voice quality. *Journal of phonetics*, 14(3-4), 477-482.
- [Klatt 1990] Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *The Journal of the Acoustical Society of America*, 87(2), 820-857.
- [Lee 1999] Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *The Journal of the Acoustical Society of America*, 105, 1455-1468.
- [Lee 2022] Lee, S. H., & Lee, G. S. (2022). Long-term Average Spectrum and Nasal Accelerometry in Sentences of Differing Nasality and Forward-Focused Vowel Productions Under Altered Auditory Feedback. *Journal of Voice*.

- [Leino 2009] Leino, T. (2009), "Long-Term Average Spectrum in Screening of Voice Quality in Speech", *Journal of Voice*, Volume 23, Issue 6, pp. 671–676
- [Linville 2001] Linville, S. E., & Rens, J. (2001). Vocal tract resonance analysis of aging voice using long-term average spectra. *Journal of Voice*, 15(3), 323-330
- [Lofqvist 1986] Lofqvist, A. (1986). The long-time-average spectrum as a tool in voice research. *Journal of phonetics*, 14(3-4), 471-475.
- [McCann 2003] McCann, J., & Peppé, S. (2003). Prosody in autism spectrum disorders: a critical review. *International Journal of Language & Communication Disorders*, 38(4), 325-350.
- [Maryn 2010] Maryn, Y., (2010). Toward improved ecological validity in the acoustic measurement of overall voice quality: combining continuous speech and sustained vowels. *Journal of voice*, 24(5), 540-555.
- [Martinez-Nicolas 2022] Martinez-Nicolas, I., (2022). Many changes in speech through aging are actually a consequence of cognitive changes. *International Journal of Environmental Research and Public Health*, 19(4), 2137.
- [Master 2006] Master, S., (2006). The long-term average spectrum in research and in the clinical practice of speech therapists. *Pró-Fono Revista de Atualização Científica*, 18, 111-120.
- [Mendoza 1996] Mendoza, E., (1996). Differences in voice quality between men and women: Use of the long-term average spectrum (LTAS). *Journal of voice*, 10(1), 59-66.
- [Menin-Sicard 2019] Menin-Sicard, A., & Sicard, E. (2019). Méthodologie d'évaluation objective de la phonologie, de la fluence et de la prosodie-Vers un bilan rapide à destination des orthophonistes. Journées de Phonétique Clinique JPC 2019. <https://hal.science/hal-02127039/>
- [Menin-Sicard 2016] Menin-Sicard, A., Sicard, E. (2016). Evaluation Et Rehabilitation De La Voix, Deboeck Solal, 2016, 978-2-35327-318-8. <https://hal.science/hal-02306658>
- [Menin-Sicard 2021] Menin-Sicard, A., (2021). Méthodologie de la segmentation pour l'évaluation objective des compétences phonologiques avec DIADOLAB3. <https://hal.science/hal-03149560/>
- [Menin-Sicard 2021b] Menin-Sicard, A., & Sicard, E. (2021). Etude de cas de voix dans le cadre de la prise en charge orthophonique. <https://hal.science/hal-03186341/>
- [Narasimhan 2017] Narasimhan, S. V., & Vishal, K. (2017). Spectral measures of hoarseness in persons with hyperfunctional voice disorder. *Journal of Voice*, 31(1), 57-61.
- [Papazian 2019] Papazian, A. (2019). Analyse acoustique de la production des consonnes et voyelles dans l'ataxie de Friedreich (Mémoire d'orthophonie Univ. Tours).
- [Peters 2017] Peters, J. (2017). Stability of long-term average spectrum measures and cepstral peak prominence in connected speech. *Proceedings P&P13*, 129-132
- [Remacle 2017] Remacle A., Genel Y., Segers M. (2017), Caractérisation de la voix d'enfants francophones euphoniques âgés de 5 ans, Congrès de la SFP & Journées Thérapeutiques de la voix, Nice, France.

- [Remacle 2020] Remacle, A., (2020). Vocal characteristics of 5-year-old children: proposed normative values based on a French-speaking population. *Logopedics Phoniatrics Vocology*, 45(1), 30-38.
- [Sergeant 2008] Sergeant, D., & Welch, G. F. (2008). Age-related changes in long-term average spectra of children's voices. *Journal of Voice*, 22(6), 658-670.
- [Sicard 2013] Sicard, E., Menin-Sicard, A. (2013). Implémentation dans VOCALAB d'indicateurs objectifs de la qualité de la voix dans le cadre de l'évaluation de la voix. *Rééducation orthophonique*, pp. 23-27. <https://hal.science/hal-00836912>
- [Sicard 2017] Sicard, E., Mauclair, J., Woisard, V. (2017). Etude de paramètres acoustiques des voix de patients traités pour un cancer ORL dans le cadre du projet C2SI, JPC 2017 Paris. <https://hal.science/hal-01510418>
- [Sicard 2020] Sicard, E., & Menin-Sicard, A. (2020). Le triangle vocalique et son application en contexte orthophonique. <https://hal.science/hal-02504513/>
- [Sicard 2020b] Sicard, E., & Menin-Sicard, A. (2020). La diadococinésie et son application en orthophonie clinique. Rapport de recherches LURCO, INSA toulouse. <https://hal.science/hal-02512856/>
- [Sicard 2021] Sicard, E., Menin-Sicard, A. (2021). Le spectrogramme et son application en clinique orthophonique. <https://hal.science/hal-03107434>
- [Sicard 2022] Sicard, E., Menin-Sicard, A., Rousteau, G. (2022) Oppositions de voyelles orales et nasales : identification des formants selon le genre. <https://hal.science/hal-03826558>
- [Sicard 2022c] Sicard, E., Menin-Sicard, A. (2022). Temps phonatoire, rapport s/z et a/z : repères en lien avec la prise en charge orthophonique. <https://hal.science/hal-03662213>
- [Sicard 2022] Sicard, E., Menin-Sicard, A., Michel, S., (2022). Evaluation et prise en charge orthophonique de la dyspraxie verbale avec DIADOLAB : approches existantes et études de cas. <https://hal.science/hal-03663252>
- [Sicard 2023] Sicard, E., Menin-Sicard, A. (2023). Fréquence fondamentale de la parole – Repères pour la prise en soin orthophonique. <https://hal.science/hal-04073008>
- [Sicard 2023b] Sicard, E., Menin-Sicard, A. (2023). DIADOLAB 3.5. Manuel d'utilisation. [www.diadolab.org](http://www.diadolab.org)
- [Sicard 2023c] Sicard, E., Menin-Sicard, A. (2023). VOCALAB 4.5. Manuel d'utilisation. [www.vocalab.org](http://www.vocalab.org)
- [Soumya 2022] Soumya, M., & Narasimhan, S. V. (2022). Correlation between subjective and objective parameters of voice in elderly male speakers. *Journal of Voice*, 36(6), 823-831.
- [Stajner-Katusic 2012] Stajner-Katusic, S., (2012). The voice of polypoid vocal folds before and after surgery. In *Investigations in Clinical Phonetics and Linguistics* (pp. 453-464). Psychology Press.
- [Svihlik 2022] Svihlik, J., (2022). Long-Term Averaged Spectrum Descriptors of Dysarthria in Patients With Parkinson's Disease Treated With Subthalamic Nucleus Deep Brain Stimulation. *Journal of Speech, Language, and Hearing Research*, 65(12), 4690-4699.

[Tanner 2005] Tanner, K., (2005). Spectral moments of the long-term average spectrum: sensitive indices of voice change after therapy?. *Journal of Voice*, 19(2), 211-222.

[Terband 2019] Terband, H., (2019). Assessment of childhood apraxia of speech: A review/tutorial of objective measurement techniques. *Journal of Speech, Language, and Hearing Research*, 62(8S), 2999-3032.

[Titze 2015] Titze, I. R., (2015). Toward a consensus on symbolic notation of harmonics, resonances, and formants in vocalization. *The Journal of the Acoustical Society of America*, 137(5), 3005-3007.

[Tjaden 2010]. Tjaden, K., (2010). Long-term average spectral (LTAS) measures of dysarthria and their relationship to perceived severity. *Journal of medical speech-language pathology*, 18(4), 125

[Yang 2018] Yang, Z., (2018). Long-term average spectra analysis of voice in children with cleft palate. *Journal of Voice*, 32(3), 285-290.

[Verdolini 2005] Verdolini, K., Rosen, C. A., & Branski, R. C. (Eds.). (2005). *Classification manual for voice disorders-I*. Psychology Press.

[Watts 2011] Watts, C. R., & Awan, S. N. (2011). Use of spectral/cepstral analyses for differentiating normal from hypofunctional voices in sustained vowel and continuous speech contexts. *Journal of Speech, Language, and Hearing Research*, 54(6), 1525-1538.

[White 1998] White, P. (1998). A study of the effects of vocal intensity variation on children's voices using long-term average spectrum (LTAS) analysis. *Logopedics phoniatrics vocology*, 23(3), 111-120.

[Yang 2018] Yang, Z., (2018). Long-term average spectra analysis of voice in children with cleft palate. *Journal of Voice*, 32(3), 285-290.