



Le spectrogramme et son application en clinique orthophonique

Etienne Sicard, Anne Menin-Sicard

► To cite this version:

Etienne Sicard, Anne Menin-Sicard. Le spectrogramme et son application en clinique orthophonique. 2021. [⟨hal-03107434⟩](#)

HAL Id: hal-03107434

<https://hal.science/hal-03107434v1>

Submitted on 12 Jan 2021

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.



HAL Authorization

Le spectrogramme et son application en clinique orthophonique

Etienne Sicard (1,2), Anne Menin-Sicard (2)

INSA-GEI, 135 Avenue de Rangueil, 31077 Toulouse (1)

Laboratoire LURCO www.lurco.unadreo.org (2)

Résumé :

Le spectrogramme et ses outils associés sont des assistants précieux, voire indispensables pour la prise en charge clinique des pathologies de la voix et de la parole. Nous décrivons dans cet article les principes de construction du spectrogramme, ses principales caractéristiques, et les données les plus significatives pouvant en être extraites. Nous détaillons ensuite quelques outils associés tels que l'allure temporelle, le profil de puissance les indicateurs d'altération et la segmentation. Quelques spectrogrammes de signaux fondamentaux tels que le son pur, le bruit, une sélection de voyelles et de consonnes sont proposés. Nous abordons enfin l'utilisation du spectrogramme en pratique clinique, pour le bilan ainsi que pour la rééducation de la voix et la parole.

Mots clé : orthophonie, évaluation de la parole, VOCALAB, DIADOLAB, PRAAT, Fondamental, FUM, altération de la voix, altération de la parole, pathologies de la voix, spectrogramme, temps réel, rééducation orthophonique, bilan orthophonique.

Introduction

Des efforts considérables ont été entrepris ces dernières années pour fournir aux orthophonistes des outils et des méthodes objectives basées des métriques éprouvées, autant pour permettre de réaliser des bilans de la voix et de la parole précis et factuels que pour rendre les rééducations ciblées, efficaces et motivantes pour le patient. Parmi les outils les plus utilisés se trouvent des logiciels tels que VOCALAB [Gerip 2020], DIADOLAB [Menin Sicard 2019] et PRAAT [Boersma 2018]. Une caractéristique commune de ces outils est de proposer une représentation de la production sonore sous forme de spectrogramme, noyau fondamental d'un véritable écosystème d'analyse de la voix et de la parole.

Le spectrogramme est une représentation permettant de voir l'ensemble de la décomposition spectrale de la voix et de la parole sur une même représentation graphique. Cet outil se révèle précis, informatif et fiable pour analyser les caractéristiques de la production sonore. L'idée du spectrogramme est très ancienne. W. Koenig [Koenig 1947] montre en temps réel la distribution de l'énergie en fréquence et en temps de différents échantillons de sons sur un « spectrographe sonore ». L'ouvrage *Speech Sciences* [Kent 1997] consacre un chapitre *Acoustic phonetics* illustré de nombreux exemples de spectrogramme. De son côté, [Baken 2000] consacre le chapitre *Sound Spectrography* au spectrogramme, en détaillant la théorie associée et le paramétrage des calculs et de la représentation graphique. Il s'agit à notre connaissance d'une des références les plus détaillées sur le sujet. L'ouvrage [Cornut 2001] en langue française donne de nombreuses illustrations à base de spectrogramme en lien avec la voix chantée. La construction et le paramétrage du spectrogramme est aussi abordé par [Ghio 2007].

Le spectrogramme a beaucoup intéressé la communauté scientifique ces dernières années comme un champ d'application possible de la reconnaissance des formes, en vue de l'aide au diagnostic. L'analyse des contrastes, profils, changements de formes dans les spectrogrammes a permis à [Sharma 2020] de séparer 50 enfants normophoniques et 50 enfants avec troubles spécifiques du langage, avec une fiabilité de détection supérieure à une dizaine de travaux antérieurs. De manière similaire, [Zahid 2020] a pu exploiter les images des spectrogrammes dans le cadre de la maladie de Parkinson pour séparer 150 témoins de 150 patients, avec une fiabilité élevée, sur la base de voyelles, de mots et de monologues. Différentes approches d'analyse d'image basées sur l'analyse des couleurs, du contraste, des ressemblances [Haralick 1973] associées à des méthodes de classification élaborées [Smola 2004] exploitent au mieux les indices de pathologie incluses dans les images de spectrogramme, ce qui prouve bien la pertinence de celui-ci pour l'analyse des pathologies de la voix et de la parole.

Il existe relativement peu de publications en langue Française qui détaillent les caractéristiques et l'utilisation du spectrogramme dans le contexte orthophonique, à l'exception notable de [Cornut 2001] du numéro dédié à *L'évaluation Vocale* [Menin-Sicard 2013], de l'article de [Ghio 2007] ou encore du mémoire d'orthophonie de [Parmentier 2015]. Pourtant, que ce soit lors du bilan ou bien tout au long des séances de rééducation orthophonique, le spectrogramme et ses modules satellites peuvent être considérés comme des assistants précieux, voire indispensables pour la prise en charge clinique des pathologies de la voix et de la parole. C'est pourquoi nous avons souhaité consacrer quelques pages et illustrations au spectrogramme et ses outils associés, que l'on peut considérer comme les éléments d'une grande famille, comme illustré à la figure 1, dans une représentation inspirée de la classification des éléments simples.

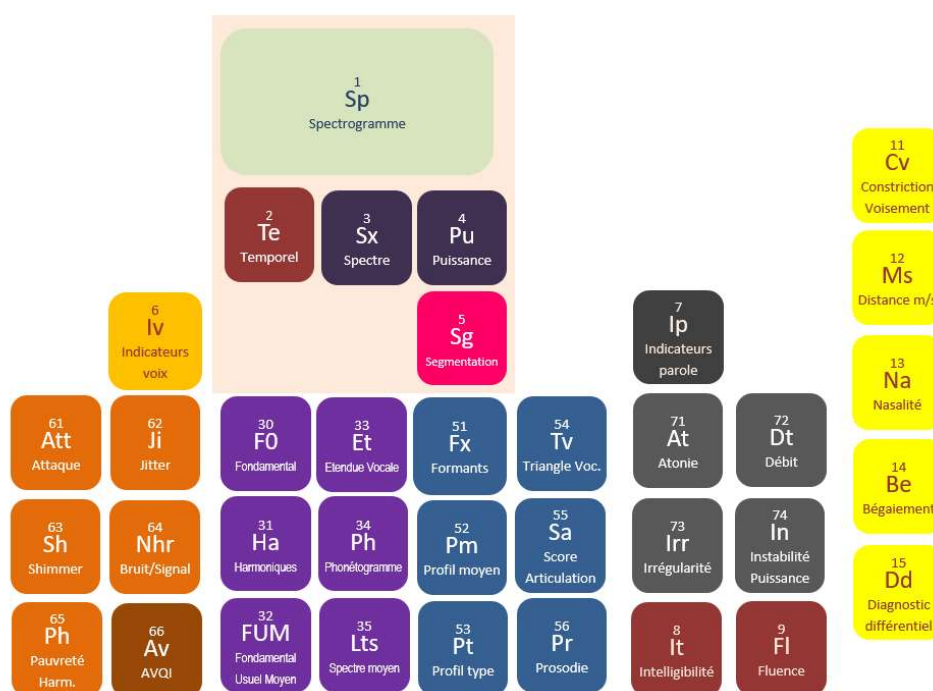


Figure 1 : Le spectrogramme, ses outils associés et métriques communes, dans une représentation inspirée de la classification des éléments

Dans un premier cercle proche, on associe au spectrogramme l'allure temporelle, le profil de puissance et la segmentation. Dans un cercle un peu plus étendu se trouvent un nombre conséquent d'indicateurs, de métriques et d'outils, parmi lesquels se trouve la fréquence fondamentale et ses dérivés, l'altération de la voix et de la parole, et de manière plus globale l'évaluation de l'intelligibilité.

Du spectrogramme dérivent aussi des analyses plus ciblées, pouvant concerner des altérations spécifiques de la parole : nasalité, constriction, voisement, voire aide au diagnostic ou à la prise en charge de certaines pathologies.

Dans cet article, nous donnons quelques rappels sur la notion de passage temps/fréquence, de construction d'un spectre instantané puis d'un spectrogramme. Nous illustrons aussi les grandeurs associées, avant de présenter les différents outils complémentaires du spectrogramme, allant de la détection de la fréquence fondamentale à la segmentation puis la construction de profils type. Enfin, dans une partie consacrée aux applications du spectrogramme en clinique orthophonique, nous abordons les aspects bilan et rééducation de la voix et de la parole, en puisant dans les protocoles et axes thérapeutiques usuels des illustrations visant à démontrer l'intérêt de cet outil.

Un peu de théorie



Le spectrogramme est une représentation que l'on qualifie parfois de « langue étrangère » au premier abord, car son graphisme nécessite un effort d'appropriation. Il s'agit en effet d'une représentation à trois dimensions (Figure 2 en haut) :

- En axe horizontal, le temps
- En axe vertical, la fréquence
- Via un jeu de couleurs, l'amplitude

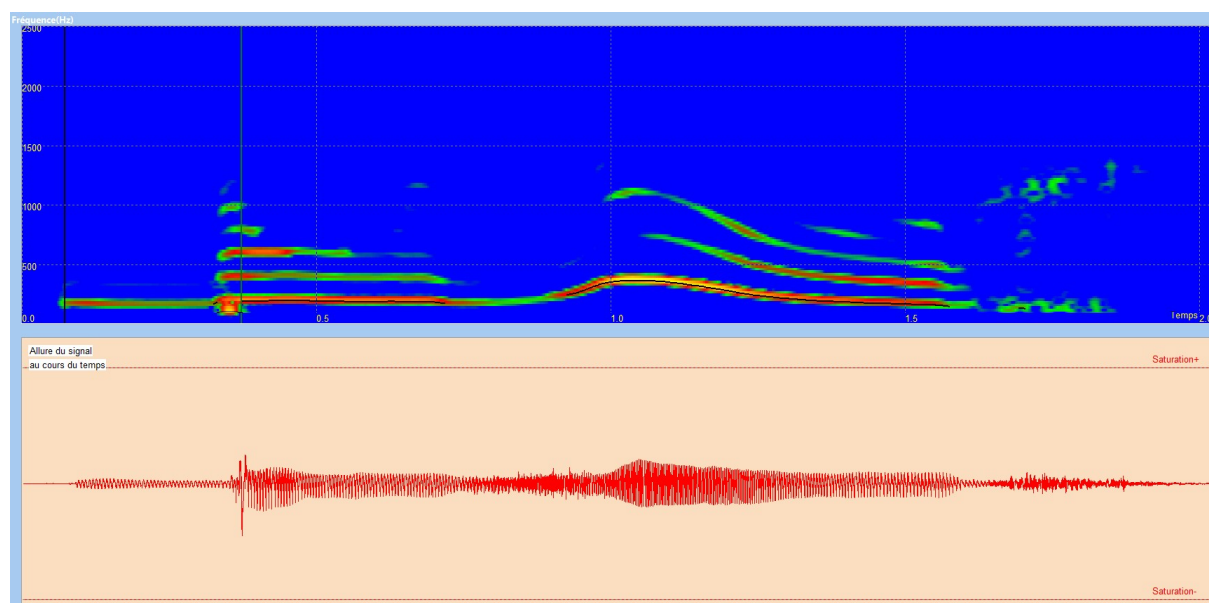


Figure 2 : /Bonjour/ : spectrogramme (en haut) et allure temporelle (en bas), femme de 46 ans, parole normale avec VOCALAB [Sicard 2013]

Allure temporelle du signal sonore



En parlant, nous faisons varier la pression acoustique de l'air. Cette variation est capturée par un microphone, se traduisant alors par une variation de tension, qui est ensuite numérisée et se traduit par une série de valeurs numériques qui correspondent au tracé de la figure 2 en bas. De manière native, nous ne disposons donc pas directement du contenu fréquentiel de la production sonore, mais seulement de son allure temporelle.

Passage temporel/fréquentiel



Le dessin du spectrogramme repose sur le calcul itératif des composantes fréquentielles de portions de son, encore appelé « spectre » grâce à une opération appelée Transformée de Fourier [Sicard 2016], et ce jusqu'à couvrir l'ensemble de la production sonore. Environ 1000 conversions temps/fréquence ont été nécessaires pour calculer le spectrogramme de la figure 2. Chaque conversion porte sur une petite portion du signal sonore, que l'on décale progressivement dans le temps. Le résultat de chaque conversion est un tableau de valeur d'énergie fonction de la fréquence, que l'on reporte sous forme d'une barrette verticale de couleurs (Figure 3). L'énergie importante est associée à une couleur chaude (jaune-blanc), une faible énergie à une couleur froide (vert-bleu).

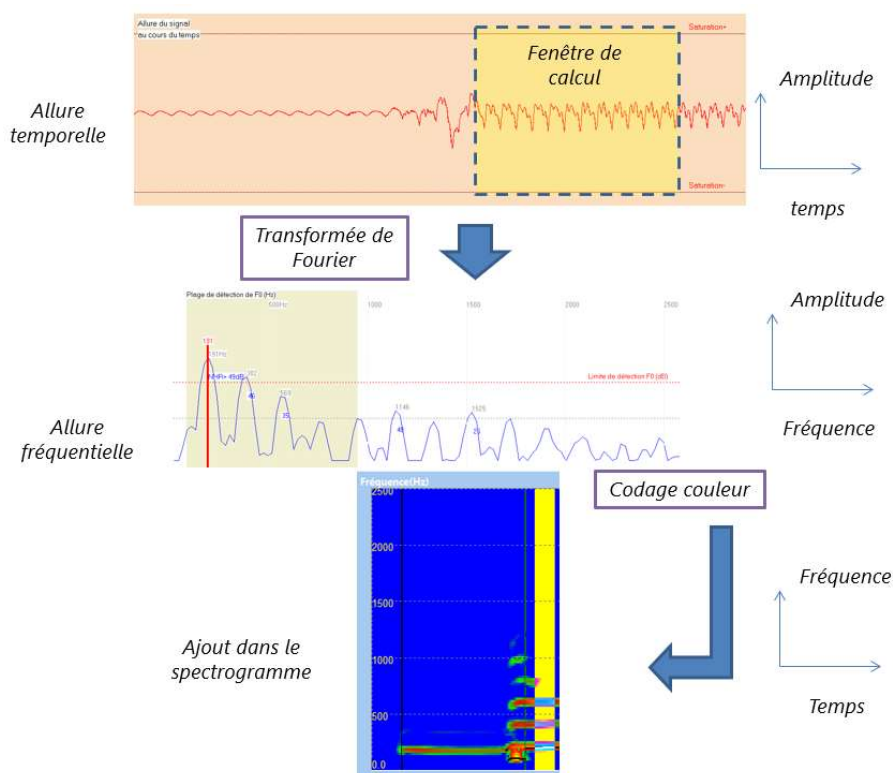


Figure 3 : Principe de construction du spectrogramme [Sicard 2017] du mot /Bonjour/

Il est possible d'associer d'autres palettes de couleurs, telles que les niveaux de gris dans PRAAT [Boersma 2018]. Dans ce cas, une énergie faible correspond à du blanc, une énergie forte correspond à du noir (Figure 4). Parmi les logiciels en libre accès qui disposent de manière native du spectrogramme, on peut aussi citer AUDACITY [Audacity 2018] (Figure 4). L'énergie faible est en blanc, l'énergie importante en bleu.

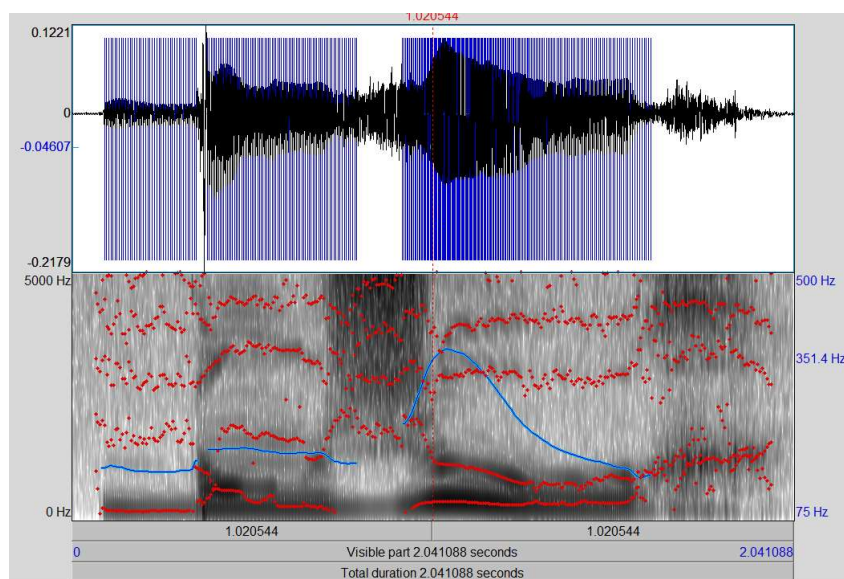


Figure 4 : /Bonjour/ : allure temporelle et spectrogramme, femme de 46 ans, parole normale avec PRAAT

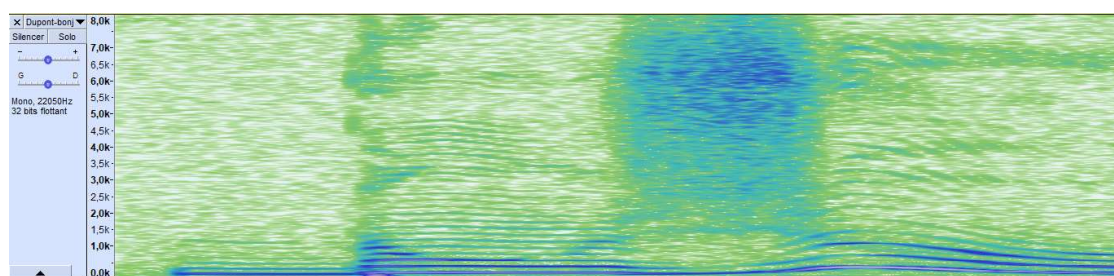


Figure 5 : spectrogramme de /Bonjour/, femme de 46 ans, parole normale avec AUDACITY [Audacity 2018]

Fondamental F0



Dans la portion de son correspondant au /on/ de /bonjour/, on constate une périodicité du son, c'est-à-dire une forme qui se répète de façon régulière au cours du temps, même si de légères variations sont observées entre chaque cycle successif. Cette période vaut ici environ 5 ms, calculée entre deux pics du signal, dans le sens horizontal qui représente le temps. En inversant cette valeur, on trouve la fréquence correspondante notée F0, égale ici à 200 Hz environ. Il s'agit de la fréquence fondamentale, tout à fait dans la norme d'une voix de femme en parole spontanée. La hauteur est reportée dans le spectrogramme dans l'axe vertical. Il est donc normal de voir apparaître un trait de couleur chaude (donc d'énergie élevée) à environ 200 Hz.

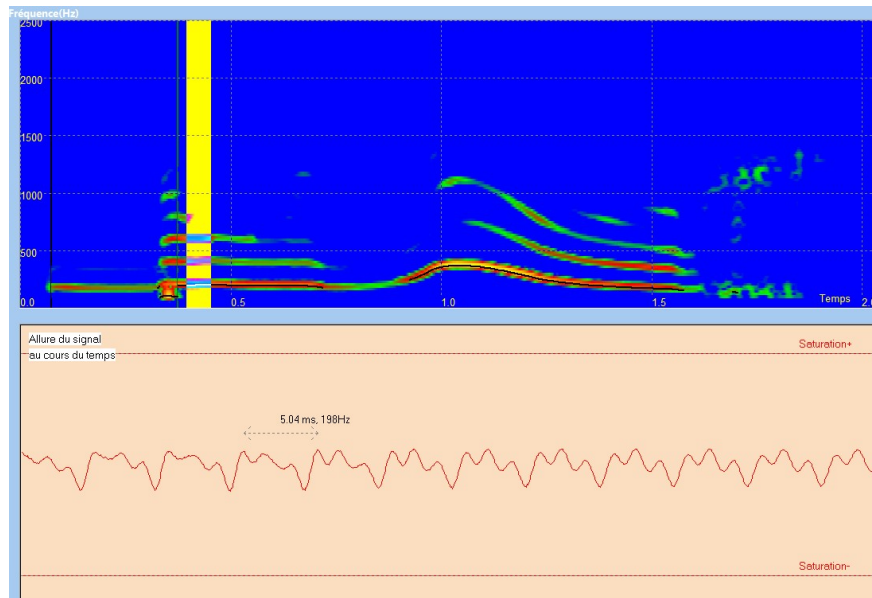


Figure 6 : Allure temporelle d'une portion du phonème /on/ de /Bonjour/ et illustration de la période fondamentale

La fréquence fondamentale existe dans tous les sons voisés, nous l'appellerons F0 ou encore fréquence fondamentale de la voix et de la parole. Elle correspond au trait noir superposé dans le spectrogramme dans VOCALAB et DIADOLAB (Figure 9 en haut). Dans PRAAT, elle apparaît en bleu, avec une échelle spécifique reportée à droite du spectrogramme (Figure 10 en bas). La cumulation des valeurs de F0 permet d'extraire le fondamental usuel moyen (FUM). L'extraction de F0 sur une sirène permet de déterminer l'étendue vocale. Enfin, la cumulation énergie fonction de F0 permet de dessiner le phonétogramme.

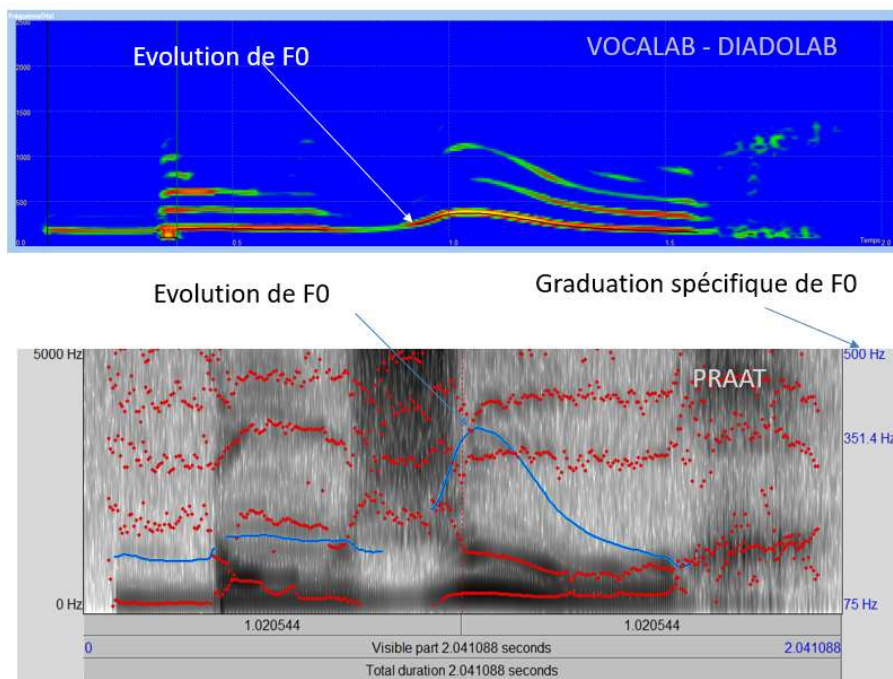


Figure 10 : Détection de la fréquence fondamentale dans PRAAT



On peut noter sur le spectrogramme de la figure 6 des traits supplémentaires situés à $2 \times F_0$ et $3 \times F_0$. Ce sont les harmoniques, qui sont des composantes multiples de la fréquence fondamentale. On peut noter que lors de la production de la consonne initiale /b/, seule F_0 apparaît distinctement. Sur la voyelle /on/, on note 3 harmoniques bien visibles. L'origine de ces harmoniques, de leur apparition ou de leur atténuation est décrite abondamment dans la littérature sous le nom de théorie source-filtre [Baken 2000]. La contraction brusque des cordes vocales au rythme de la fréquence fondamentale F_0 ne génère pas qu'une seule fréquence, mais se traduit aussi par un grand nombre de multiples de F_0 ($2 \times F_0$, $3 \times F_0$, etc.). Si l'on assimile la contraction à un phénomène périodique triangulaire, la décomposition harmonique est infinie, bien que l'atténuation des composantes soit assez rapide. Le conduit vocal et la position des articulateurs sert de filtre qui déforme ces harmoniques en atténuant ou exagérant certaines zones. Autrement dit, l'énergie des certaines harmoniques peut être considérablement atténuée par rapport à d'autres selon la forme du conduit vocal et le couplage des cavités entre elles.

La notion énergie (ou d'intensité) correspond dans le spectrogramme à une palette de couleurs. Ainsi, dans VOCALAB et DIADOLAB, le bleu-vert correspond à une énergie faible, le jaune-blanc à une énergie importante.

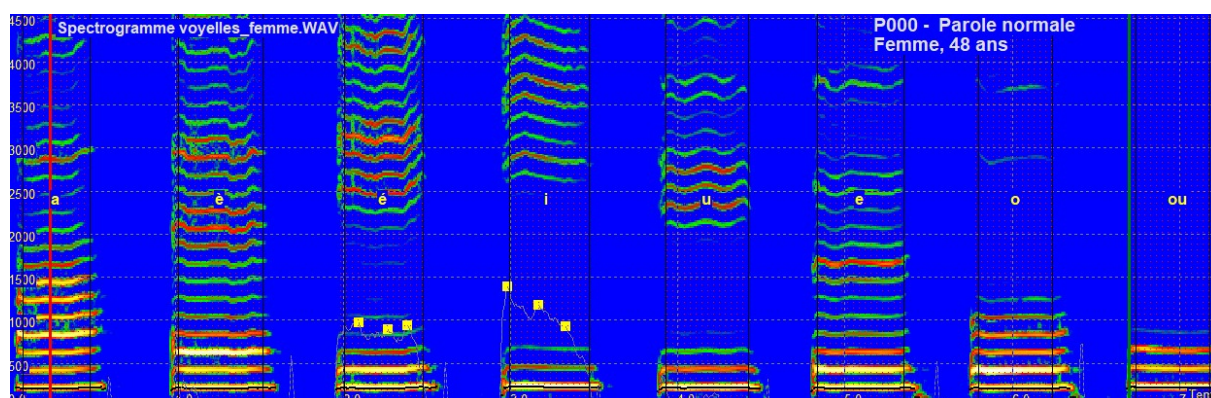
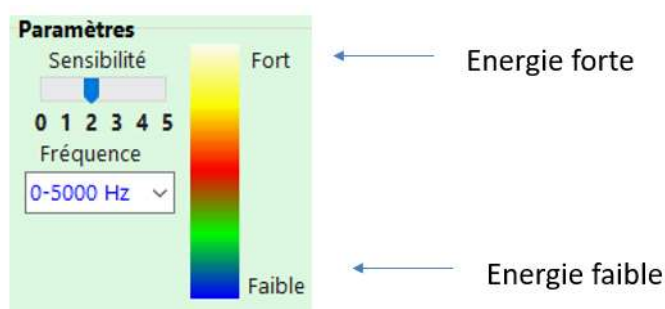


Figure 7 : Allure temporelle de voyelles isolées, mettant en évidence des structures harmoniques différentes et des énergies réparties différemment selon les fréquences

L'énergie relative des harmoniques par rapport à F0 est très dépendante du phonème qui est prononcé. Le spectrogramme de la figure 7 concernant des voyelles isolées /a, è, é, i, u, e, au, ou/ prononcées par la même personne que pour la figure 5. On constate de grandes différences de couleur et de position des harmoniques. Le /a/ et le /è/ sont les voyelles parmi les plus riches en termes d'harmoniques, autrement dit les plus énergétiques.

Paramètres du spectrogramme

D'un point de vue spectral, l'échantillonnage à une fréquence F_e impose une limitation intrinsèque dans le calcul du spectre F_{max} , qui ne peut s'étendre qu'à la moitié. Cette règle est connue sous le nom de théorème de Shannon. Si l'on choisit $F_e=22$ KHz comme fréquence d'échantillonnage, le spectre correspondant est limité à $F_{max}=11$ KHz. Avec $F_e=44$ KHz (CD audio), F_{max} monte à 22 KHz. On peut considérer que $F_e=22$ KHz est une limite acceptable pour la mesure des voix pathologiques, la limite 11 KHz englobant une très large part des composantes de la voix et de la parole. C'est pourquoi les logiciels VOCALAB et DIADOLAB sont configurés par défaut à $F_e=22$ KHz.

Certains auteurs distinguent les notions de bande étroite et de bande large [Baken 2000][Ghio 2007]. Nous utilisons par défaut dans VOCALAB et DIADOLAB la bande étroite, permettant de bien distinguer les harmoniques et évaluer le bruit synonyme de souffle, ou les sous-harmoniques synonymes d'éraillure, tandis que PRAAT utilise par défaut la bande large, plus ciblée sur les formants. Nous utilisons une échelle de représentation de l'énergie en décibel (échelle logarithmique et non linéaire) qui est plus proche de la perception humaine.

Les outils associés au spectrogramme

Allure Temporelle



Le compagnon presque indispensable au spectrogramme est l'allure temporelle. On ne s'étonnera pas que PRAAT propose une représentation par défaut avec l'allure temporelle dans la partie supérieure, le spectrogramme dans la partie inférieure. Dans VOCALAB, C'est l'onglet « Temporel » qui permet de superposer le spectrogramme et l'allure temporelle. Dans DIADOLAB, l'outil phonologie inclut aussi une fenêtre temporelle qui apparaît en bas à gauche, comme illustré figure 8. La correspondance temps/fréquence se révèle très pertinente en formation initiale et continue, lorsque l'on analyse les sons fondamentaux, mais il est aussi un compagnon très précieux pour l'analyse des phénomènes transitoires telles que les consonnes plosive ou encore le phénomène de contamination.

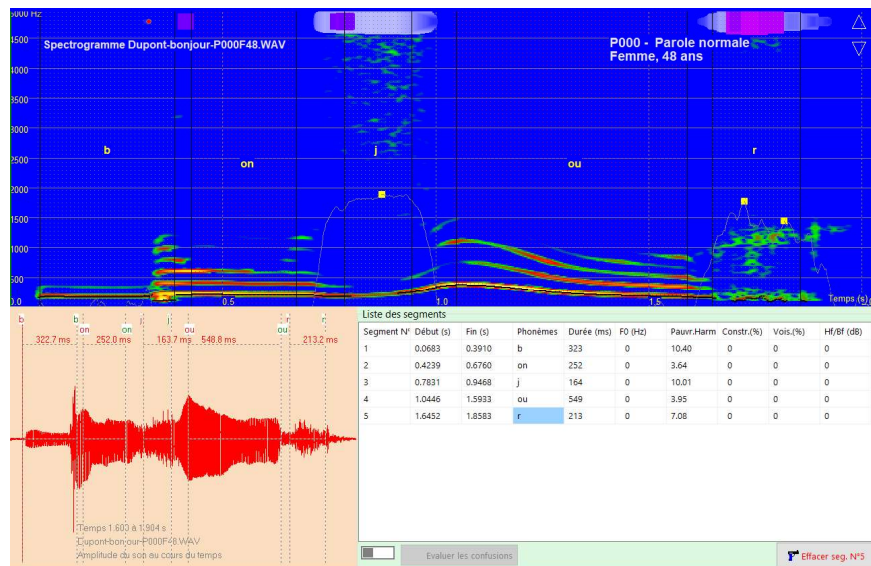


Figure 8 : Association du spectrogramme et de l'allure temporelle pour identifier les frontières des phonèmes et leurs caractéristiques transitoires, mot /bonjour/ prononcé par une femme de 48 ans (DIADOLAB)

L'outil temporel permet par exemple d'ajouter très précisément les frontières, alors que la segmentation dans le spectrogramme ne permet qu'un découpage assez grossier (Figure 9).

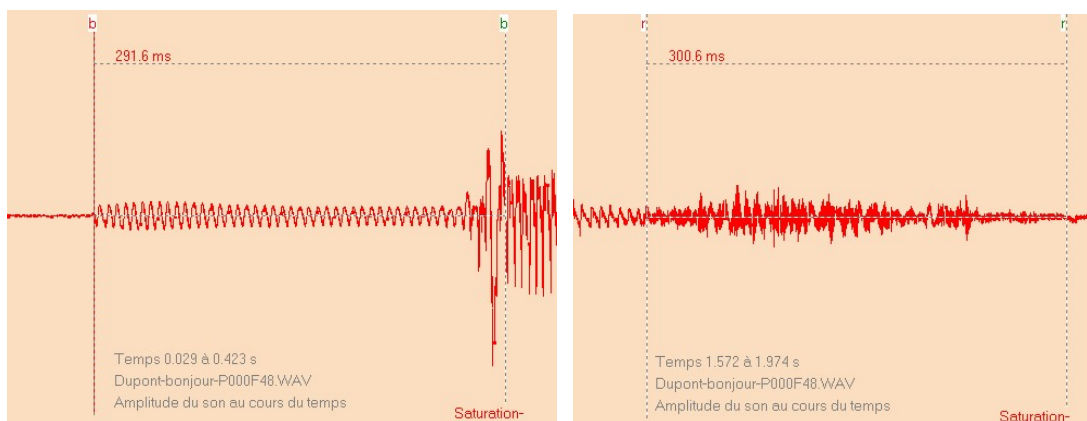
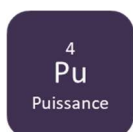


Figure 9 : Ajustement des frontières du /b/ et du /r/ dans /bonjour/ grâce à la fenêtre temporelle

Profil de puissance



Le profil de puissance est aussi une source d'information complémentaire au spectrogramme, même s'il s'agit d'une information un peu redondante, mais présentée différemment. Dans VOCALAB, l'onglet « Puissance » permet de superposer la puissance du signal sonore, qui n'est autre que la cumulation des énergies réparties dans les différentes harmoniques, en séparant les énergies en dessous de 1300 Hz et celles au-dessus de 1300 Hz. On voit bien apparaître le phénomène de constriction dans le /j/ de /bonjour/, ainsi que le bruit lors de la prononciation du /r/. On peut aussi identifier le moment précis de l'explosion du /b/ qui correspond à un pic de puissance au temps 0.3 s (Figure 11).

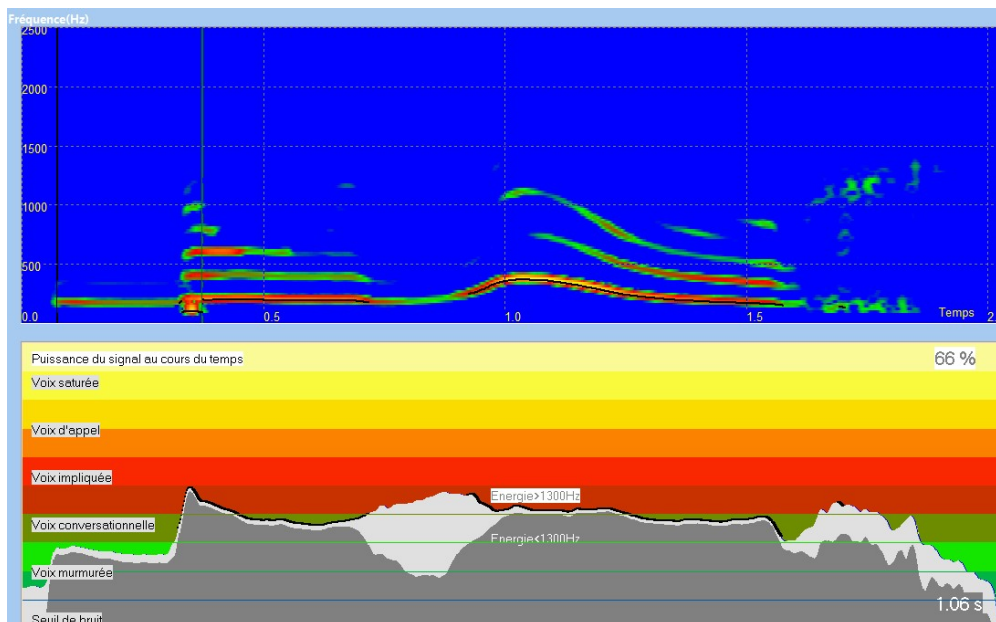


Figure 11 : Association du profil de puissance au spectrogramme, permettant de mettre en valeur les phénomènes de constriction et de bruit

Indicateurs d'altération de la voix



Il est difficile, même pour une oreille experte, d'évaluer séparément, de manière fiable et objectif les différents paramètres d'altération de la voix après écoute, ou par analyse du spectrogramme. Par exemple, si l'on prend le cas de la figure 11 correspondant au /a:/ tenu, la voix est-elle instable, pauvre, bruitée ? Si oui, de combien ? A quelle distance de la limite normal/altéré ?

Pour répondre à ces questions, de nombreux scientifiques ont développé des indicateurs, les plus célèbres d'entre eux étant sans doute le *Jitter*, *Shimmer* et *HNR* (*Harmonic Noise Ratio*). On peut aussi citer le AVQI qui est particulièrement utilisé en Belgique [Barsties 2018]. Les métriques de base se retrouvent dans des outils tels que PRAAT, EVA2 et MDVP pour ne citer que les plus connus [Sicard 2015]. VOCALAB et DIADOLAB proposent des formulations similaires dénommées instabilité en hauteur, en intensité, et rapport bruit/signal, associés à d'autres indicateurs tels que l'altération de l'attaque et la pauvreté harmonique (Figure 12).

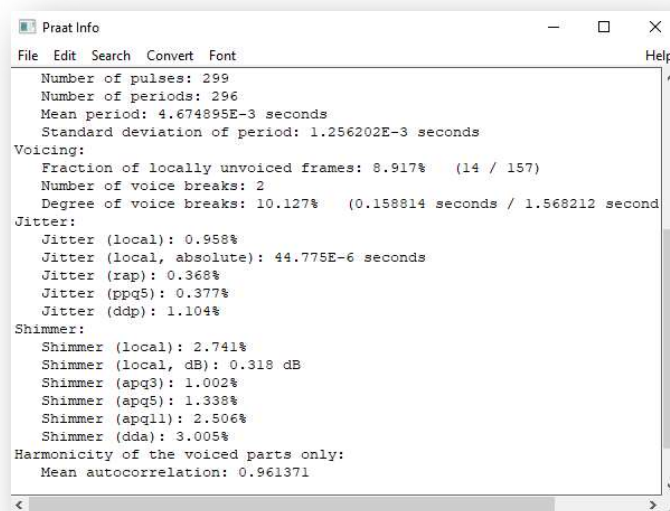
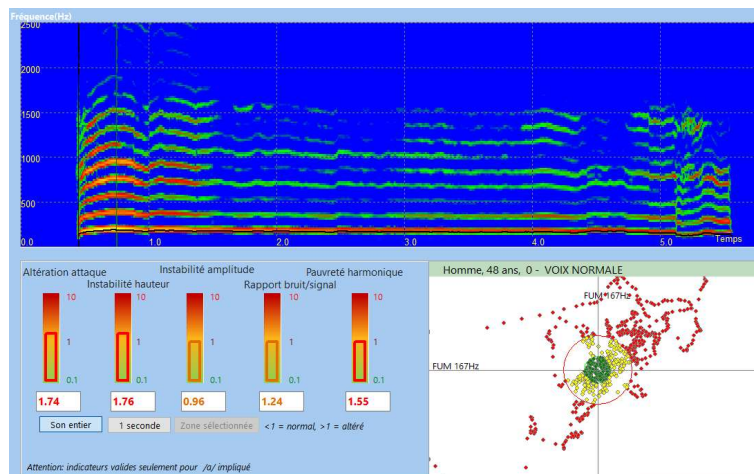
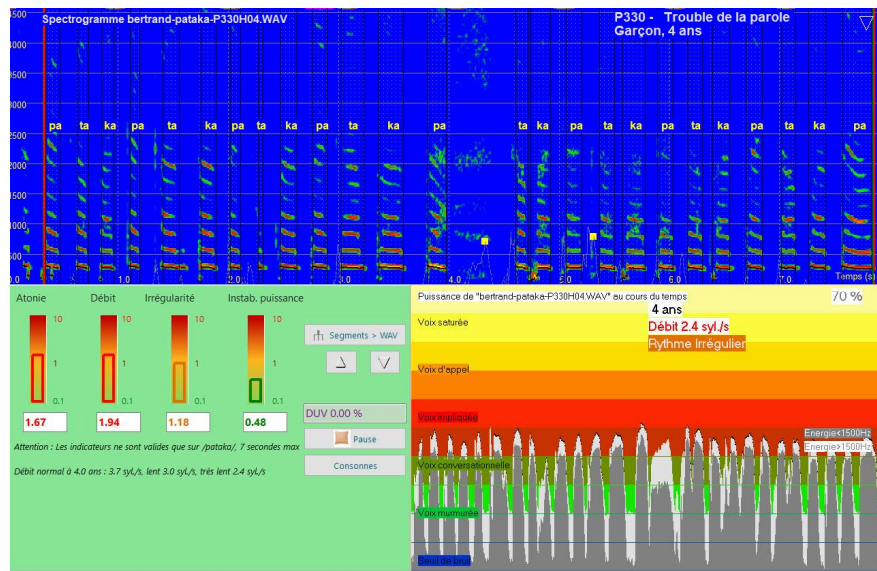


Figure 12 : spectrogramme d'un /a:/ tenu altéré et indicateurs d'altération de la voix de VOCALAB. Les mesures d'instabilité peuvent être rapprochées du Jitter et Shimmer calculés par PRAAT, même si les algorithmes de calculs et seuils normal/altéré sont différents

Indicateurs d'altération de la parole



De même que pour l'analyse objective du /a:/ présenté précédemment, une question similaire se pose lors de l'évaluation d'une épreuve très commune en bilan de la parole : l'épreuve de diadococinésie [Sicard 2020]. Comment évaluer de manière fiable et objectif les différents paramètres d'altération de la parole et du contrôle phono-articulatoire ? Par exemple, si l'on prend le cas de la figure 13 correspondant à une série de /Pataka/, le débit est-il instable, irrégulier, l'articulation atone ? Si oui, de combien ? A quelle distance de la limite normal/altéré ?



Pour répondre à ces nouvelles questions, nous avons développé dans DIADOLAB des indicateurs d'altération de la diadococinésie, dénommés *atonie*, *débit*, *irrégularité* et *instabilité en puissance*. Ces indicateurs se sont révélés particulièrement pertinents pour les cas de retards de parole, de dyspraxie, de bredouillement, ou encore de dysarthries [Sicard 2020].

Segmentation du spectrogramme



L'analyse de la parole se passe difficilement d'une étape de segmentation de la production sonore, afin de préciser les frontières de chaque phonème pour permettre de faire porter ensuite des analyses ciblées telles que l'extraction des formants, des profils moyens, de la constriction, etc. Un exemple de segmentation de la phrase /Bonjour Monsieur Tralipau/ [Sicard 2020] est reporté figure 14. La segmentation a été facilitée d'une part par les changements d'allure temporelle et fréquentielle, ainsi que par la fourniture d'indices telles que la constriction (taches blanches) et la richesse harmonique (taches rouges).

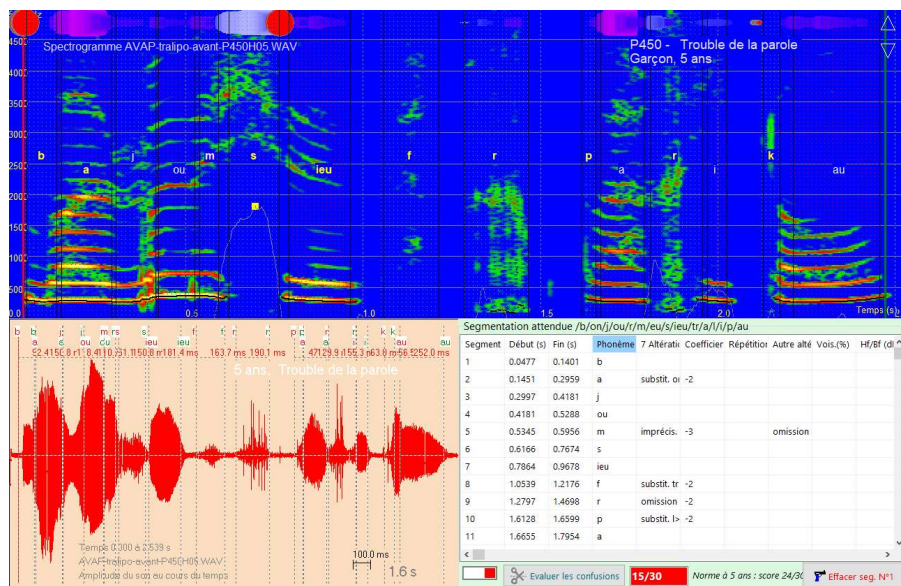


Figure 14 : spectrogramme de la phrase /Bonjour Monsieur Tralipau/, segmentation et calcul du score d'articulation avec DIADOLAB

Profil moyen d'un phonème et extraction des formants



Le profil moyen d'un phonème correspond à la moyenne du spectrogramme dans l'intervalle donné. Ce profil permet d'analyser les caractéristiques du phonème par comparaison avec un modèle de référence basé sur une analyse statistique, ou de comparer deux phonèmes entre eux, comme par exemple le /m/ et le /s/ [Sicard 2016] ou les oppositions orales/nasales telles que /è-in/ de la figure 15 [Sicard 2021]. Il permet aussi de clairement distinguer les formants et antiformants, par exemple ici le formant F2 du /in/ autour de 2000 Hz, et F2 du /è/ autour de 2600 Hz.

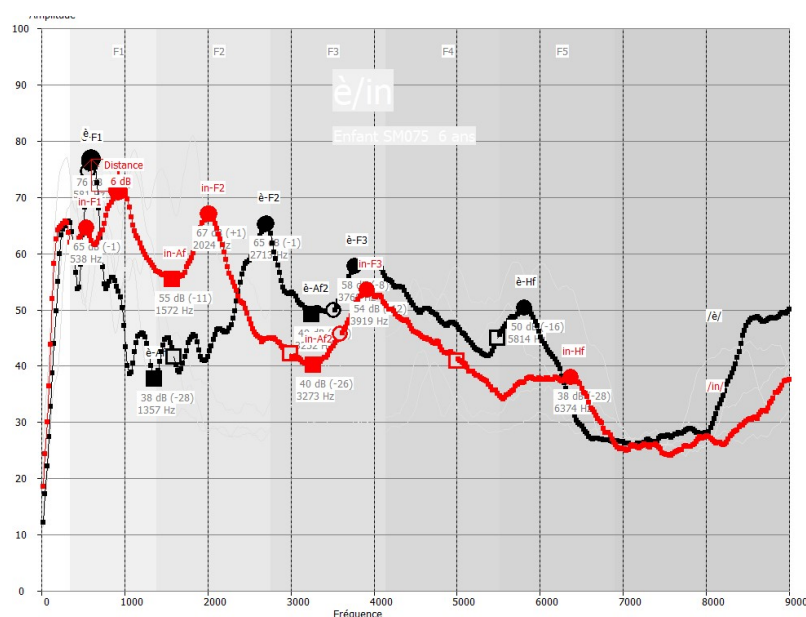
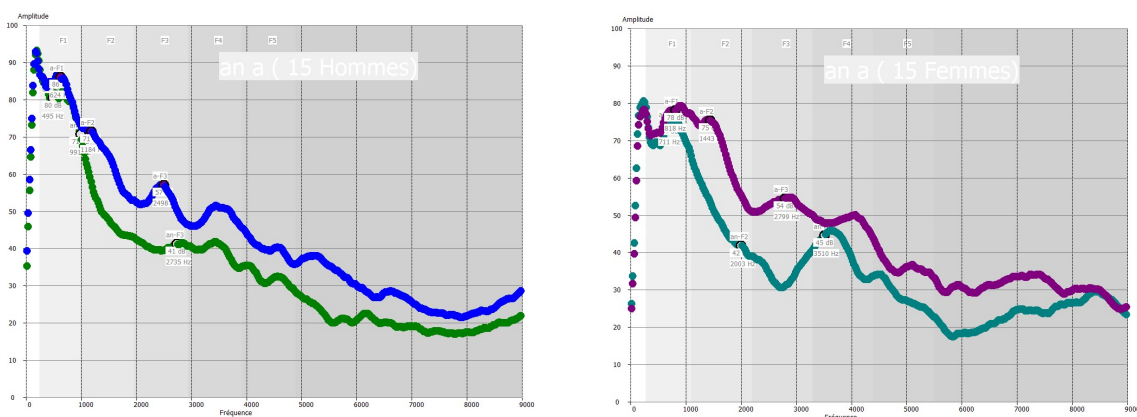


Figure 15 : Profils comparés du /è/ (courbe noire) et du /in/ (courbe rouge) d'un enfant normo phonique

Profil type d'un phonème

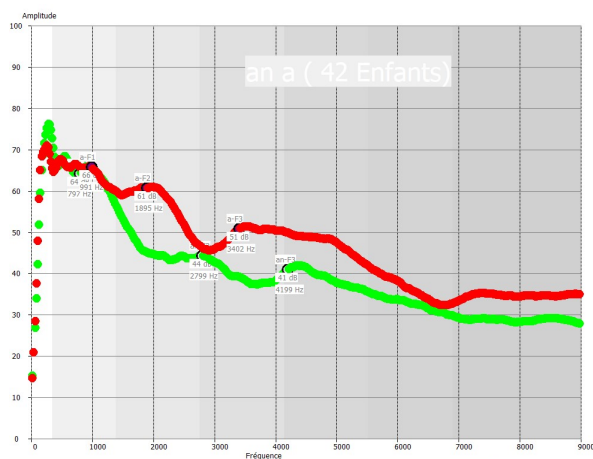
43
Pt
Profil type

Par analyse d'un grand nombre de phonèmes produits par plusieurs dizaines de personnes, il devient possible avec Diadolab de construire le profil type des phonèmes afin de comparer leurs allures, comme le /a-an/ pour hommes, femmes et enfants dans le cadre de l'imprécision orale/nasale [Sicard 2021] (Figure 15). L'approche statistique permet de séparer ce qui est récurrent de ce qui est propre à un individu donné. Ici, on distingue assez clairement la place des formants, les antiformants et la position respective des formants dans les oppositions.



a) Hommes /a/ (courbe bleue) /an/ (courbe verte)

b) Femmes /a/ (courbe violette) /an/ (courbe verte)



c) Enfants /a/ (courbe rouge) /an/ (courbe verte)

Figure 16 : Profils comparés du /a/ et du /an/, femmes, hommes & enfants, parole normale

Triangle vocalique

44
Tv
Triangle Voc.

Toujours en lien avec la segmentation, il est possible de positionner les voyelles, voire les consonnes dans un espace à deux dimensions, tel que F1/F2 ou F2/F3 (Figure 17). Dans la représentation F1/F2,

les réalisations des phonèmes /a,i,ou/ se situent en général aux extrêmes de ce qui est communément appelé le triangle vocalique [Sicard 2020]. La représentation permet de situer les réalisations du patient par rapport à des normes. Dans le cas de la figure 16, le triangle est un peu plus resserré que la norme, mais le patient ne fait aucune erreur d'articulation.

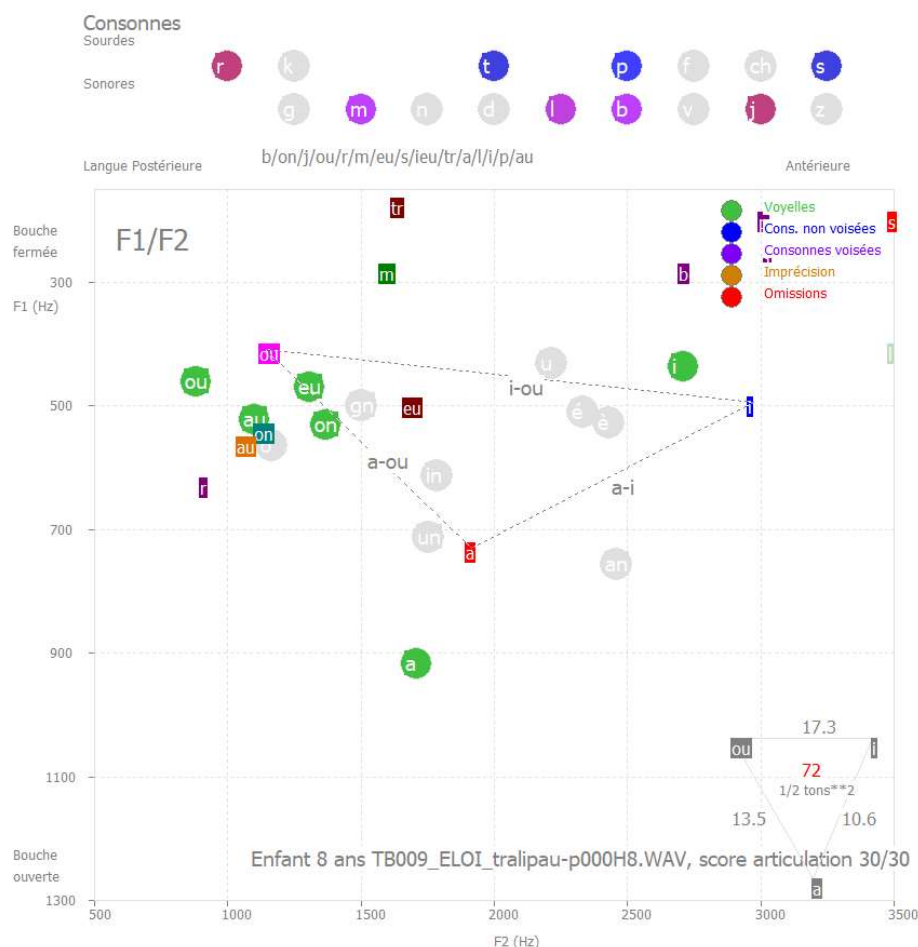


Figure 17 : Positionnement des voyelles et des consonnes dans l'espace des formants F1/F2.

Spectre moyen long terme



Le spectre moyen long terme (*Long Term Average Spectrum, LTAS*) est en quelque sorte la moyenne du spectrogramme, ou d'un ensemble de spectrogrammes portant sur quelques secondes ou dizaines de secondes de parole, permettant d'inclure une large variété de combinaisons de phonèmes. Le LTAS peut aussi combiner un groupe de personnes ayant une caractéristique commune, par exemple des patients atteints d'une certaine pathologie. Le LTAS a été utilisé pour l'identification et la reconnaissance des locuteurs, l'évaluation de la qualité de la voix ou encore la détection de la dysarthrie [Tjaden 2010] [Kacha 2020]. Un exemple de comparaison de spectre long terme de témoins (Figure 18) et de patients ayant été atteints de cancer ORL est donné figure 18 [Sicard 2017].

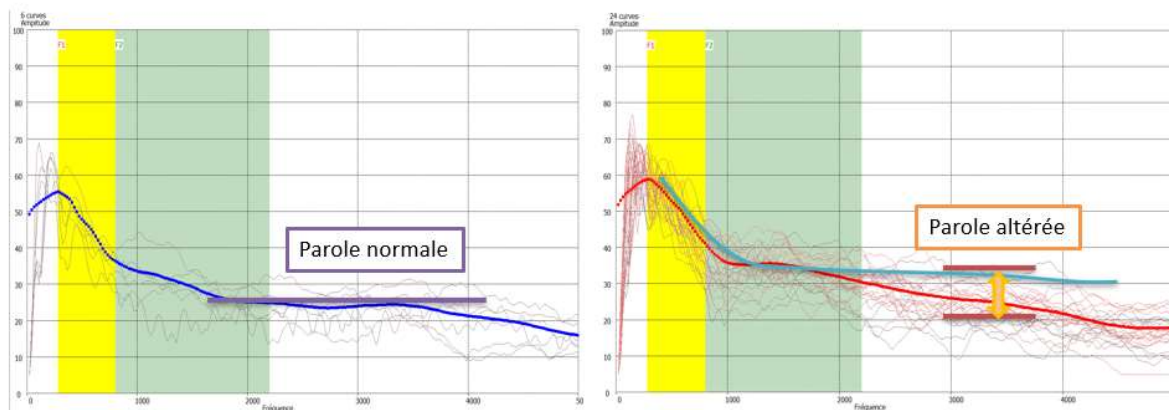


Figure 18 : Comparaison de spectres moyens long terme en comparant la parole de témoins et celle de patients ayant été atteints de cancer ORL, projet C2SI [Sicard 2017]

Spectrogrammes fondamentaux

Spectrogramme d'un son pur

Un son pur équivaut à un signal sonore d'une fréquence unique. En temporel, ce signal correspond à une simple sinusoïde de période $1/F_0$, et à trait horizontal dans le spectrogramme de fréquence F_0 (ici 440 Hz). La médiathèque de VOCALAB donne l'accès à cet échantillon par le bouton « Son Pur ».

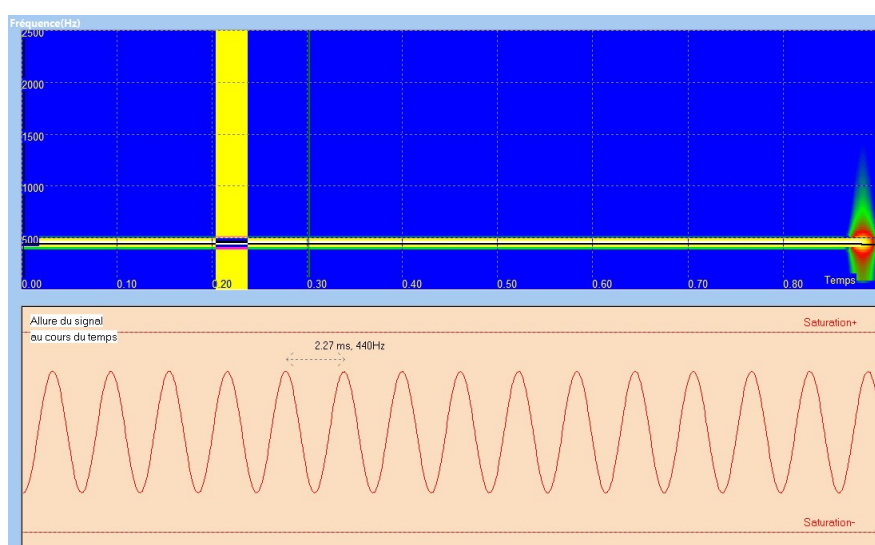


Figure 18 : Spectrogramme d'un son pur et allure temporelle de la portion sélectionnée

Spectrogramme d'un bruit pur

A l'inverse du son pur, le bruit pur, encore appelé bruit blanc par analogie avec la lumière (blanc=toutes les couleurs, ici blanc=toutes les fréquences), le spectrogramme est aléatoire, de puissance uniforme mais sans structure. L'allure temporelle est toute aussi aléatoire, sans période ni récurrence de quelque nature. La médiathèque de VOCALAB donne l'accès à cet échantillon par le bouton « Bruit blanc » (Figure 19).

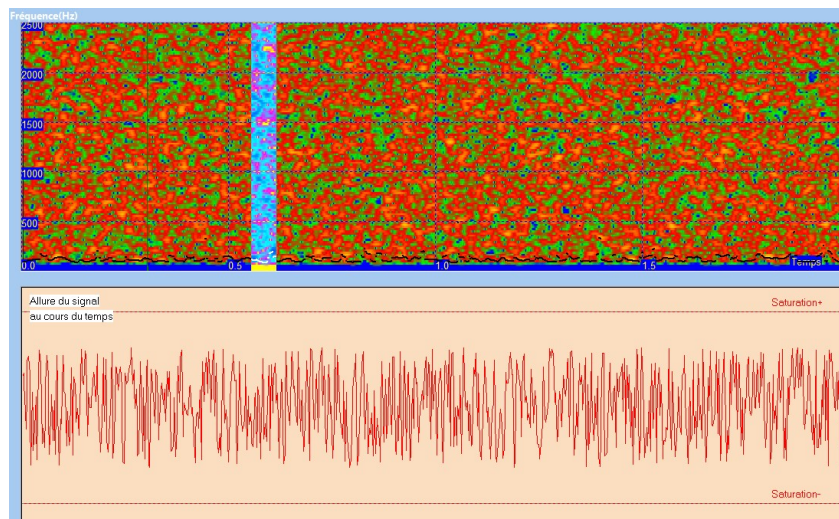


Figure 19 : Spectrogramme d'un bruit pur et allure temporelle de la portion sélectionnée montrant la structure aléatoire du signal temporel

Spectrogramme du /a:/ tenu

La voyelle /a/ fait partie des voyelles les plus énergétiques, avec le contenu harmonique très riche, et fait partie intégrante de nombreux bilans de voix et de parole. C'est aussi une voyelle universelle très étudiée dans la littérature scientifique. L'allure temporelle est assez complexe car s'additionnent différents contenus harmoniques, bien visibles sous forme de traits horizontaux dans le spectrogramme, associé au bruit de souffle (Figure 20).

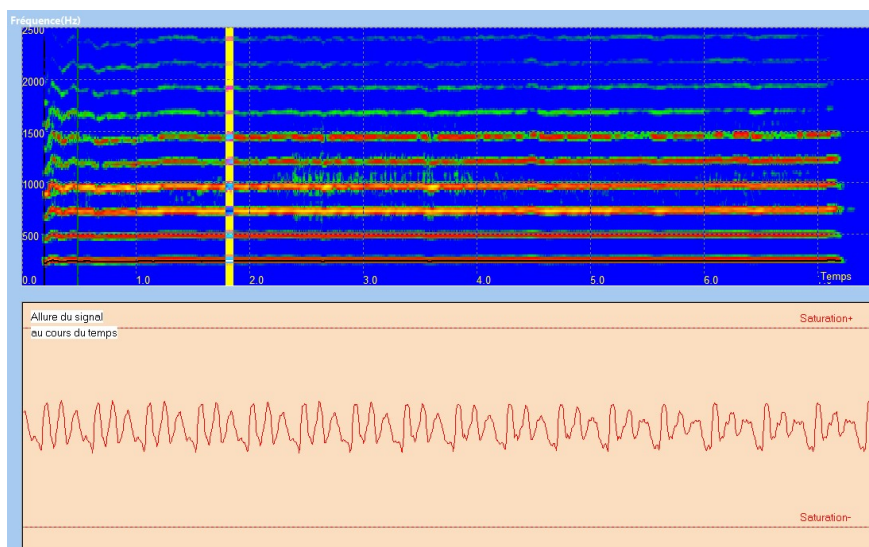


Figure 20 : Spectrogramme d'un /a:/ et allure temporelle de la portion sélectionnée

Spectrogramme des voyelles

Différentes voyelles prononcées par un homme de 46 ans sont reportées à la figure 21. Le resserrement de l'espace vertical entre les harmoniques est lié au F0 de la voix d'homme, autour de 100 Hz, deux fois plus faible que pour une femme. Les paquets d'énergie, équivalents aux formants, se déplacent du fait du changement de position de la langue, de l'ouverture de la bouche, de la position des lèvres et du mouvement d'ouverture du voile du palais. La place des formants change en fonction du genre, mais les positions relatives sont assez voisines.

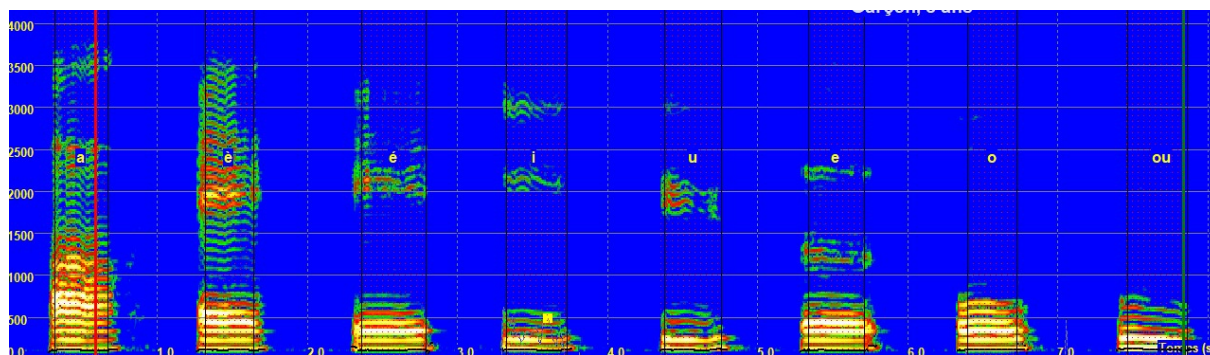


Figure 21 : Spectrogramme de voyelles prononcées par un homme

Les oppositions orales/nasales correspondant à une femme de 22 ans montrent aussi des différences significatives entre deux oppositions (a-an par exemple) et entre les différents groupes (Figure 22). Toutefois, une analyse plus précise, en particulier l'évaluation fine de la valeur des formants et des rapports d'énergie nécessite l'utilisation de l'outil de profil moyen, présenté à la section précédente.

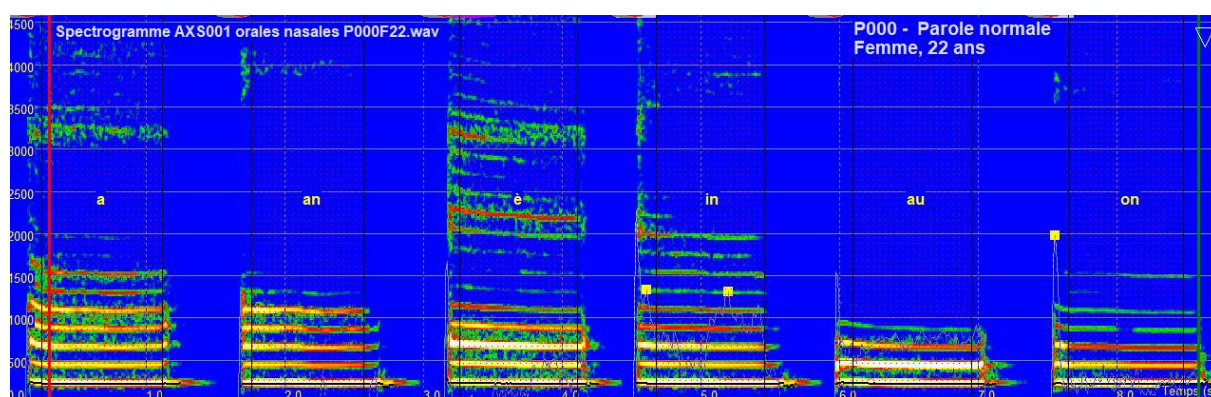


Figure 22 : Spectrogramme des oppositions de voyelles orales/nasales, femme de 22 ans, parole normale

Spectrogramme des consonnes

Consonnes occlusives sourdes

Les phonèmes /p/, /t/, /k/ correspondent à une occlusion, mise en pression suivies de phénomènes transitoires brefs, de l'ordre de quelques dizaines de millisecondes. Du point de vue spectral, on peut noter de légères différences entre les trois réalisations (Figure 23), sans toutefois pouvoir faire un lien clair entre le spectrogramme et les différents points de contact des occlusives.

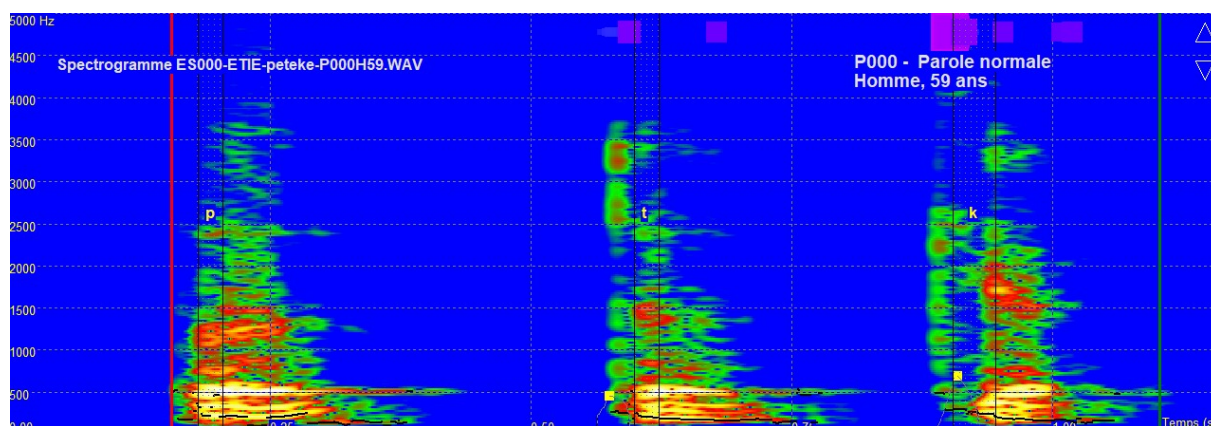


Figure 23 : Spectrogramme de /Pe, Te, Ke/, homme de 59 ans, parole normale

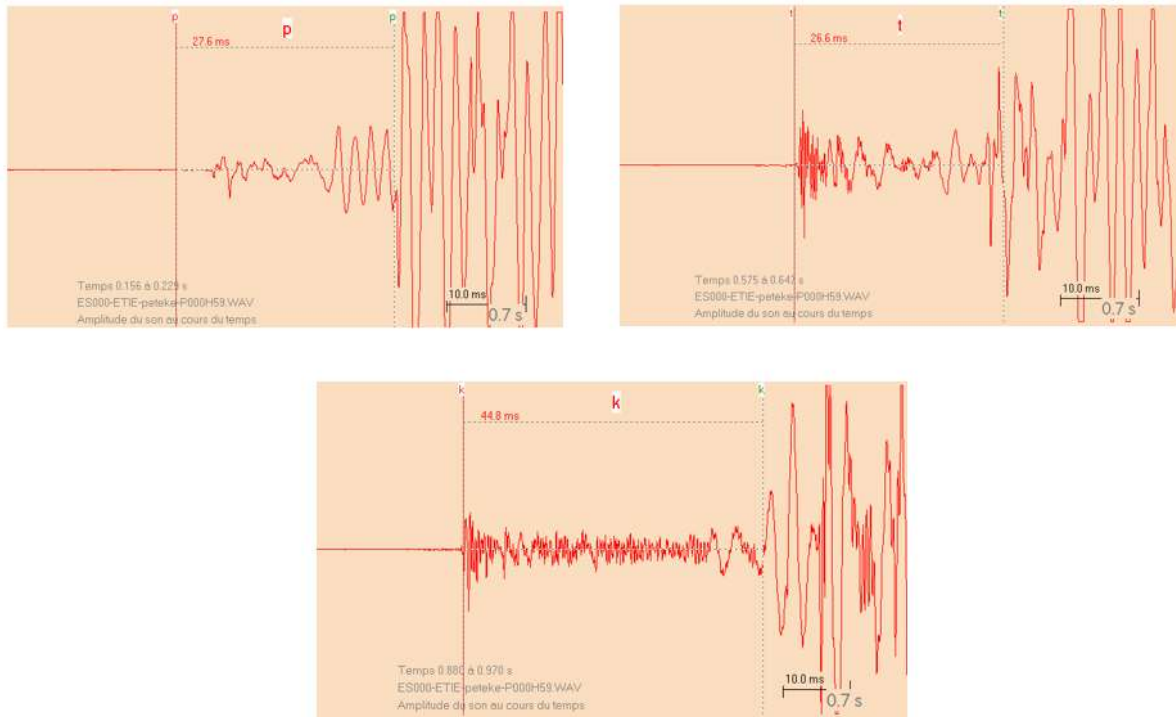


Figure 24 : Allure temporelle des plosives dans /Pe, Te, Ke/, homme de 59 ans, parole normale

On note en revanche, au niveau temporel, des différences de structures nettes, notamment en termes de durée, de forme du signal et de séquençage du signal, qu'illustre mal le spectrogramme car faisant reposer ses calculs sur trop grand nombre de points.

Fricatives sourdes et sonores

Autant les contrastes entre les plosives sont peu visibles sur le spectrogramme, autant les différences sont beaucoup plus nettes concernant les fricatives /f, ch, s/ et les oppositions sonores /v, j, z/. On constate un positionnement du bruit à des zones bien distinctes, ainsi que l'apparition nette du phénomène de voisement à très basse fréquence, dans la zone F0 (Figure 25).

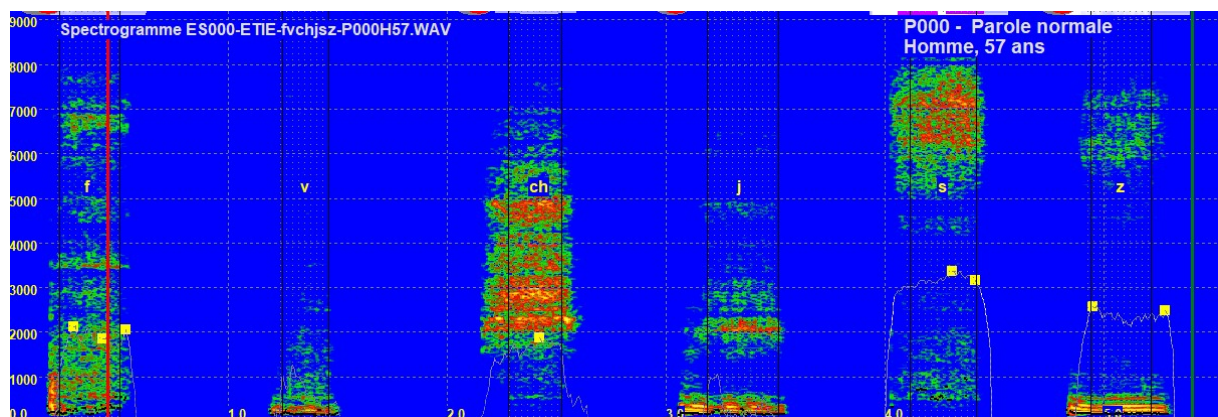


Figure 25 : Spectrogramme de /f,v,ch,j,s,z/, homme de 57 ans, parole normale

Le spectrogramme et ses applications cliniques

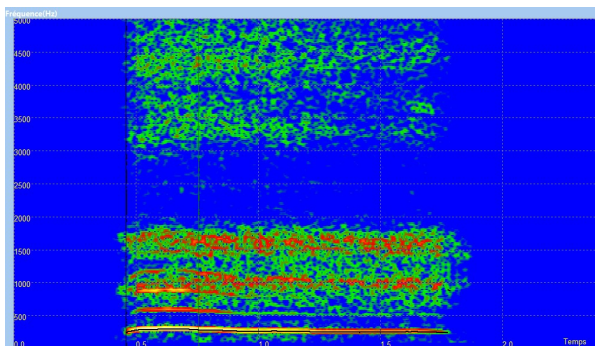
Nous donnons ici quelques pistes sur les applications du spectrogramme en clinique orthophonique.

Le spectrogramme en bilan de la voix & parole

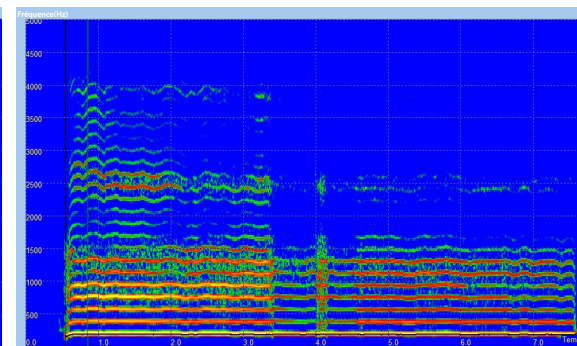
Analyse du /a:/ tenu

De très nombreux auteurs cités notamment par [Baken 2000][Kent 2000], et plus récemment [Jeffery 2018] utilisent la voyelle /a:/ tenue comme véhicule de différentes analyses de l'altération de la voix et de la parole. L'intérêt du /a:/ par rapport à des voyelles moins ouvertes est principalement sa grande richesse harmonique. Différents phénomènes peuvent être mis en évidence, tels qu'illustrés figure 26 grâce à des exemples puisés dans la base de données de voix en ligne de l'ERU 15 [Lurco 2020] :

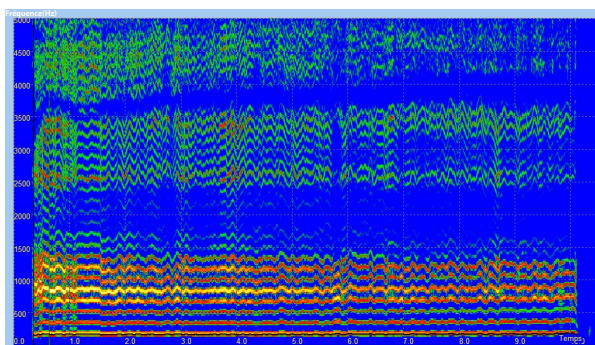
- /a:/ soufflé
- Déstructuration harmonique
- Instabilité en hauteur marquée
- Voix éraillée, avec présence de sous-harmoniques [Jeffery 2018]



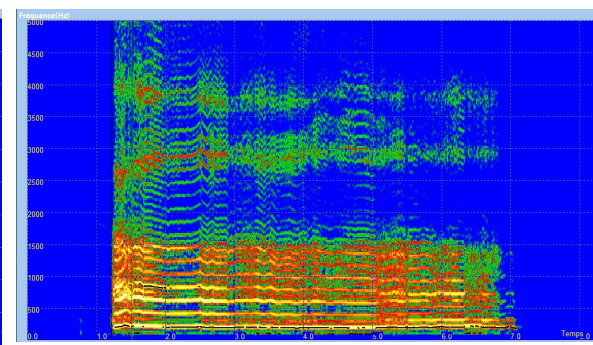
a) /a/ soufflé – AS036



b) Déstructuration harmonique - AS013



d) Voix instable en hauteur – AS031



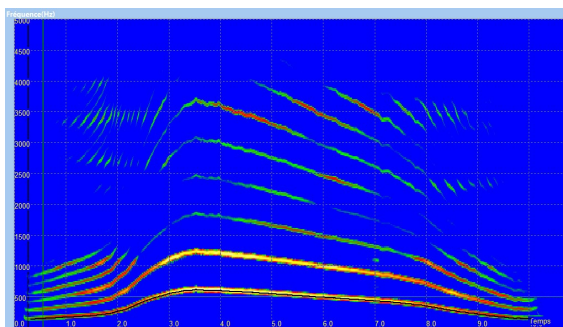
e) Voix éraillée – AS033

Figure 26 : Spectrogramme de voix altérées ou pathologiques

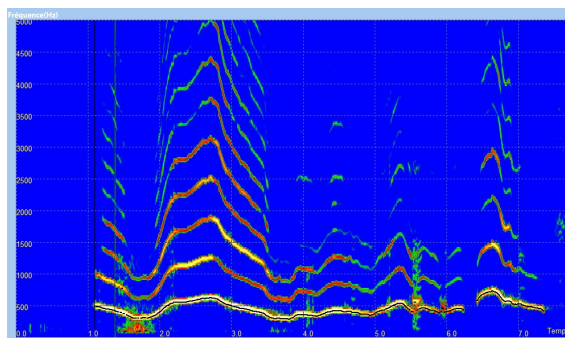
Le spectrogramme se révèle l'outil le plus performant pour relever une grande variété d'indices d'altération de la phonation. On peut encore citer le coup de glotte, le serrage, la pauvreté harmonique, autant de phénomènes bien identifiables dans la représentation spectrale.

Analyse de la sirène

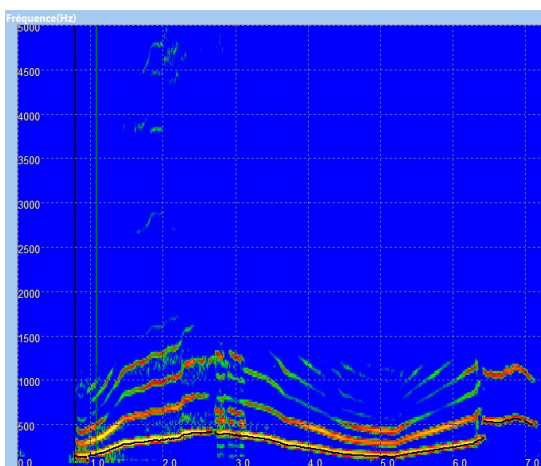
En effectuant une sirène sur un /ou/ (figure 27), on constate le déplacement progressif du fondamental F0 vers les hautes fréquences. On constate souvent une perte de résonance au milieu de la sirène liée au passage du mécanisme 1 au mécanisme 2.



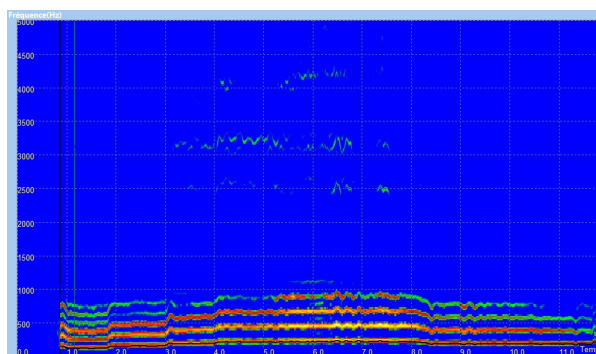
a) Sirène normale – AG001



b) Sirène avec difficultés de contrôle de hauteur – AS009



b) Sirène avec instabilité et éraillure – AS022



d) Sirène de faible étendue fréquentielle – AS031

Figure 27 : Spectrogramme de sirènes : normale (a), instable (b), éraillée (c) et de faible étendue fréquentielle (d)

Le spectrogramme en rééducation de la voix & parole

Spectrogramme en temps réel

Nous avons vu jusqu'ici des spectrogrammes construits d'après des enregistrements, ce qui est particulièrement utile en bilan ou pour établir un bilan d'évolution. Il est aussi possible d'utiliser le spectrogramme en temps réel, c'est-à-dire en calculant en direct les composantes fréquentielles de la voix ou de la parole capturées par le microphone. Dans l'exemple de la figure 28, un patient ayant des difficultés à prononcer une voyelle /a:/ tenue est encouragé à augmenter la durée de phonation et la puissance. Le patient réussit ici à passer de 3.5, 4.7 à 6.8 secondes au fil des réalisations. Ce type d'exercice est au cœur de méthodes de rééducation de la dysarthrie telles que LSVT™ [Fox 2012].

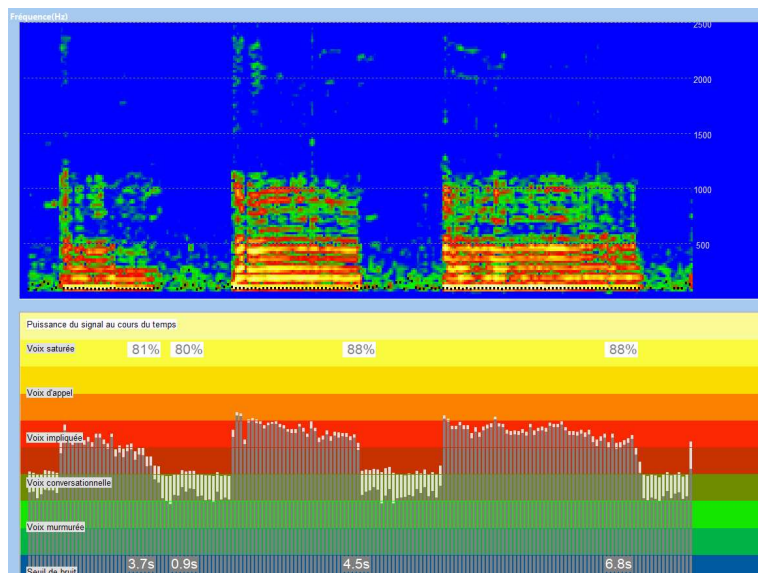


Figure 28 : Utilisation du spectrogramme en temps réel pour produire une voyelle /a:/ tenue de plus en plus longtemps, avec une voix forte

Travail de la résonance

Le travail de la résonance est un élément central des thérapies de la voix. Dans le cadre de nos formations au bilan et la rééducation de la voix [AMS 2020], nous insistons sur l'importance du travail de résonance vocalique suivant une méthodologie bien précise.

Dans le spectrogramme temps réel de la figure 29, le 1^{er} /a:/ est sourd, avec très peu d'harmoniques significatives au-dessus de 1500 Hz. Le 2^{ème} /a:/ est forcé mais inefficace : l'énergie se concentre à basse fréquence. Le travail en dissociation des mouvements permet au 3^{ème} /a:/ d'être beaucoup plus riche en harmoniques de rang élevé, tout en équilibrant la voix sans pour autant forcer.

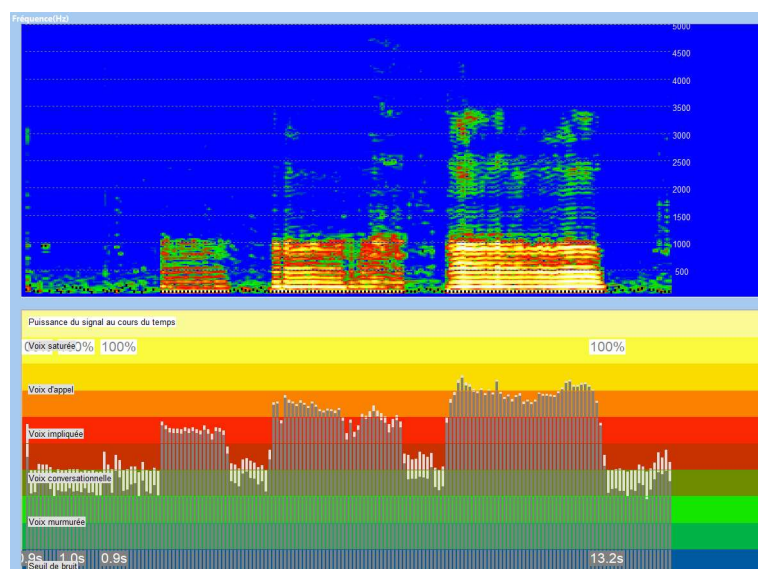


Figure 29 : Utilisation du spectrogramme en temps réel pour travailler la résonance sur une voyelle /a:/ tenue de hauteur confortable

Travail des oppositions constriction/voisement

Un autre exemple illustrant l'intérêt du spectrogramme concerne les oppositions de consonnes fricatives non voisées et voisées, telles que /f-v/. L'orthophoniste demande au patient d'enchaîner /f-v-f-v.. / sans interruption et en tentant de ne pas diminuer le niveau de puissance. On constate sur le graphique une amélioration des performances.

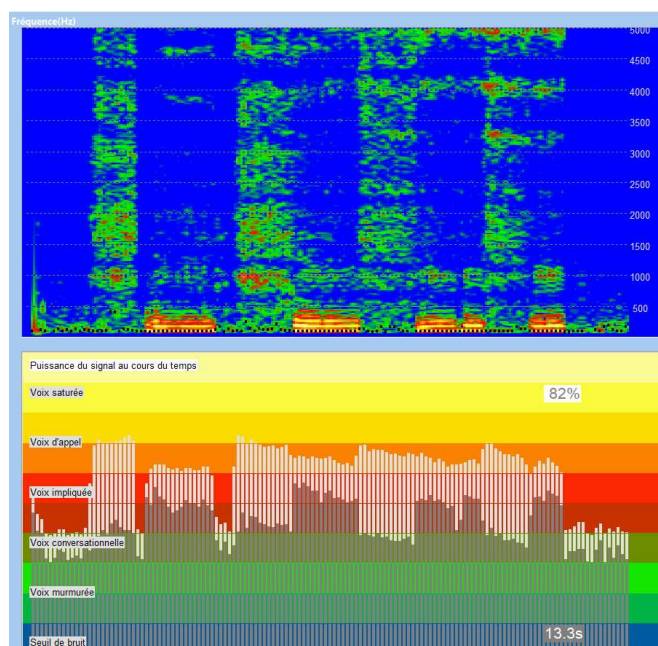


Figure 29 : Travail des oppositions /f-v/ en recherche de continuité entre les portions voisées et non voisées.

La rééducation ne doit pas se focaliser sur la pression sous-glottique mais plutôt sur la pression sus-glottique en optimisant le rôle des articulateurs et leurs mouvements dissociés. L'approche AMCO en rééducation de la phonation se focalise non pas sur le larynx mais sur ce qui va contribuer à le soulager. Ceci dans le cas des dysphonies dites dysfonctionnelles non lésionnelles. La visualisation immédiate des progrès sous l'effet des modifications apportées à l'articulation motive le patient et rend le suivi plus efficace.

Conclusion

Le spectrogramme, qui donne une visualisation des composantes spectrales dans le temps, peut être considéré comme l'un des points forts de l'analyse de la voix et de la parole. Très informatif, impartial, d'une précision bien supérieure à l'oreille humaine, le spectrogramme est un peu déroutant à décrypter, de par sa représentation temps/fréquence. L'apprentissage de la lecture des différentes composantes fréquentielles des sons se fait en général étape par étape, du son pur au bruit en passant par les voyelles, consonnes, syllabes et mots. Le spectrogramme est un outil très utile en bilan, robuste et peu sensible aux conditions d'utilisation. Il se révèle aussi particulièrement pertinent, voire indispensable en rééducation, en mode temps réel, servant de véritable miroir du contrôle de l'intensité, du souffle, de la résonance, qui sont au cœur de multiples approches thérapeutiques. Un certain nombre d'outils complémentaires a été décrit dans cet article, parmi lesquels on peut citer la représentation temporelle, l'évolution de la puissance, la segmentation, ainsi que les indicateurs ciblant des altérations spécifiques de la voix et de la parole.

Remerciements

Nous saluons aussi les efforts de mise en archive ouvertes d'un très grand nombre de résultats scientifiques, en particulier dans le domaine de la parole, ce qui permet une diffusion très large, simple et sans droit d'accès, autorisant de ce fait des analyses comparatives et méta-analyses qui accélèrent ainsi leur mise en application clinique. Enfin, nous remercions 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 et 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, ainsi que de MICROWIND et IC-EMC en lien avec la conception des circuits intégrés. Il a co-écrit une vingtaine d'ouvrages et plus de 200 publications scientifiques dans le domaine de l'analyse de la voix et parole, des circuits intégrés, et de la compatibilité électromagnétique. Il a été nommé en 2006 *Distinguished Lecturer* de la société savante IEEE.

Site : www.etienne-sicard.fr

Email : 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 (DeBoeck, 2017) ainsi que de nombreuses publications dans le domaine de la voix et de la parole pathologique.

Site : <https://www.formationsvoixparole.fr>

Email : anne.sicard2@orange.fr

Références

[AMS 2020] <https://www.formationsvoixparole.fr>

[Audacity 2018] <https://www.audacityteam.org/>

[Baken 2000] Baken, R. J., & Orlikoff, R. F. (2000). Clinical measurement of speech and voice. Cengage Learning.

[Barsties 2019] Barsties v. Latoszek, B., Ulozaitė-Stanienė, N., Petrauskas, T., Uloza, V., & Maryn, Y. (2019). Diagnostic accuracy of dysphonia classification of DSI and AVQI. The Laryngoscope, 129(3), 692-698.

[Boersma 2018] Boersma P. & Weenink D. (2018): Praat: doing phonetics by computer [Computer program]. Version 6.0.37, retrieved 14 March 2018 from <http://www.praat.org/>

[Cornut 2001] Cornut, G., & Arom, S. (2001). Moyens d'investigation et pédagogie de la voix chantée: Actes du colloque, Lyon. Symétrie.

- [Fox 2012] Fox, C. M., & Boliek, C. A. (2012). Intensive voice treatment (LSVT LOUD) for children with spastic cerebral palsy and dysarthria. *Journal of Speech, Language, and Hearing Research*.
- [Gerip 2020] VOCALAB 4.1 (2020). Edité par GERIP <https://www.gerip.com>
- [Ghio 2007] Ghio, A. (2007). L'onde sonore : réalités physiques et perception. Auzou P.; Rolland V.; Pinto S. ; Ozsancak C. Les dysarthries, Solal, 2007. (hal-01616690)
- [Jeffery 2018] Jeffery, T., Cunningham, S., & Whiteside, S. P. (2018). Analyses of sustained vowels in down syndrome (DS): a case study using spectrograms and perturbation data to investigate voice quality in four adults with DS. *Journal of Voice*, 32(5), 644-e11.
- [Haralick 1973] Haralick, R. M., Shanmugam, K. and Dinstein, I. (1973) "Textural features for image classification," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-3, no. 6, pp. 610–621, Nov 1973
- [Kacha 2020] Kacha, A., Grenez, F., Orozco-Arroyave, J. R., & Schoentgen, J. (2020). Principal component analysis of the spectrogram of the speech signal: interpretation and application to dysarthric speech. *Computer Speech & Language*, 59, 114-122.
- [Kent 1997] Kent, R. D., (1997). *Speech Sciences*. Thomson Delmar Learning.
- [Kent 2000] Kent, R. D., & Ball, M. J. (Eds.). (2000). *Voice quality measurement*. Singular.
- [Koenig 1947] Koenig, W. (1947), *The Sound Spectrograph*. Thirty-First Meeting of the Acoustical Society of America, 1947
- [Lefol 2018] Lefol, G. (2018) "Vers une évaluation objective de l'articulation en pratique clinique orthophonique", Mémoire de Grade Master 2 en Orthophonie, Université de Tours, Faculté de médecine, Centre de Formation Universitaire en Orthophonie, dumas-01815287, 2018
- [Menin-Sicard 2013] Menin-Sicard, A., (2013). L'évaluation vocale, Rééducation orthophonique N° 254.
- [Menin-Sicard 2016] Menin-Sicard, A., Sicard, E. (2016). Évaluation et réhabilitation de la voix - Approche clinique et objective avec VOCALAB, 288 pages, ISBN 9782353273188, DeBoeck Supérieur
- [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, Mai 2019, Mons, Belgique. <https://hal.archives-ouvertes.fr/hal-02127039>
- [Parmentier 2014] Parmentier, F. Marié-Bailly, I., Pillot-Loiseau, C. (2014). A l'écoute de sa voix : apports du spectrogramme comme rétrocontrôle visuel en rééducation vocale. *Glossa, UNADREO - Union Nationale pour le Développement de la Recherche en Orthophonie*, 2014, Glossa, 116, pp.18 - 32. (hal-01099451)
- [Sharma 2020] Sharma, G., Prasad, D., Umapathy, K., & Krishnan, S. (2020, July). Screening and analysis of specific language impairment in young children by analyzing the textures of speech signal. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* (pp. 964-967). IEEE.

[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, Ortho édition, 2013, pp.23-27. <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. 7èmes Journées de Phonétique Clinique, Laboratoire de Phonétique et Phonologie UMR 7018, Jun 2017, Paris, France. <hal-01510418>

[Sicard 2019b] Sicard E., Menin-Sicard A. (2019). Analyse de la qualité des phrases pour un bilan objectif de la parole. [Rapport de recherche] INSA Toulouse. 2019. <https://hal.archives-ouvertes.fr/hal-02389764>

[Sicard 2020a] Sicard E., Menin-Sicard A. (2020). DIADOLAB: bilan court et objectif de la dysarthrie dans le cadre de la prise en charge orthophonique. Journées Nationales de Neurologie en langue Française, Avril 2020, Lyon, France. <https://hal.archives-ouvertes.fr/hal-02396136>

[Sicard 2020b] Sicard, E., & Menin-Sicard, A. (2020). Le triangle vocalique et son application en contexte orthophonique. Rapport de recherches en archives ouvertes. INSA Toulouse, LURCO/UNADREO. <https://hal.archives-ouvertes.fr/hal-02504513/>

[Smola 2004] Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. Statistics and computing, 14(3), 199-222.

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

[Yu 2012] Yu, H., & Kim, S. (2012). SVM Tutorial-Classification, Regression and Ranking. Handbook of Natural computing, 1, 479-506.

[Zahid 2020] Zahid, L., & al. (2020). A Spectrogram-Based Deep Feature Assisted Computer-Aided Diagnostic System for Parkinson's Disease. IEEE Access, 8, 35482-35495.