



HAL
open science

Le triangle vocalique et son application en contexte orthophonique

Etienne Sicard, Anne Menin-Sicard

► **To cite this version:**

Etienne Sicard, Anne Menin-Sicard. Le triangle vocalique et son application en contexte orthophonique. [Rapport de recherche] INSA Toulouse. 2020. <hal-02504513>

HAL Id: hal-02504513

<https://hal.science/hal-02504513v1>

Submitted on 10 Mar 2020

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 triangle vocalique et son application en contexte 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)

etienne.sicard@insa-toulouse.fr

anne.sicard2@orange.fr

Résumé :

Cet article présente le triangle vocalique et son application possible dans le cadre de l'évaluation et de la rééducation de la parole en orthophonie. Après avoir rappelé les principes fondamentaux de la résonance des tubes, nous illustrons le lien entre voyelles et formants, puis analysons leurs valeurs en tentant d'expliquer les variabilités importantes dans les publications. Nous comparons aussi les résultats obtenus par les logiciels PRAAT et DIADOLAB, avant de donner différents exemples d'études comparatives utilisant les triangles vocaliques. Nous définissons ensuite les métriques principales telles que les distances entre les extrêmes et l'aire vocalique, en donnant quelques exemples de valeurs pour homme, femme et enfant. Nous proposons aussi une extension aux consonnes, permettant de représenter l'ensemble des phonèmes sur un même graphique pour visualiser les altérations de la parole ainsi que les progrès accomplis grâce à la prise en charge orthophonique. Quelques statistiques utilisant le triangle vocalique complètent cet article.

Mots clé : orthophonie, évaluation de la parole, processus, triangle vocalique, extrêmes consonantiques, occurrence, phonotactique

Introduction

De nombreuses études sur la parole utilisent le formalisme du triangle vocalique pour synthétiser des résultats permettant d'illustrer certaines caractéristiques de parole, dans différents contextes d'analyse tels le débit de parole [Gendrot 2005], l'âge des enfants [Kent 2018], la régionalisation [Picart 2010], le bégaiement [Hirsch 2008] ou encore la parole dysarthrique [Roland 2017] [Audibert 2012]. Le triangle vocalique repose sur l'hypothèse que les différentes voyelles peuvent être placées dans un espace à deux dimensions, en lien avec la valeur des deux premières zones de résonance, encore appelés formants F1 et F2, eux-mêmes corrélés à la position des articulateurs (principalement mâchoire et langue). Les voyelles /a,i,ou/ se positionnent alors comme les extrêmes d'un triangle appelé triangle vocalique.

Il nous a paru intéressant d'étudier l'intérêt potentiel du triangle vocalique dans le contexte particulier de l'évaluation et de la rééducation orthophonique de la parole, en tentant d'évaluer le degré de fiabilité et de pertinence de cette représentation et de ses métriques associées. Un tour d'horizon des publications incluant des représentations à base de triangle vocalique a été fait au préalable. Nous avons aussi porté une attention particulière à la voix des sages telles que celle de R.D. Kent sur le sujet [Kent 2018], grâce à des méta-analyses particulièrement informatives et étayées.

Nous nous sommes aussi posé la question de la fiabilité des approches logicielles concernant les valeurs des formants, notamment dans le contexte de la parole de l'enfant ou de la parole altérée et pathologique qui est au cœur des préoccupations des orthophonistes.

Dans cet article, nous repons quelques bases théoriques de la résonance des tubes (partie 1), nous illustrons le lien entre voyelles et formants avec la position de la langue et de la mâchoire (partie 2) qui nous amène à présenter le triangle vocalique. Les valeurs de F1 et F2 issues de la littérature montrent une variabilité surprenante que nous tentons d'expliquer. Nous comparons aussi les valeurs obtenues par les logiciels PRAAT et DIADOLAB sur un corpus identique, avant de donner différents exemples d'études comparatives utilisant les triangles vocaliques, concernant le débit de parole, l'effet de la régionalisation, le discours hypo-hyper articulé, ainsi que différentes pathologies telles que le bégaiement ou la dysarthrie. Nous définissons en section 3, les métriques principales telles que les distances entre les extrêmes et l'aire vocalique, en donnant quelques exemples de valeurs pour homme, femme et enfant. Nous proposons à la section 4 une extension aux consonnes permettant de représenter l'ensemble des phonèmes de la langue française sur un même graphique, de visualiser les altérations de la parole ainsi que les progrès accomplis grâce à un suivi orthophonique. Quelques exemples d'études statistiques utilisant le triangle vocalique, notamment autour de la voix de l'enfant de 5 ans, complètent cet article.

La résonance des tubes

Mise en formule

Si l'on considère la distance l entre les cordes vocales et les lèvres (Figure 1), et que l'on approxime le trajet du son à un simple conduit, le tube correspondant comporte des résonances naturelles à des fréquences F_n appelées formants. Dans l'équation 1, c représente la vitesse de propagation du son dans l'air à 37°C, soit 356 m/s.

$$F_n = \frac{(2n - 1)}{4l} c$$

(Eq. 1)

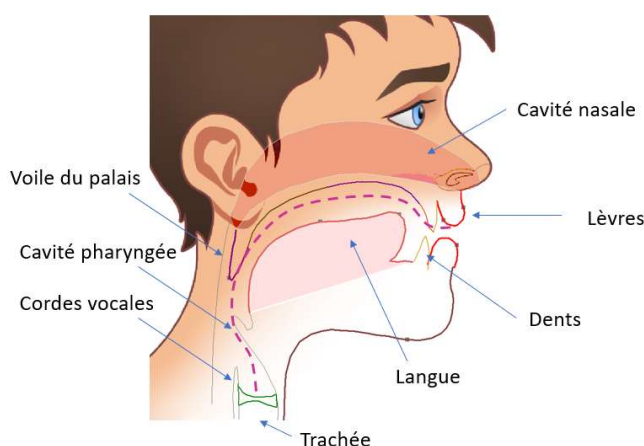


Figure 1 : trajet depuis les cordes vocales jusqu'aux lèvres, permettant d'évaluer la distance l .

En appliquant l'équation 1 pour 3 longueurs type représentative des enfants de 5 ans, femmes et hommes, nous obtenons les chiffres du tableau 1. Les valeurs obtenues sont très différentes selon le genre, allant du simple au double entre homme et enfant. La série des formants et des fréquences de résonance associées F_1, F_2, \dots, F_n est à priori sans limite, si ce n'est la fréquence maximale d'audition de l'oreille, mais on étend en général les calculs jusqu'à 10 KHz environ. Cependant, on se focalise principalement sur F_1 et F_2 , ce qui donne une plage d'analyse encore plus restreinte, de 0 à 6000 Hz environ.

	L (CM)	F1 (HZ)	F2 (HZ)	F3 (HZ)
Enfant	8	1106	3318	5531
Femme	12	737	2212	3687
Homme	18	492	1475	2458

Tableau 1 : Résonances naturelles d'un tube de longueur 8, 12 ou 18 cm

Si l'on compare avec l'étude bibliographique de [Kent 2018], on retrouve des données analogues. La figure 2 précise aussi F_1 autour de 500 Hz pour les hommes, 750 Hz pour les femmes et 1000 Hz pour les enfants de 5 ans.

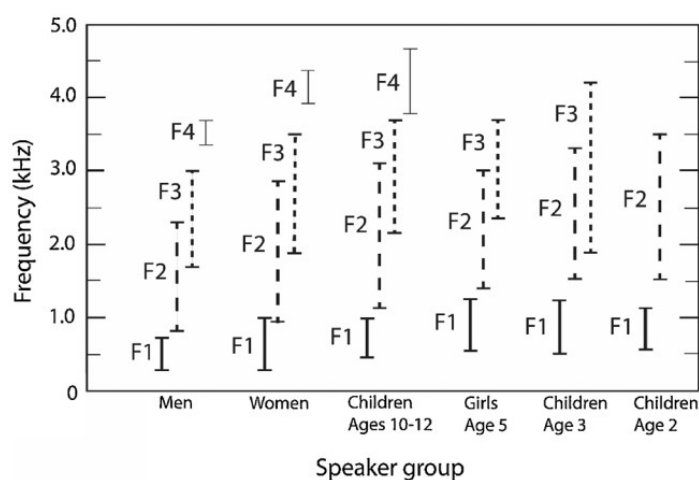


Figure 2 : compilation de différentes données de la littérature Anglaise par [Kent 2018] sur les valeurs des formants

Le spectre moyen de la parole

Une illustration de la notion de formant est proposée sur la voyelle /e:/ prononcée par un homme de 59 ans (Figure 3) à l'aide du logiciel DIADOLAB [Diadolab 2019]. Nous notons /e:/ la voyelle /e/ tenue pendant plusieurs secondes. Le spectrogramme de la figure 3 est un calcul au cours du temps des composantes fréquentielles de la voix. Ici, la plage de fréquence (axe des Y) est de 0 à 4500 Hz. Les traits horizontaux représentent les harmoniques, les couleurs représentent la notion d'énergie (bleu : pas d'énergie, rouge : énergie moyenne, blanc : énergie forte).

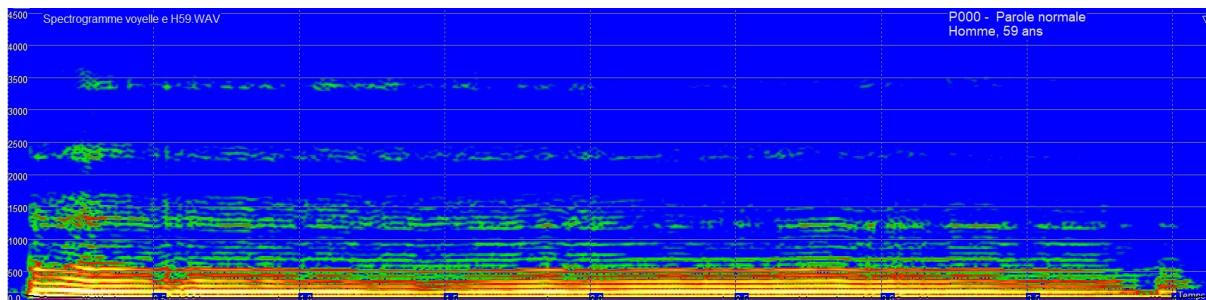


Figure 3 : spectrogramme d'une voyelle /e:/ tenue avec DIADOLAB, homme de 59 ans

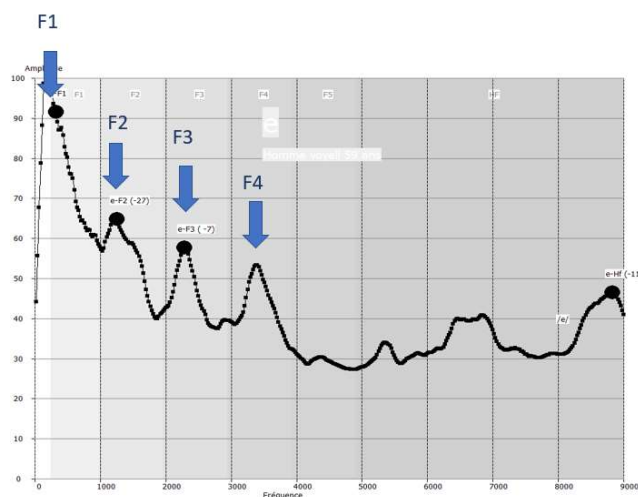


Figure 4 : spectre moyen d'une voyelle /e:/ tenue prononcée par un homme de 59 ans montrant les formants F1..F4

Si l'on considère la moyenne du spectrogramme sur l'ensemble de la production sonore du /e:/, on obtient le dessin de la figure 4. Sur un long échantillon de parole, cette représentation s'appelle aussi *Long Term Average Speech Spectrum* (LTASS) [Byrne 1994]. Sur une seule voyelle /e:/, on peut distinguer des résonances à différentes fréquences, que l'on peut considérer comme les formants F1, F2, F3, etc. A noter que la représentation de l'axe des X est ici en linéaire, qui montre bien le côté périodique des formants, tel que proposé par l'équation 1. L'amplitude Y est par contre en échelle log (dB). Les valeurs du tableau 1 (F1 : 492 Hz, F2 : 1475 Hz, F3 : 2458 Hz) sont relativement proches des valeurs observées ici. Il devient en revanche difficile d'identifier dans ce cas précis les pics au-delà de F4 avec netteté.

Prise en compte de la résonance nasale

Si l'on considère plus attentivement la structure anatomique de l'oro-pharynx, on peut considérer une première structure basée sur la cavité pharyngée, suivie de la cavité orale dans les voyelles ou consonnes orales. Dans le cas des voyelles ou consonnes nasales, on doit considérer un chemin supplémentaire via la zone de naso-pharynx, les sinus et le nez. Différents auteurs ont mis en évidence des résonances et anti-résonances supplémentaires en lien avec la structure décrite par [Pruthi 2007], dont la modélisation est reproduite figure 3. Les résonances deviennent alors plus complexes, comme en témoigne la grande variété des indices de nasalités répertoriées dans la méta-analyse de [Baken 2000].

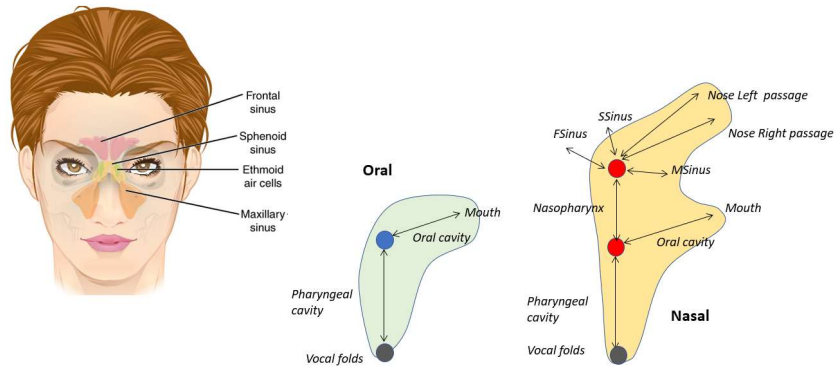


Figure 3 : Dans le cas de voyelles et consonnes nasales, 3 tubes et des résonances multiples doivent être considérées [Pruthi 2007]

Le triangle vocalique

Une étude historique de [Vilain 2015] décrit les différentes représentations des voyelles, en remontant aux descriptions de Hellwag en 1781 (Fig. 4 à gauche) ou Chladni en 1809 (Fig. 4 à droite). L'étude cite notamment la modélisation à base de tubes du conduit vocal et la notion de fréquences de résonance associée (Figure 5) par Fant en 1960, qui introduit aussi la notion de nomogramme.

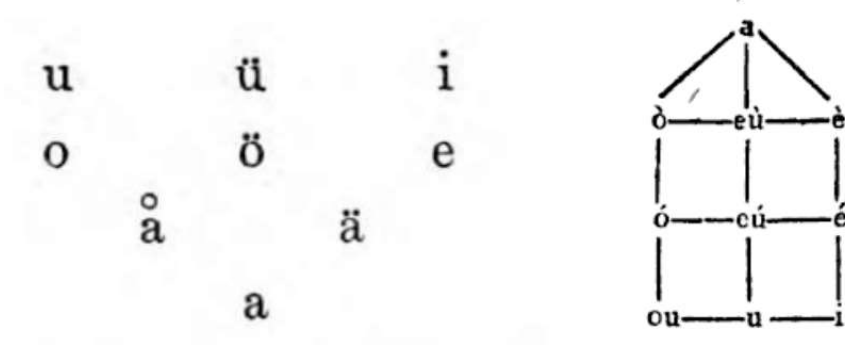


Figure 4 : Les représentations les plus anciennes du triangle vocalique au XIIIème siècle [Vilain 2015]

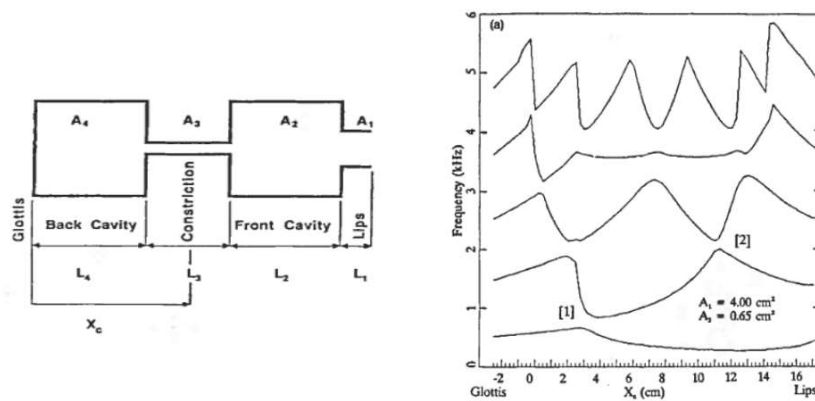


Figure 5 : Représentation du conduit vocal et notion de nomogramme par Fant en 1960 [Vilain 2015]

Lien entre voyelles et formants

L'ouverture de la bouche comme dans le /a/, l'avancée de la langue et le rétrécissement des lèvres comme dans le /i/, ou le recul de la langue et la projection des lèvres comme dans le /ou/ modifient considérablement la forme du conduit vocal, ce qui a pour effet de modifier les valeurs des formants F1, F2, F3, etc. autour de leur position moyenne (Figure 6).

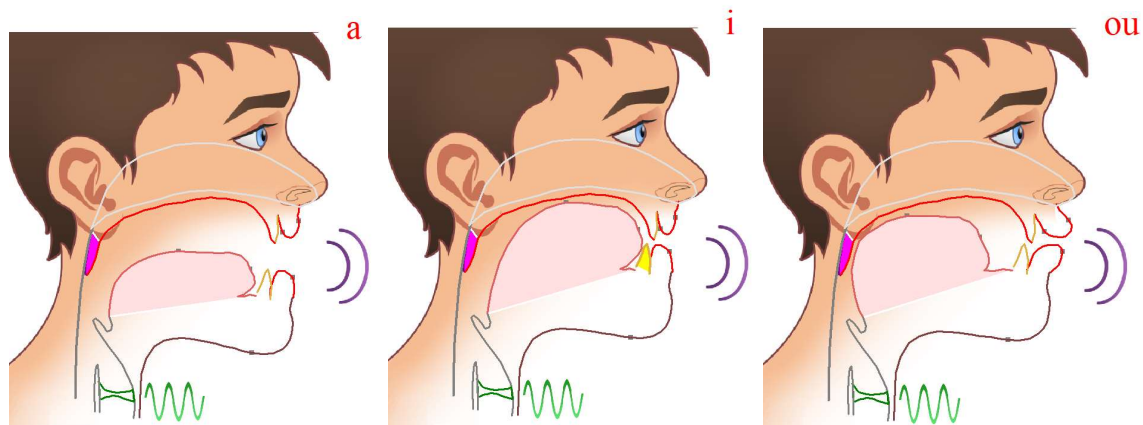
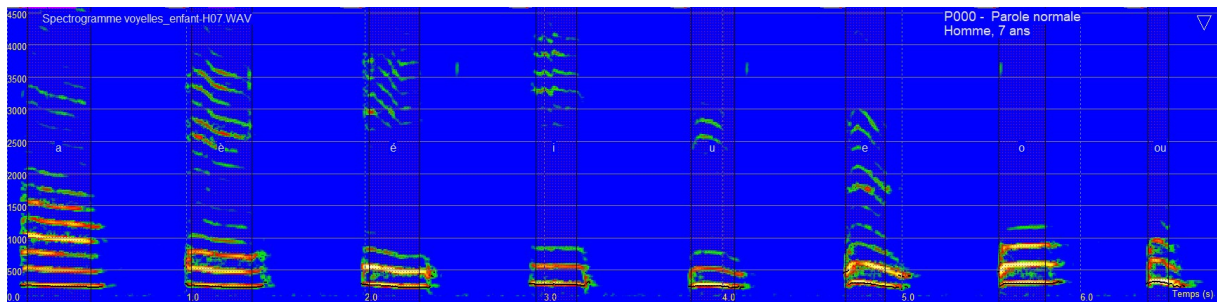


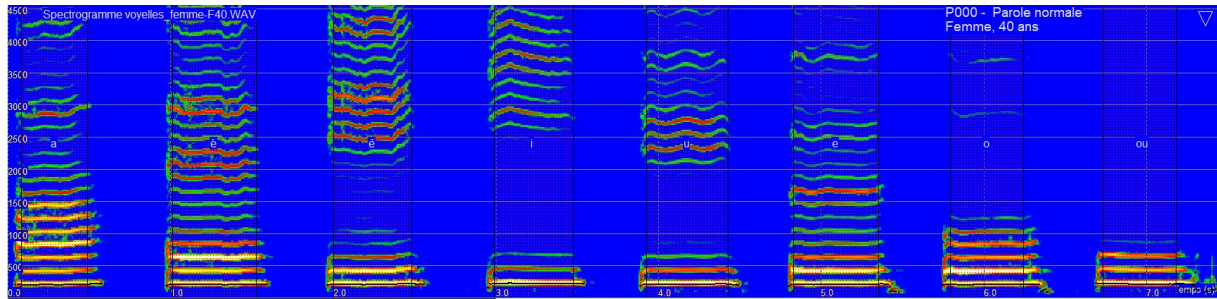
Figure 6 : Position de la langue et de la mâchoire pour /a,i,ou/ illustré avec Diadolab

Le spectrogramme de différentes voyelles prononcées par un enfant, une femme et un homme est reporté figure 7. On constate un nombre de traits horizontaux plus faible chez l'enfant, simplement parce que le fondamental usuel moyen (FUM), correspondant au trait horizontal noir qui matérialise la trajectoire du fondamental de la voix est plus haut (264 Hz). Concernant l'homme, les harmoniques sont resserrées car multiples d'un FUM beaucoup plus bas (110 Hz). Les 3 valeurs de FUM sont proches des normes usuelles en voix conversationnelle pour homme, femme et enfant.

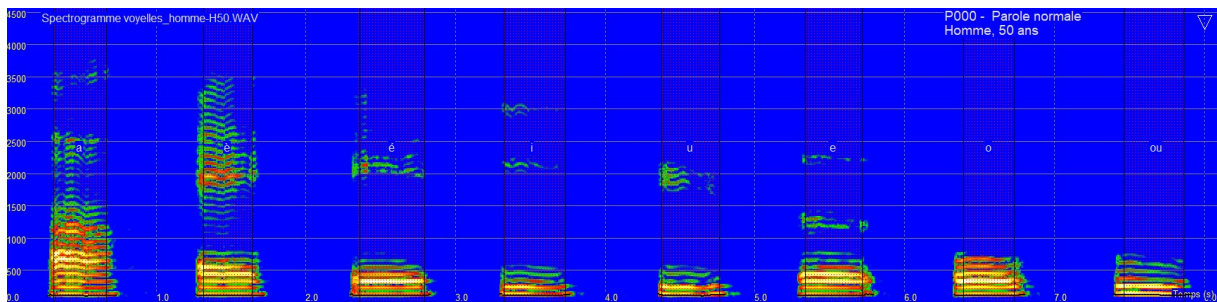
On peut constater des regroupements d'harmoniques dans des zones fréquentielles significativement différentes selon les voyelles. Ces regroupements correspondent aux notions de formants. Les formants F1, F2 (ainsi que les suivants) sont donc différents selon les voyelles. Les regroupements d'énergie sont positionnés de manière similaire pour homme, femme et enfant, mais on peut constater des valeurs plus faibles pour les hommes : par exemple, le formant F2 pour le /é/ démarre au-dessus de 3000 Hz pour l'enfant, 2500 Hz pour la femme et 2000 Hz pour l'homme, ce qui est conforme à la théorie des tubes.



a-Garçon de 7 ans – FUM 264 Hz



b-Femme de 40 ans – FUM 210 Hz



c-Homme de 50 ans – FUM 110 Hz

Figure 7 : Spectrogramme des voyelles /a,è,é,i,u,e,o,ou/ prononcées par (a) un garçon de 7 ans (b) une femme de 40 ans et (c) un homme de 50 ans

Représentation F1-F2

Le changement de position de la langue et l'ouverture de la bouche modifient au premier ordre les valeurs des formants F1 et F2. Une représentation commune des voyelles consiste à extraire la valeur de F1 et de F2 et positionner la valeur correspondante dans un espace à deux dimensions, où X représente la position de la langue avant/arrière, et Y l'ouverture de la bouche. Trois valeurs extrêmes se distinguent, comme dessiné figure 8 :

- Le /a/ en bas
- Le /i/ à droite
- Le /ou/ à gauche

Si les valeurs changent considérablement selon homme, femme ou enfant, les positions respectives des voyelles dans l'espace F1,F2 restent à peu près les mêmes, et les extrêmes du triangle sont toujours /a,i,ou/.

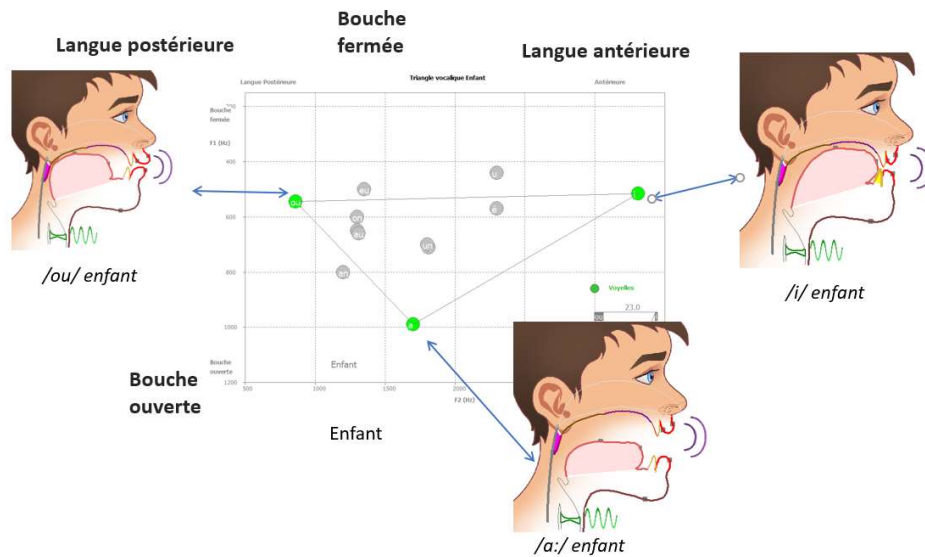


Figure 8 : Représentation des différentes voyelles de la langue française mettant en évidence les extrêmes d'un triangle /a,i,ou/, valeurs pour un enfant. L'axe X correspond à la valeur du formant F2, l'axe Y au formant F1.

Valeurs de F1 et F2

De très nombreux auteurs ont proposé des valeurs de F1 et F2, notamment pour /a,i,ou/, principalement en langue Anglaise [Baken 2000] [Meunier 2007] [Nicolaidis 2016] [Eichhorn 2018] [Deterding 1997] [Hillenbrand 1995], mais aussi en langue Française [Gendrot 2005] [Audibert 2018] [Boula 2007][Picart 2010][Kahn 2011]. Des méta-analyses telles que [Kent 2018] permettent de compiler un nombre considérable de données en lien notamment avec les valeurs de formants, dont très peu concernent malheureusement la langue Française. On peut constater, d'après les différents points de la figure 9, chacun représentant une source d'information, des écarts importants entre les valeurs publiées, laissant supposer des différences de protocoles, d'outils d'extraction et de corpus étudiés.

Dans [Kent 2018], plus de 15 articles scientifiques sont cités en lien avec la valeur des formants pour des adultes, ainsi que des enfants de 1 à 18 ans, en particulier ceux de [Assman 2008] et [Lee 1999], avec encore une fois des disparités importantes, notamment sur le formant F1 du /ou/ et F2 du /a/.

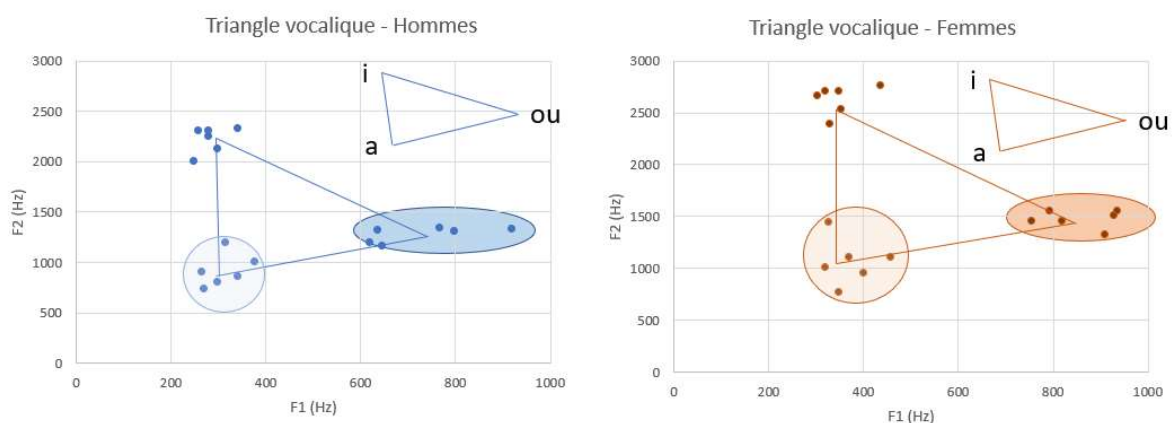


Figure 9 : Valeurs de F1 et F2 pour /a,i,ou/ selon diverses publications – Hommes à gauche, Femmes à droite

Origine de la variabilité

On peut tenter de donner quelques hypothèses sur l'origine de la variabilité constatée dans les valeurs des formants. Tout d'abord, l'extraction des fréquences des formants se fait pour des voyelles produites dans des contextes différents selon les études, tels que :

- La production de voyelles isolées comme illustré figure 7 ou étudiés dans [Roland 2017].
- La production de syllabes rattachées à des voyelles. Le mémoire de G. Lefol [Lefol 2018] ainsi que celui de A. Papazian [Papazian 2019] se basent sur la forme consonne-voyelle (CV), notamment avec la consonne /p/. En langue Anglaise [Kent 2018], la structure /hVd/ est parmi les plus étudiée car le /h/ minimise les articulations supraglottiques qui pourraient influencer les caractéristiques de la voyelle qui lui succède. De même, l'influence de la consonne finale doit être minimale, et si possible totalement neutre vis-à-vis des valeurs de formants de la voyelle qui lui précède.
- La production d'une phrase. Les mémoires de T. Bardelang [Bardelang 2020] et C. Grenier [Grenier 2020] utilisent la phrase « Bonjour Monsieur Tralipau » proposée dans le cadre du bilan orthophonique de la parole [Sicard 2019]. Les phénomènes de co-articulation naturelle peuvent avoir une influence sur les voyelles cible : /jour/, /tral/ et /lip/
- La lecture d'un texte, la répétitions d'une phrase avec ou sans modèle sonore, ou encore la parole spontanée.

D'autres facteurs tels que l'état physique et émotionnel du patient, le stress vis-à-vis de la procédure d'enregistrement, le style de parole utilisé (voix conversationnelle, voix impliquée, voix projetée), la complexité des mots à prononcer ou encore la fatigabilité du patient en lien avec la durée du protocole de test global peuvent influencer les résultats. La question du modèle pré-enregistré fourni est aussi un facteur pouvant influencer le résultat ainsi que les consignes du protocole. D'autre part, l'extraction des formants fait appel à des algorithmes de calculs qui, sur un même échantillon de parole, peuvent donner des résultats assez différents [Woerling 2007].

- PRAAT utilise l'approche Linear Predictive Coding (LPC), sans autoadaptation au sexe et à l'âge du patient. Une discussion sur l'optimisation de PRAAT, notamment sur l'ordre de la LPC selon le sexe est reportée dans [Skarnitzl 2015]. On peut supposer que la configuration par défaut (ordre 10 soit 5 formants entre 0 et 5500 Hz) n'est pas adaptée à la voix de l'enfant, qui n'en comporte que 3 dans cette bande.
- DIADOLAB utilise l'analyse de la Transformée de Fourier (FFT) selon des cibles prédéfinies, obtenues par analyse spectrale long terme sur des centaines de patients, avec adaptation au sexe et l'âge du patient. Autant l'extraction de formants de certaines voyelles semble évidente (Fig. 10 - /a/), autant des doutes peuvent apparaître quant à la vraisemblance de certaines détections (Fig. 10 /ou/) où certains formants tel que le F2 du /ou/ pourraient être placés à différents endroits, aucun maximum évident n'apparaissant dans la zone F2.

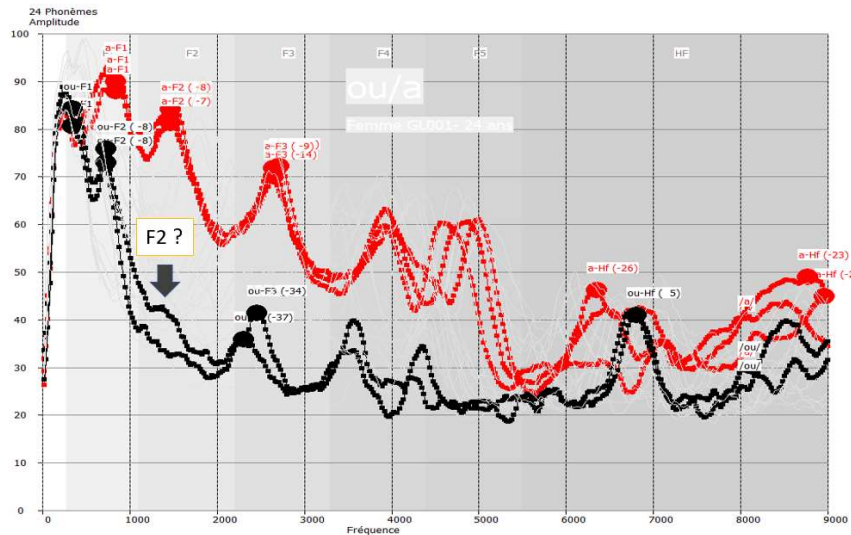


Figure 10 : Spectre moyenne de 3 /ou/ et 3 /a/ pour un même patient. L'extraction de F2 du /ou/ avec DIADOLAB donne une valeur dans la zone F1.

Les critères de sélection des frontières de segmentation varient considérablement selon les personnes et les outils de segmentation, pouvant ainsi induire des différences significatives d'estimation des fréquences des formants. Pour garantir une certaine fiabilité et reproductibilité, les mesures des formants devraient idéalement porter sur une partie stable de la voyelle, à intensité maximale, en l'absence de phénomène de co-articulation ou contaminations, conditions rarement remplies dans la pratique.

Comparaison PRAAT-DIADOLAB

Nous avons comparé les valeurs de formants obtenues avec PRAAT et DIADOLAB pour /a,i,ou/ segmentés dans de la parole lue, avec le texte « Il les perdait toutes de la même façon », pour 23 hommes et 17 femmes, dont 1/3 de témoins et 2/3 de patients atteints de dysarthrie. Nous avons converti au préalable les fréquences F en $\frac{1}{2}$ tons st en utilisant la formule suivante :

$$st = \frac{12 \times \log (F)}{\log (2)} - 80$$

(eq. 2)

La figure 11 indique globalement des valeurs similaires entre PRAAT et DIADOLAB, ce qui est plutôt rassurant. Toutefois, des différences significatives apparaissent sur le F2 du /ou/, mais aussi sur F1 du /a/ des hommes. On voit donc qu'il est difficile d'atteindre une cohérence de valeurs élevée entre des outils différents, la « réalité terrain » pouvant par ailleurs être encore différente, comme décrit dans [Skarnitzl 2015]. De plus certains patients ont une parole altérée ou pathologique avec une pauvreté harmonique marquée, ce qui complique grandement la recherche de formants, et mettre en défaut les algorithmes de recherche de zones de résonances, quand celles-ci sont noyées dans le bruit.

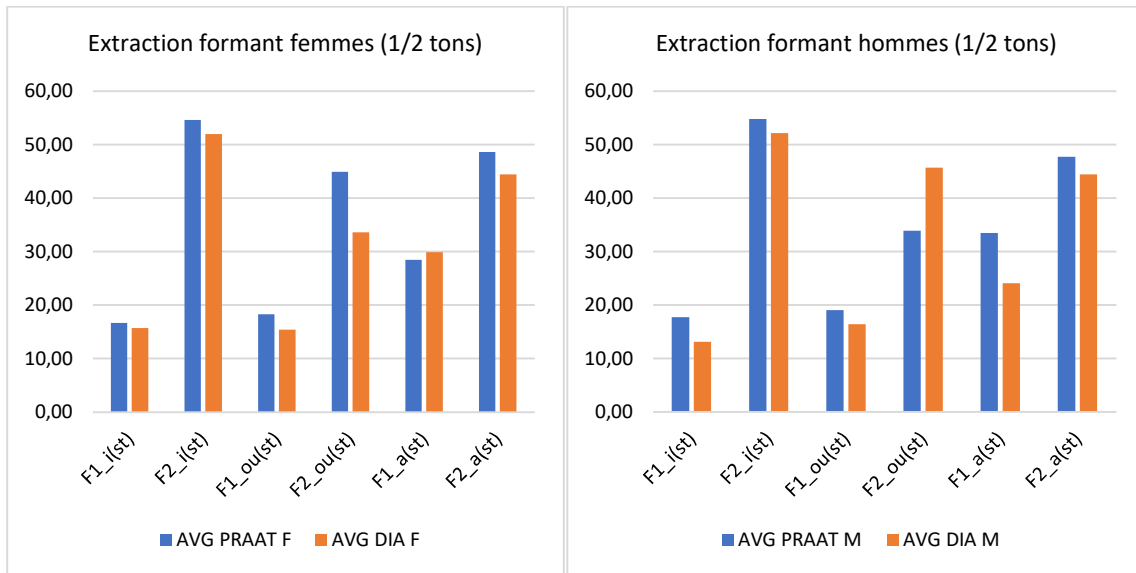


Figure 11 : Comparaison des valeurs de formants en ½ tons pour 23 hommes et 17 femmes entre PRAAT et DIADOLAB pour /a,i,ou/ en contexte de texte lu.

Exemple d'études comparatives sur le triangle vocalique

De très nombreuses études utilisent le triangle vocalique comme illustration synthétique de différentes tendances. Nous avons sélectionné des publications en archive ouverte qui illustrent différents exemples d'analyses, soit en variant le type de parole, soit en comparant des groupes de caractéristiques différentes. Trois cas concernent la parole altérée et pathologie : bégaiement, parkinson et différents types de dysarthrie.

Variation de débit de parole

Dans [Gendrot 2005] un grand nombre de voyelles orales extraites de 2 heures de discours par 15 hommes et 15 femmes sont analysées. On constate une réduction de l'aire du triangle vocalique avec la réduction de la durée de phonation, principalement liés aux phénomènes de co-articulation.

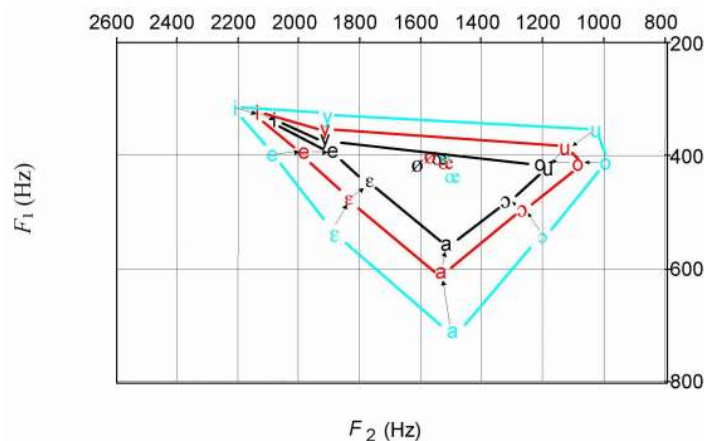


Figure 12 : durée de phonation [30 – 50ms] : noir, [60 - 80] : rouge, [90 - 110] : bleu [Gendrot 2005]

Analyse du bégaiement

Dans [Hirsch 2008], différentes populations *control group* (CG), *stutterers* (NTS), *treated stutterers* (TS) sont analysés en parole à rythme normal (Fig. 13 à gauche) ou à rythme élevé (Figure 13 à droite). On constate que les patients qui bégaiement, sans traitement, ont une aire vocalique réduite.

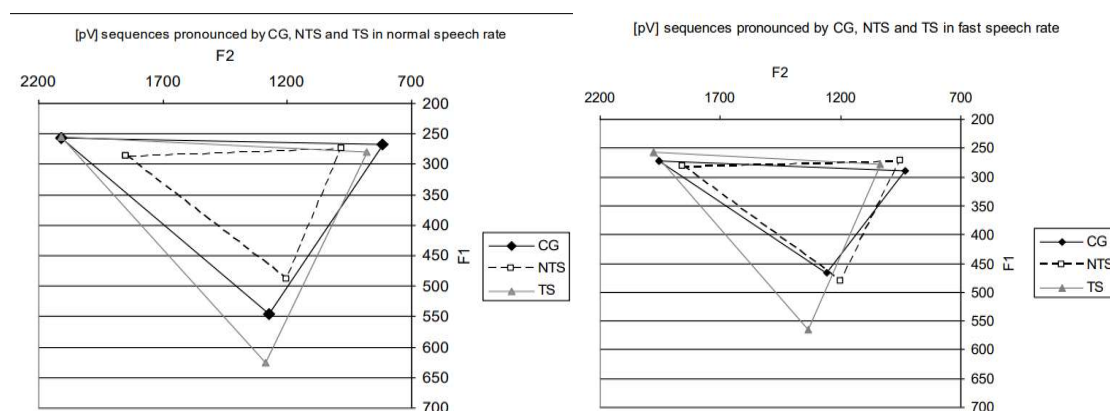


Figure 13 : Comparaison entre témoins (CG), patients qui bégaiement avec suivi (TS) ou sans suivi (NTS), pour la parole normale (à gauche) et parole rapide (à droite).

Lecture d'une histoire à un enfant

Lorsque les adultes parlent aux nourrissons, ils changent radicalement de structure prosodique et acoustique de la parole [Audibert 2018]. Les formants de 14 femmes lisant une histoire à leur enfant ou à l'expérimentateur ont montré que l'espace vocalique ne différait pas selon la situation, mais que la dispersion, c'est-à-dire la variabilité des formants, était plus élevée en présence de l'enfant (Fig. 14).

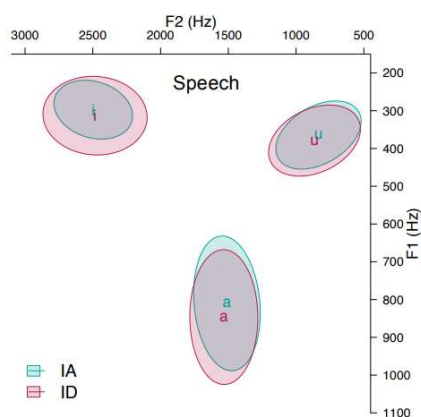


Figure 14 : Lecture d'une histoire infant-directed (ID) ou adult-directed (IA). La plus forte dispersion est pour ID, sans que les valeurs moyennes ne soit significativement modifiées [Audibert 2018]

Comparaison selon les régions

Dans [Boula 2007], le triangle vocalique est utilisé pour comparer la parole de plus de 100 personnes, la moitié séjournant dans le Nord de la France, l'autre moitié dans le Sud. On peut constater une différence significative sur les valeurs des formants du /a/ et le /o/ (Fig. 15).

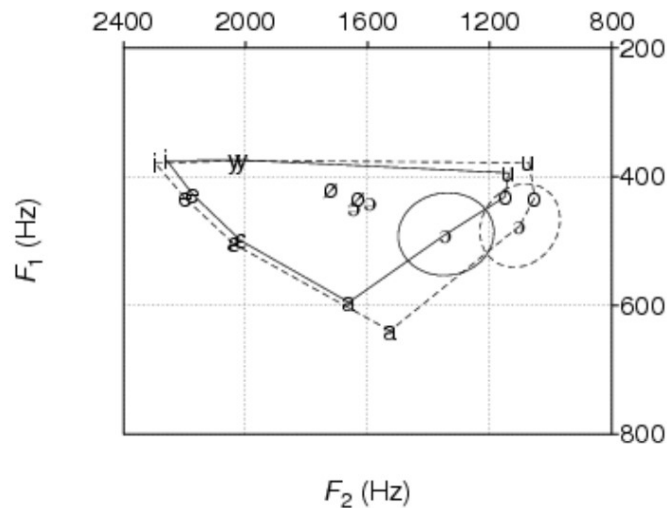


Figure 15 : Comparaison des triangles de personnes originaires du Nord et du Sud de la France [Boula 2007]

Discours hyper et hypo-articulé

L'analyse de la parole neutre, hypo et hyper-articulée a été conduite par [Picart 2010] pour un homme prononçant 1359 phrases selon les 3 modes, en observant notamment les modifications du triangle vocalique, ainsi que les ellipses de dispersion (Fig. 16). On peut constater une réduction significative de l'espace vocalique quand la parole devient moins articulée. Les trajectoires articulatoires sont moins marquées et les valeurs de F1 et F2 sont bien moins séparés dans l'espace vocalique, ce qui peut être relié à la faible intelligibilité de la parole hypo-articulée.

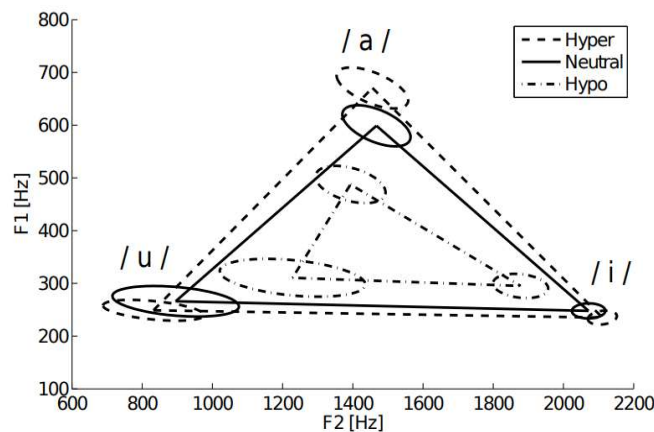


Figure 16 : Comparaison entre parole neutre, hypo et hyper-articulée par un homme prononçant 1359 phrases montrant des modifications importantes de l'aire vocalique [Picart 2010]

Variabilité

Sur la base de phrases lues par 64 femmes et 47 hommes francophones, l'extraction des formants par la méthode LPC avec le logiciel PRAAT a été conduite par [Kahn 2011] sur plus de 300.000 voyelles. L'analyse de la dispersion des voyelles entre les orateurs a été étudiée, permettant d'extraire pour les femmes des valeurs reportées figure 17. Les ellipses représentent la dispersion de valeurs entre les 64 femmes, qui est remarquablement faible. On note cependant un écart sur le F2 du /i/, pour lequel la détection automatique est plus sujette aux erreurs que pour les autres voyelles orales.

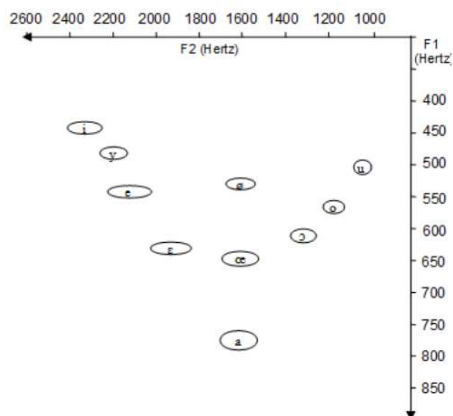


Figure 17 : Analyse de la dispersion des valeurs de F1-F2 pour 11 voyelles par analyse de phrases lues par 64 femmes, représentant un total de 150,000 phonèmes [Kahn 2011]

Maladie de Parkinson

Une comparaison entre la structure du triangle vocalique pour 9 patients atteints de la Maladie de Parkinson avec des témoins est proposée par [Roland 2017], sur la base de 3 voyelles /a :, i :, ou:/ tenues. On peut constater que l'espace articulatoire est beaucoup plus variable chez les parkinsoniens que chez les sujets sains (Fig. 16).

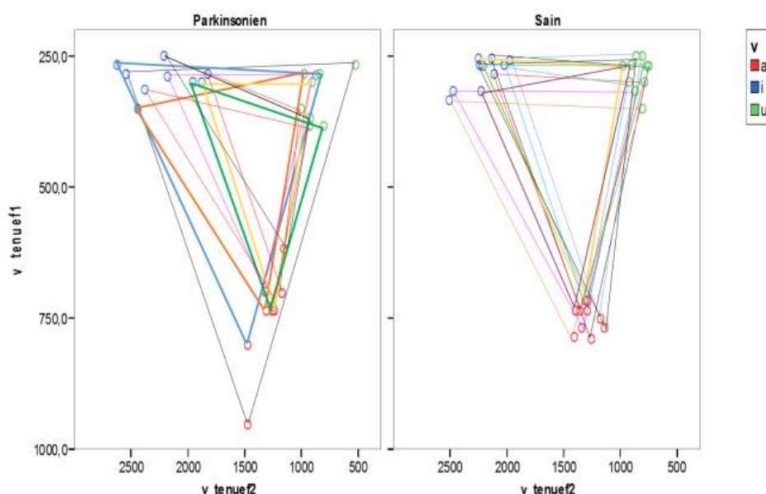


Figure 16 : Comparaison de triangle vocalique (voyelles /a :, i :, ou:/ tenues) pour des patients Parkinsoniens et sains [Roland 2017].

Dysarthries

L'étude du triangle vocalique a été conduite par [Audibert 2012] pour 22 patients présentant une dysarthrie ataxique (GrCereb), 30 patients souffrant de la maladie de Parkinson (GrPark), 26 patients atteint de Sclérose Latérale (GrSLA), et 26 témoins couvrant la distribution d'âge des patients (GrTem). Le groupe GrPark se distingue des témoins et des autres groupes dysarthriques par une réduction de F1 et un décalage vers le haut (Figure 17). On constate aussi une réduction de F2 pour les patients du groupe cérébelleux (décalage vers la droite).

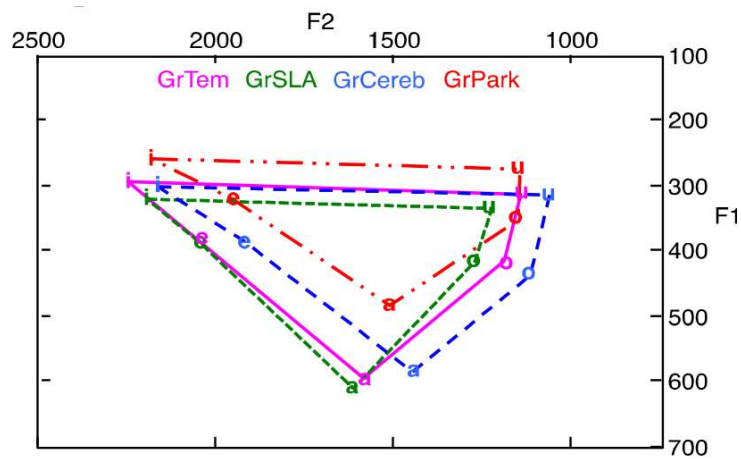


Figure 17 : Comparaison de triangles vocaliques extraits de 70 patients dysarthriques et 26 témoins [Audibert 2012]

Aire du triangle vocalique

Une fois les valeurs de formants extraites et positionnées dans l'espace, l'étape suivante est de trouver une métrique simple permettant d'extraire des tendances, par exemple entre la parole normale et la parole pathologique, ou encore l'évolution du triangle vocalique avec l'âge de l'enfant.

Métriques principales

Il existe différentes mesures concernant la taille et la géométrie de l'espace acoustique. Dans [Kent 2018], la section 10 répertorie les mesures les plus couramment utilisées dans la littérature scientifique. On peut considérer 3 distances entre les valeurs moyennes ou instantanées de chaque couple F1-F2, et par voie de conséquence, le périmètre et l'aire du triangle vocalique, comme illustré figure 18.



Figure 18 : Définition de métriques de distance, de dispersion et d'aire du triangle vocalique

Aire du triangle vocalique

Dans DIADOLAB, nous utilisons une métrique inspirée du *Triangular Vowel Space Area* (tVSA or VSA-3) proposé par [Fox 2008], à l'aide de la formule de Heron. Les valeurs sont exprimées en ½ tons et non en Hz, pour limiter l'influence du genre et de l'âge de l'enfant.

$$T_{VSA} = \sqrt{s(s-a)(s-b)(s-c)}$$

Eq. 3

$$s = \frac{a+b+c}{2}$$

Eq. 4

Où a, b, c sont les côtés du triangle en ½ ton. TVSA est donc exprimé en ½ ton carré. La formule de passage d'une fréquence F en Hz en valeur st exprimée en ½ tons est celle de l'équation 2. Quelques valeurs de référence sont données pour les hommes, femmes et enfants, d'après l'exploitation de différents corpus de voyelles tenues [Lefol 2018][Papazian 2019] ainsi que d'un corpus spécifique de voix d'enfants de 5 ans [Remacle 2017][Sicard 2019]. Les aires vocaliques sont entre 120 et 145 ½ ton carrés (Figure 19).



Figure 19 : Aire du triangle vocalique pour homme, femme et enfant

Le triangle vocalique tend à se rétrécir avec l'âge de l'enfant et tendre vers le triangle vocalique d'une femme (avant la mue du garçon), comme le montrent les triangles vocaliques d'enfants de 7 et 9 ans (Figure 20). La différenciation fille/garçon du point de vue du triangle vocalique commence à apparaître vers 11-12 ans [Kent 2018].

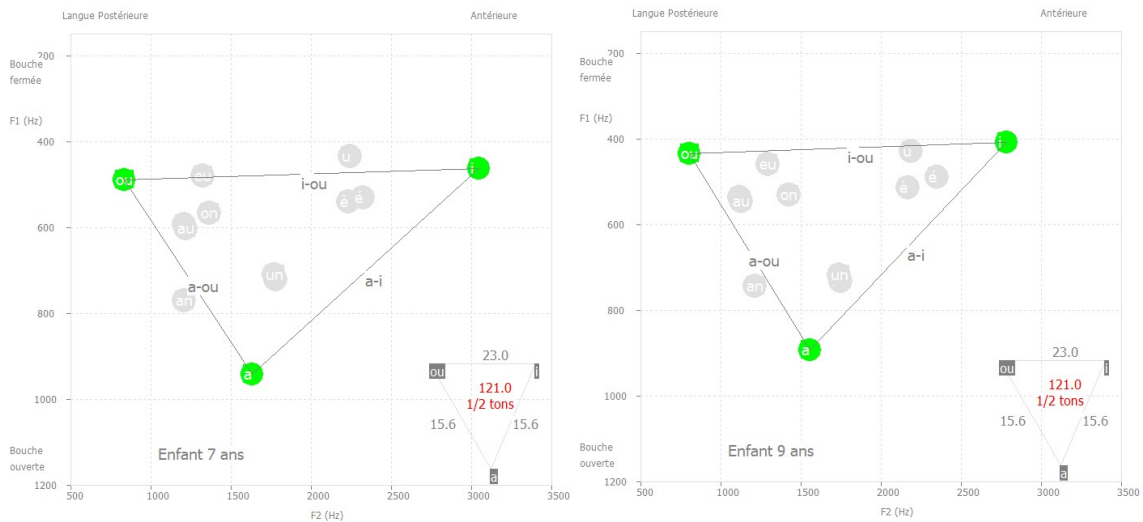


Figure 20 : Aire du triangle vocalique d'enfants de 7 et 9 ans

Extension aux consonnes

Principe

Le triangle vocalique peut être complété par les consonnes, en rapprochant les phonèmes ayant des positions d'articulateurs similaires, notamment des points de contact langue palais comme illustré Figure 21.

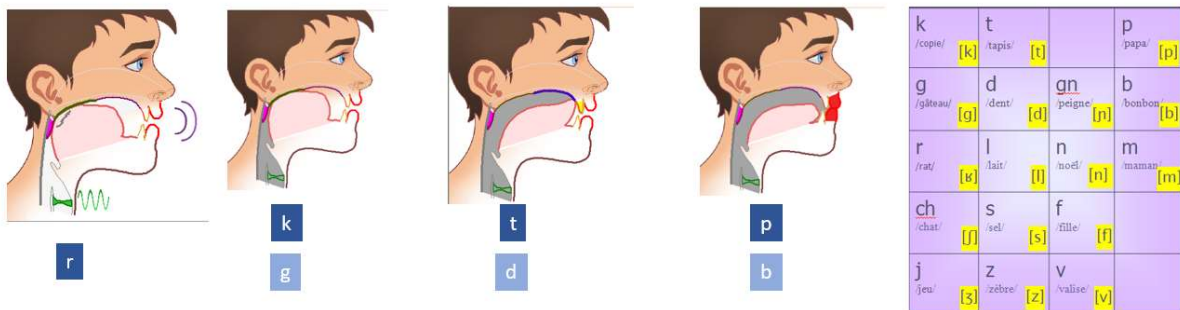


Figure 21 : Position des articulateurs de certaines consonnes

On peut alors représenter les consonnes et voyelles sur un graphique similaire concernant l'axe horizontal, qui correspond à la position de la langue. Nous séparons les consonnes voisées et non voisées. Dans la figure 22 apparaissent les consonnes et voyelles incluses dans la phrase « *Bonjour Monsieur Tralipau* » [Sicard 2019], les consonnes et voyelles non couvertes apparaissant en gris.

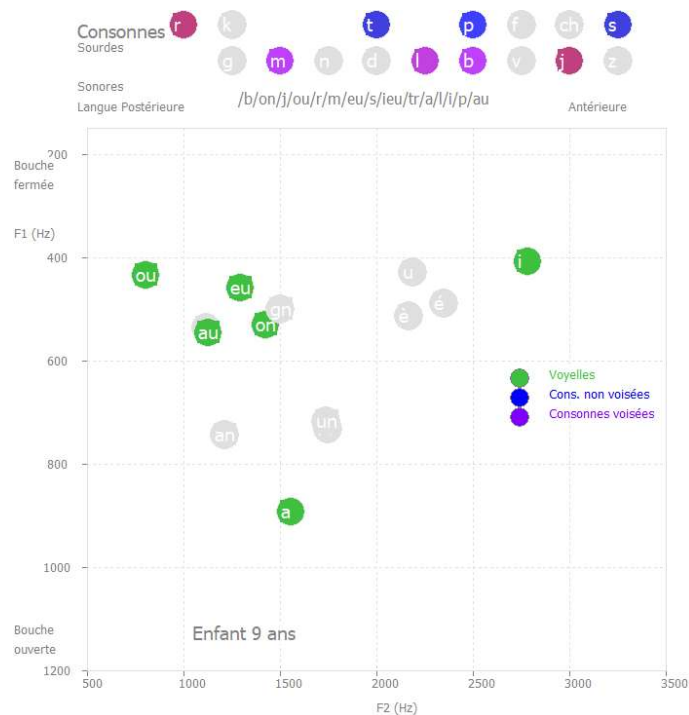


Figure 22 : Ajout des consonnes au-dessus du triangle vocalique, selon la position postérieure ou antérieure de la langue

Application à l'analyse de la parole

Dans le contexte de l'évaluation et la prise en charge orthophonique de la parole, nous pouvons utiliser la représentation du triangle vocalique pour :

- Calculer l'aire vocalique du patient, à condition que la phrase segmentée comporte bien /a,i,ou/ ce qui est le cas de phrases telles que « Bonjour Monsieur Tralipau » ou « Il les perdait toutes de la même façon ».
- Visualiser avec des couleurs spécifiques les erreurs sur des phonèmes (en rouge, erreur sur l, p, s, on), en orange imprécision du r, k, au et i.
- Identifier les zones à problème, comme par exemple dans le cas de la figure 23 l'omission de consonnes avec langue en position antérieure (l, p, s).

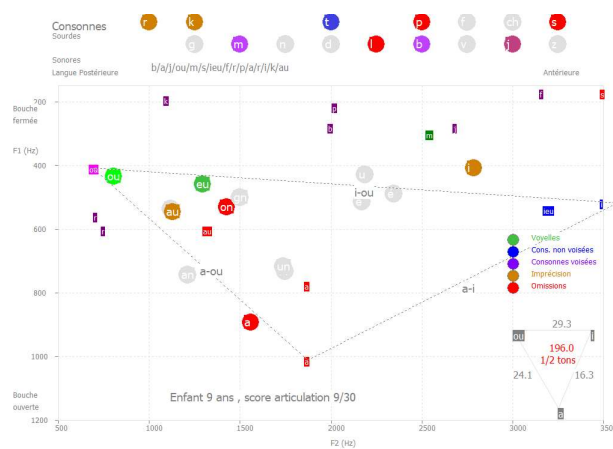


Figure 23 : Analyse d'un échantillon de parole

Progrès en rééducation orthophonique

Le triangle vocalique est utilisé dans le module *Evolution* de DIADOLAB [Diadolab 2019] pour visualiser les progrès en articulation suite à la prise en charge du trouble de la parole. On constate une amélioration importante du score d'articulation (13 points sur 30) grâce à une bonne prononciation des voyelles l,p,s, bien que des problèmes persistent sur certains phonèmes (Figure 24).



Figure 24 : Evolution des progrès du patient dans le cadre de la prise en charge orthophonique d'un enfant de 5 ans avec retard de parole [Diadolab 2019]

Statistiques à l'aide du triangle vocalique

Le calcul des formants pour un nombre important de cas de parole permet d'extraire des données statistiques telles que les valeurs moyennes et les dispersions de valeur. Sur la base du corpus de parole d'enfants de 5 ans de l'Université Catholique de Louvain [Remacle 2017], nous avons étudié le /a ; i ; ou:/ en contexte isolé pour 25 filles et 26 garçons et produit les valeurs reportées à la figure 25. Chaque échantillon est placé sur l'espace F1-F2, la moyenne et la dispersion apparaissent pour les 3 voyelles. Bien que les enfants soit normo-phoniques et tous âgés de 5 ans, la dispersion de valeurs est assez notable sur le /a/.

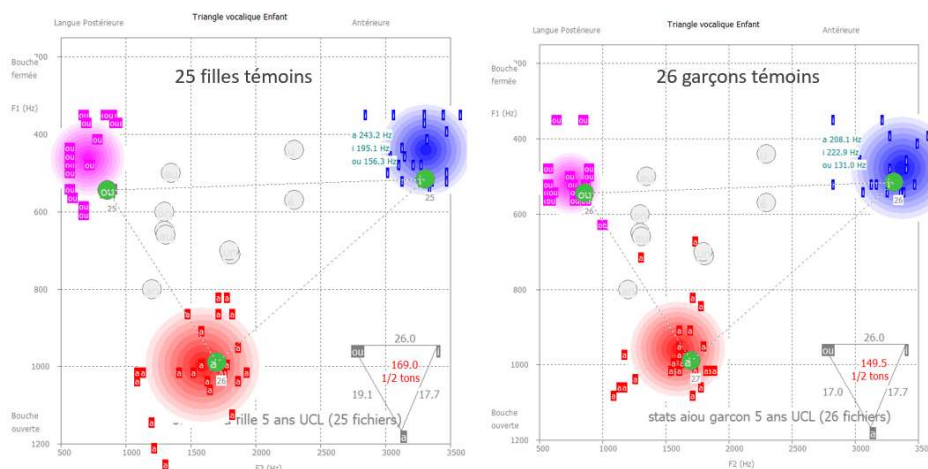


Figure 25 : Triangle vocalique d'enfants normo-phoniques de 5 ans (Corpus UCL [Remacle 2017])

Conclusion

Nous avons décrit dans cet article le principe du triangle vocalique et illustré son utilisation dans différentes situations, dont certaines au cœur des préoccupations des orthophonistes, telles que la parole de l'enfant, les troubles de la fluence ou encore la parole dysarthrique. Nous avons pu observer une certaine variabilité dans les résultats publiés. Nous avons tenté d'expliquer l'origine de ces variabilités, en mettant en avant la grande diversité des protocoles d'enregistrement et les contextes d'études très divers. Nous avons aussi comparé deux approches logicielles pour identifier les similitudes et les biais. Parmi les différentes métriques existantes, nous avons choisi le calcul de l'aire vocalique sur la base de distances en ½ tons, et avons souhaité étendre le concept du triangle vocalique en y associant les consonnes, qui permet d'avoir un panorama synthétique des troubles du patient en vue d'illustrer de manière claire les progrès de la prise en charge orthophonique.

Références

- [Assmann 2008] Assmann, P. F., Nearey, T. M., & Bharadwaj, S. (2008). Analysis and classification of a vowel database. *Canadian Acoustics*, 36, 148–149
- [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).
- [Audibert 2012] Audibert, N., & Fougeron, C. (2012, June). Distorsions de l'espace vocalique: quelles mesures? Application à la dysarthrie. In *Proceedings of the Joint Conference JEP-TALN-RECITAL 2012*, volume 1: JEP (pp. 217-224).
- [Baken 2000] Baken, R. J., & Orlikoff, R. F. (2000). *Clinical measurement of speech and voice*. Cengage Learning.
- [Bardelang 2020] Bardelang T. (2020), L'apport d'une phrase supplémentaire au protocole de DIADOLAB3. Mémoire d'orthophonie. Université de Lyon.
- [Boula 2007] Boula de Mareüil, P., Adda-Decker, M., & Woehrling, C. (2007). Analysis of oral and nasal vowel realisation in northern and southern French varieties. In *16th International Congress of Phonetic Sciences* (pp. 2221-2224).
- [Byrne 1994] Byrne, et al. (1994). An international comparison of long-term average speech spectra. *The journal of the acoustical society of America*, 96(4), 2108-2120.
- [Diadolab 2019] Menin-Sicard, A., Sicard, E. (2019) DIADOLAB 3 - Logiciel d'évaluation et de rééducation de la parole - Manuel d'utilisation. 2019. (hal-02073314)
- [Deterding 1997] Deterding, D. (1997). The Formants of Monophthong Vowels in Standard Southern British English Pronunciation. *Journal of the International Phonetic Association*. 27. 47 - 55
- [Eichhorn 2018] Eichhorn, et al (2018) Effects of Aging on Fundamental Frequency and Vowel Formants, *Journal of Voice* Volume 32, Issue 5, September 2018, Pages 644.e1-644.e9

- [Fant 1960] Fant G. (1960). Acoustic theory of speech production. Mouton, The Hague.
- [Fox 2008] Fox, R. A., & Jacewicz, E. (2008). Analysis of total vowel space areas in three regional dialects of American English. *Acoustics'08*, pp. 495–500.
- [Gendrot 2005] Gendrot, C., & Adda-Decker, M. (2005). Impact of duration on F1/F2 formant values of oral vowels: an automatic analysis of large broadcast news corpora in French and German. In *Ninth European Conference on Speech Communication and Technology*.
- [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.
- [Hillenbrand 1995] Hillenbrand, J., Getty, L., Clark, M., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America*, 97, 3099-3111.
- [Hirsch 2008] Hirsch, F., et al. (2008). Formant structures of vowels produced by stutterers in normal and fast speech rates. 8th International Seminar On Speech Production, Dec 2008, France.
<https://halshs.archives-ouvertes.fr/halshs-00716583/>
- [JPC 2019] Anne Menin-Sicard, Etienne Sicard. 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, May 2019, Mons, Belgique. (hal-02127039)
- [Kahn 2011] Kahn, J., Audibert, N., Bonastre, J. F., & Rossato, S. (2011, August). Inter and Intra-speaker Variability in French: An Analysis of Oral Vowels and Its Implication for Automatic Speaker Verification. In *ICPhS* (pp. 1002-1005).
- [Kent 2018] Kent, R. D., & Vorperian, H. K. (2018). Static measurements of vowel formant frequencies and bandwidths: A review. *Journal of communication disorders*, 74, 74-97.
- [Lefol 2018] Lefol, G. (2018). Vers une évaluation objective de l'articulation en pratique clinique orthophonique. Mémoire de Master d'orthophonie. Université de Tours, Juin 2018.
- [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.
- [Lemaitre 2020] emaitre M. (2020), Élaboration d'une échelle d'évaluation vocale des adultes implantés cochléaires. Mémoire de Master d'orthophonie. Univ. de Tours.
- [Meunier 2007] Meunier C. (2007) Phonétique acoustique : Phonétique acoustique. Auzou P. Les dysarthries, Solal, pp.164-173, 2007. (hal-00250272)
- [Nicolaidis 2016] Nicolaidis K., Sfakianaki A. (2016) Acoustic characteristics of vowels produced by Greek intelligible speakers with profound hearing impairment I: Examination of vowel space, *International Journal of Speech-Language Pathology*, 18:4, 378-387
- [Papazian 2019] Papazian A. (2019). Analyse acoustique de la production des consonnes et voyelles dans l'ataxie de Friedreich. Mémoire de Master d'orthophonie. Université de Tours, Juin 2019.

- [Praat] Boersma, Paul & Weenink, David (2019). Praat: doing phonetics by computer [Computer program]. Version 6.1.06, retrieved 8 November 2019 from <http://www.praat.org/>
- [Picart 2010] Picart, B., Drugman, T., & Dutoit, T. (2010). Analysis and synthesis of hypo- and hyperarticulated speech. In Seventh ISCA Workshop on Speech Synthesis.
- [Pruthi 2007] Pruthi, T., Espy-Wilson, C. Y., & Story, B. H. (2007). Simulation and analysis of nasalized vowels based on magnetic resonance imaging data. *The Journal of the Acoustical Society of America*, 121(6), 3858-3873.
- [Roland 2017] Roland, V; et al (2017), Dynamique phonétique et contrôle moteur dans la maladie de Parkinson: analyse du contrôle de la production des glides, JPC 2017 Paris
- [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.
- [Sicard 2019] Sicard E., Meyrieux L., Moreau M., Remacle A. (2019). L'analyse acoustique des voix d'enfants de 5 ans : Proposition de valeurs de référence pour les logiciels PRAAT et VOCALAB. Journées de Phonétique Clinique 2019, May 2019, Mons, Belgique. (hal-02434863)
- [Stoel-Gammon 2010] Stoel-Gammon, C. (2010). The Word Complexity Measure: Description and application to developmental phonology and disorders. *Clinical linguistics & phonetics*, 24(4-5), 271-282.
- [Skarnitzl 2015] Skarnitzl, Radek & Vaňková, Jitka & Boril, Tomas. (2015). Optimizing the Extraction of Vowel Formants. In book: Tackling the Complexity in Speech, Chapter: 9, Publisher: Faculty of Arts, Charles University, pp.165-182
- [Vilain 2015] Vilain, C., Berthommier, F., & Boë, L. J. (2015, September). A brief history of articulatory-acoustic vowel representation. <https://hal.archives-ouvertes.fr/hal-01197460>
- [Woehrling 2007] Woehrling, C., Boula de Mareüil, P. (2007). Comparing Praat and Snack formant measurements on two large corpora of northern and southern French. In Eighth Annual Conference of the International Speech Communication Association.