

**Développement et validation d'outils de mesures de la
qualité de la voix dans le logiciel VOCALAB**

Etienne Sicard*, Stéphanie Perrière**, Anne Menin-Sicard***

* professeur, INSA, Toulouse
etienne.sicard@insa-toulouse.fr

** orthophoniste, Institut de la Face et du Cou, Nice
stephanie.perriere@gmail.com

*** orthophoniste, Toulouse
anne.sicard2@orange.fr

Résumé :

Dans cet article, nous décrivons la méthodologie suivie pour sélectionner et valider un ensemble limité d'indicateurs de pathologie de la voix, accessibles aux orthophonistes, permettant une évaluation objective des progrès en rééducation. Nous décrivons leur prototypage dans le logiciel VOCALAB ainsi que le développement d'un outil statistique. Nous détaillons cinq indicateurs, justifions les choix par rapport à l'état de l'art et décrivons notre démarche de validation sur des voix normales et pathologiques. La construction d'une base de données de cas de voix pathologiques est décrite, de même que l'analyse statistique des résultats en comparaison avec l'outil PRAAT. Les indicateurs se révèlent pertinents et fiables, relativement peu sensibles au mode opératoire et assez tolérants à la variabilité de qualité des conditions d'enregistrement.

Mots clés : rééducation de la voix, analyse objective, auto-évaluation, bilan, base de données, marqueurs de pathologie.

Development and validation of tools for measuring the quality of the voice using the software VOCALAB

Summary:

In this paper, we describe the methodology used to select and validate a limited set of indicators of pathology of voice, allowing an objective assessment of progress in rehabilitation. We describe the prototyping in VOCALAB software and the development of a statistical tool. We focus on a selection of five indicators, justify the choice compared to the state of the art, and describe our validation approach on normal and pathological voices. The steps to build a database of pathological voices is described, as well as statistical analysis of the results in comparison with PRAAT tool. The indicators reveal relevant and reliable, relatively insensitive to the operating and fairly tolerant to variable voice recording conditions.

Key words : voice pathology, objective analysis, auto-evaluation, assessment, data-base, pathology indicators.

----- INTRODUCTION -----

Pour gagner en efficacité dans la prise en charge des pathologies de la voix, les orthophonistes se retrouvent limités par les méthodes d'évaluation perceptives, du fait de leur variabilité, leur manque de fiabilité et de crédibilité vis-à-vis du patient. Disposer d'une plate-forme d'outils simples, fiables et normalisés leur permet de mesurer la qualité de la voix, autant en phase de bilan pour les guider dans leurs choix thérapeutiques, qu'en phase de rééducation pour donner des repères objectifs immédiats sur les progrès accomplis.

L'analyse objective à l'aide d'outils spécialisés tels qu'EVA (Ghio et al., 2008) et MDVP (Deliyski, 2012) a longtemps été réservée à des orthophonistes experts, souvent dans le cadre hospitalier. L'outil gratuit PRAAT (Boersma, Weenink, 2009) est pertinent pour des recherches et calculs statistiques, mais se révèle moins adaptable à la clinique orthophonique. Son principal atout est en effet son langage de programmation très puissant, cependant peu accessible aux orthophonistes sans une sensibilisation à la programmation et au traitement du signal.

D'importants développements ont été conduits autour du logiciel VOCALAB (Sicard, Menin-Sicard, 2013-a) afin de prototyper un ensemble réduit de marqueurs de qualité de la voix. Dans le cadre d'une démarche de recherche au laboratoire LURCO de l'UNADREO, une étude bibliographique a été menée sur les différents outils et indicateurs de qualité de la voix, en tentant de classer les marqueurs utilisés par grandes familles, afin de faire ressortir les points communs et les différences (Sicard, Menin-Sicard, 2013-b). Cette étude nous a guidés dans le choix d'indicateurs pertinents et leur implémentation prototype.

Le but de cet article est de présenter une synthèse de l'étude bibliographique, de détailler les indicateurs de VOCALAB, de présenter les étapes de construction de la base de données, enfin de présenter les résultats obtenus en terme de corrélation entre les marqueurs, et comparaison entre les logiciels. Cet article discute des résultats obtenus et ouvre les pistes de recherche future.

----- SYNTHÈSE DE L'ÉTUDE BIBLIOGRAPHIQUE -----

L'analyse d'ouvrages de référence tels que ceux de Baken et Orlikoff (2000) et Kent et Ball (2000), ainsi que d'une quarantaine d'articles portant sur des analyses d'un nombre significatif de cas de voix normales et pathologique a été conduite (Sicard, Menin-Sicard, 2013-b). Nous nous sommes intéressés aux logiciels EVA (Ghio et al., 2008), PRAAT (Boersma, Weenink, 2009), MDVP (Deliyski, 2012) et VOCALAB (Menin-Sicard, Sicard, 2009), qui sont, à notre connaissance, les plus utilisés par les orthophonistes francophones. Ces outils ont en commun, à quelques variantes près, la possibilité d'évaluer les marqueurs listés dans la figure 1.

<i>PRAAT</i>	<i>EVA</i>	<i>MDVP</i>	<i>VOCALAB</i>
<ul style="list-style-type: none"> Fondamental usuel moyen Etendue vocale Temps maximum de phonation (TMP) 	<ul style="list-style-type: none"> Fondamental usuel moyen Etendue vocale Temps maximum de phonation (TMP) 	<ul style="list-style-type: none"> Fondamental usuel moyen Etendue vocale Temps maximum de phonation (TMP) 	<ul style="list-style-type: none"> Fondamental usuel moyen Etendue vocale Temps maximum de phonation (TMP)
<ul style="list-style-type: none"> Instabilité en hauteur: Jitter, Jita, DDP, ratio, RAP, PPQ5, jDDP Instabilité en amplitude: Shim, ShdB, APQ3, APQ5, APQ11, ADDP Bruit: HNR 	<ul style="list-style-type: none"> Aire dynamique vocale Instabilité en hauteur: JittA, Jitt Factor, Jitt Ratio, vF0, RAP Instabilité en amplitude: Shim Factor, APQ Bruit: Sr, Sr>1KHz 	<ul style="list-style-type: none"> Instabilité en hauteur: PPQ, Jitt, RAP, Jita, sPPQ, STD, vf0, Instabilité en amplitude: APQ, Shim, ShdB, sAPQ, vAm Bruit: VTI, NHR, SPI, FTRI, ATRI Harmoniques: DSH, NSH 	<ul style="list-style-type: none"> Rapports s/z, a/z Instabilité en hauteur Instabilité en amplitude Rapport Signal/Bruit Pauvreté harmonique
<p><i>Marqueurs spécifiques :</i></p> <ul style="list-style-type: none"> Ruptures et irrégularités : FLUF, DVB <i>Langage de programmation très puissant permettant le calcul statistique et l'extraction de tous types de données</i> 	<p><i>Marqueurs spécifiques :</i></p> <ul style="list-style-type: none"> Débit : <i>pression, rendement, efficacité, fuite, résistance glottique</i> 	<p><i>Marqueurs spécifiques :</i></p> <ul style="list-style-type: none"> Ruptures et irrégularités : DVB, DUV, NVB, NUV 	<p><i>Marqueurs spécifiques :</i></p> <ul style="list-style-type: none"> Altération de l'attaque

Figure 1. Analyse des indicateurs de pathologie dans les systèmes EVA, MDVP, PRAAT et VOCALAB.

Les marqueurs de pathologie de la voix utilisés dans ces différents logiciels sont listés par famille, en s'inspirant des classifications proposées par Pützer (2001). On retrouve dans ces logiciels quatre grandes familles d'analyse : l'instabilité en hauteur (Jitter et dérivés), l'instabilité en amplitude (Shimmer et dérivés), l'analyse du bruit (HNR et dérivés), et l'évaluation de phénomènes transitoires parasites. Selon les logiciels, d'autres types d'analyses peuvent être conduits (en italique sur la figure) : analyse du débit d'air dans le système EVA, analyse des ruptures et irrégularités dans PRAAT et MDVP, altération de l'attaque dans VOCALAB.

L'instabilité de la hauteur de la voix. Le plus connu est le *Jitter*, qui caractérise l'instabilité de la fréquence fondamentale à court terme, selon de multiples variantes (Baken, 2000). Le coefficient de Lyapunov (Teston, 2004) correspond à une instabilité à plus long terme. C'est le même type de calcul qui a été implémenté dans VOCALAB (Sicard, Menin-Sicard, 2013-a) afin de dessiner le portrait de phase.

La figure ci-dessous donne l'exemple d'un spectrogramme (temps en X, fréquence en Y, énergie avec une palette de couleurs), avec le tracé superposé (trait noir) de la fréquence fondamentale instantanée. On note des variations importantes de F0, qui se traduisent par un portrait de phase (en bas à droite) fortement décentré du cercle central, qui indique une hauteur stable en fréquence.

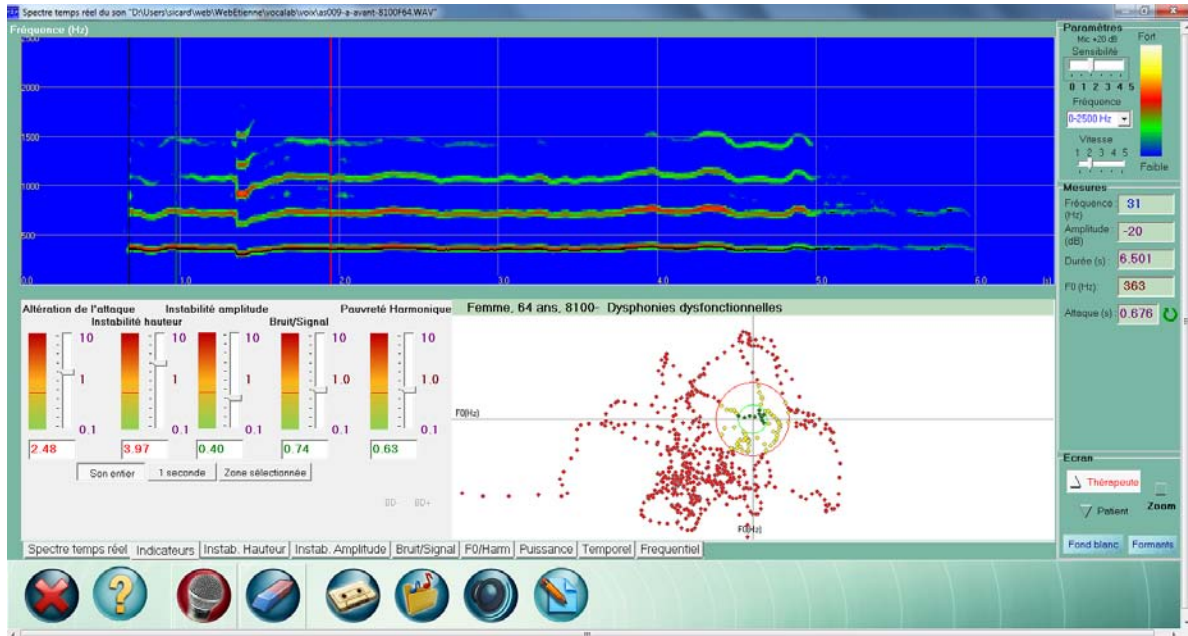


Figure 2. Instabilité en hauteur du cas AS009 de la base de données et portrait de phase associé, fortement décentré du fait des variations importantes du fondamental F0.

Tous les auteurs s'accordent sur le fait que la valeur du *Jitter* est fortement dépendante de l'algorithme utilisé pour la détection du fondamental de la voix. Au sein d'un même outil, plusieurs algorithmes peuvent être proposés, associés à plusieurs variables d'ajustement, qui requièrent de ce fait une bonne connaissance de leur utilisation selon les contextes et de leurs limites.

Dans le cas de voix éraillées, soufflées ou simplement dans des conditions d'enregistrement de basse qualité, les indicateurs varient du simple au triple selon les logiciels utilisés, comme démontré par Maryn et al. (2009). Les seuils définissant une frontière entre voix normales et pathologiques diffèrent selon les études, les protocoles et les logiciels utilisés (Kent et al., 2003). La plupart des études se focalisent sur l'analyse d'une portion de voyelle stable d'un /a/ (en général une seconde), d'autres considèrent l'ensemble de l'échantillon (Revis et al., 1999) pour inclure l'attaque. Werth et al. (2010) démontre que les différents indicateurs de Jitter sont fortement corrélés entre eux.

Selon certains auteurs tels que Parsa et Jamieson (2001) et Wolfe et al. (1995), le Jitter est peu corrélé avec le degré de sévérité des dysphonies, ce qui fait planer un doute sur la pertinence de ce marqueur dans la démarche de rééducation de la voix. Nos premiers essais d'implémentation du Jitter dans VOCALAB ont été peu convaincants. Pour s'approcher du jugement perceptif auditif, nous avons focalisé l'instabilité en hauteur de VOCALAB sur le moyen (100 ms) et le long terme (1000 ms) et non le court terme comme le font les logiciels

EVA, PRAAT et MDVP. Ce choix s'est rapidement imposé lors des phases de test, certaines voix étudiées étant très instables à l'audition malgré un Jitter faible, d'autres stables malgré un Jitter très élevé.

L'instabilité de l'amplitude de la voix. Le plus connu est le *Shimmer* qui évalue la variation d'amplitude du signal au cours du temps (voir figure ci-dessous) et se décline en différentes versions telles que le Shimmer instantané, Shimmer relatif ou moyen sur 3, 5, 11 cycles, etc.

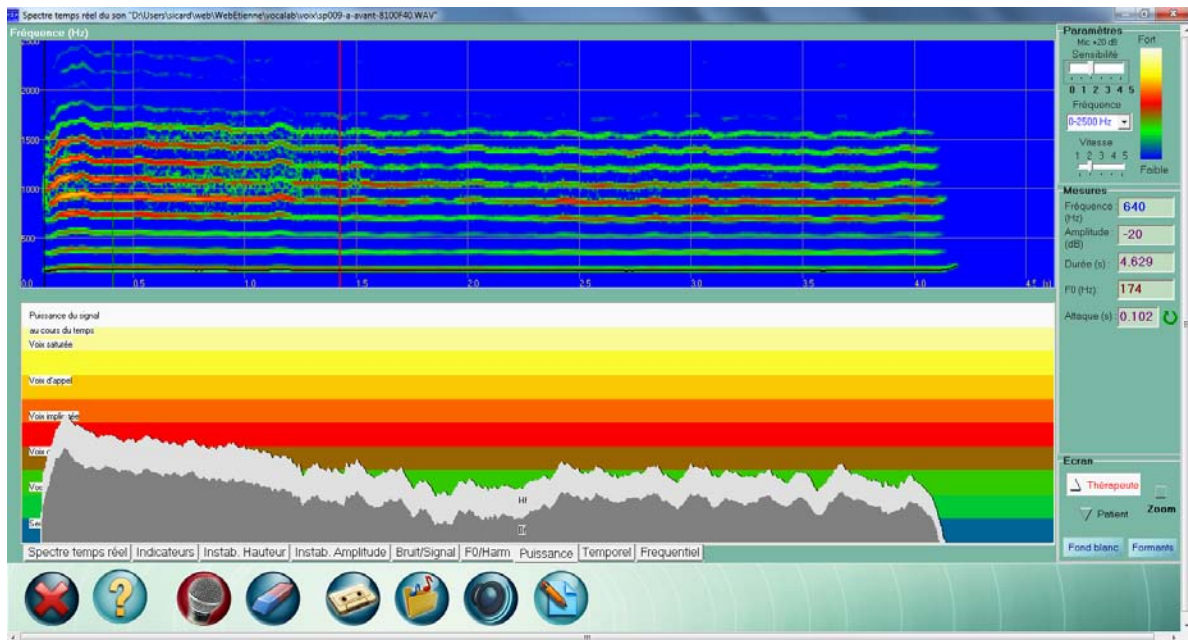


Figure 3. Illustration de la variation de l'amplitude du signal au cours du temps (écran du bas), montrant de nombreuses irrégularités dans le contrôle de puissance, qui est à la base du calcul du Shimmer (Cas SP009 de la base de données en ligne).

Les algorithmes mis en jeu dans l'évaluation du Shimmer influent considérablement sur les valeurs obtenues dans le cas de voix dysphoniques (Maryn, 2009). Les différents Shimmer sont aussi fortement corrélés entre eux (Werth, 2010) et se révèlent sensibles au bruit, que ce soit du bruit provenant de la chaîne d'enregistrement ou de la déperdition d'air liée à un accolement glottique insuffisant (Sicard, 2013-a in RO).

Dans VOCALAB, nous avons décidé de nous concentrer sur l'analyse des variations d'amplitude de l'harmonique F0, éliminant ainsi la dépendance avec le bruit. Contrairement au Shimmer de PRAAT, EVA ou MDVP, l'analyse de la stabilité en amplitude n'est pas à court terme (échelle de quelques ms), mais sur les moyen (100 ms) et long termes (1000 ms). Ce choix s'est révélé pertinent sur différents cas de voix perturbées par des bruits basses fréquences (50, 100 Hz) ou encore des enregistrements effectués avec des chaînes d'acquisition du son de qualité moyenne à faible ou encore en présence de bruits divers tels que des conversations proches (salle d'attente) et bruits d'ordinateurs.

L'évaluation du bruit. Yumoto (1982) a été l'un des premiers à proposer une technique simple de calcul du rapport Harmonique/Bruit (H/N en Anglais). Pour le groupe normal, ce

rapport est centré sur 12 dB (harmoniques significativement plus énergétiques que le bruit) et peut devenir négatif pour des pathologies sévères (harmoniques noyées dans le bruit). Yu et al. (2001) utilise dans EVA deux marqueurs d'évaluation du rapport Signal-Bruit (Sr, en %, et Sr>1KHz en %), relativement peu corrélés entre eux. Le marqueur Sr décroît de 60% à 20% avec la sévérité, Sr>1KHz de 10 à 5%, avec peu de différences Homme/Femme.

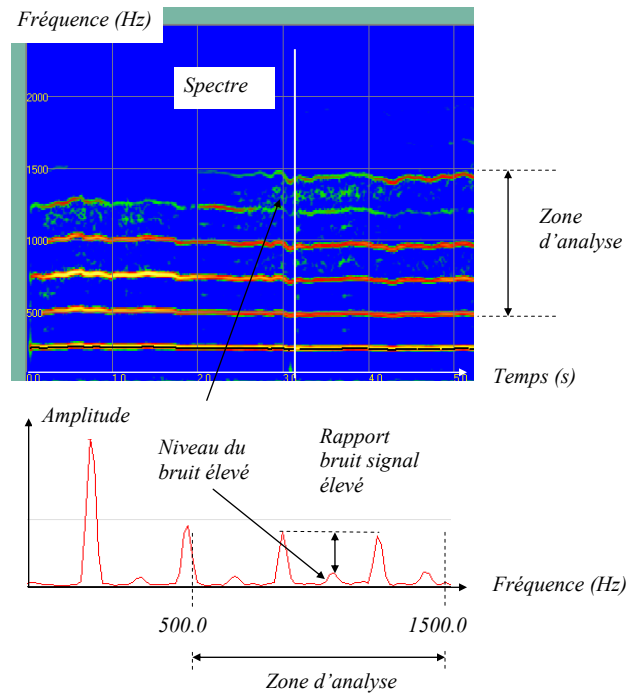


Figure 4. Détection du bruit et de son amplitude par rapport au signal.

L'outil MDVP propose un marqueur Noise-to-harmonic ratio (NHR) qui évalue le rapport entre l'énergie non-harmonique de l'intervalle 1500-4500 Hz avec l'énergie harmonique de 70 à 4500 Hz. La valeur commune de NHR est autour de 0.1, et augmente avec la sévérité des dysphonies (Xue, Deliyski, 2001). Werth et al. (2010) analysent un très grand nombre de cas de voix pathologiques à l'aide du logiciel Dr. Speech. Les valeurs typiques de NHR sont autour de 20 dB pour les cas normaux et ne baissent en moyenne que de 2 dB pour les voix pathologiques. Le facteur NNE (Normalized Noise Energy), corrélé à NHR (-0.67) passe de -8 dB à -6dB pour ces mêmes cas. En résumé, on constate une certaine variabilité de conditions de calculs, de normes et de résultats selon les outils utilisés.

VOCALAB utilise une mesure proche du NHR de MDVP. Le marqueur est sensible au bruit ambiant, le logiciel n'étant pas capable de dissocier le bruit de la voix de celui du lieu d'enregistrement. VOCALAB détecte relativement bien les traces de souffle (Figure 5, cas SP021), qui apparaissent comme des traces aléatoires, sans lien avec les harmoniques. Les éraillures (Figure 5, cas SP093) sont des traces entre les harmoniques, d'une durée souvent courte et d'amplitude instable que VOCALAB considère aussi comme du bruit.

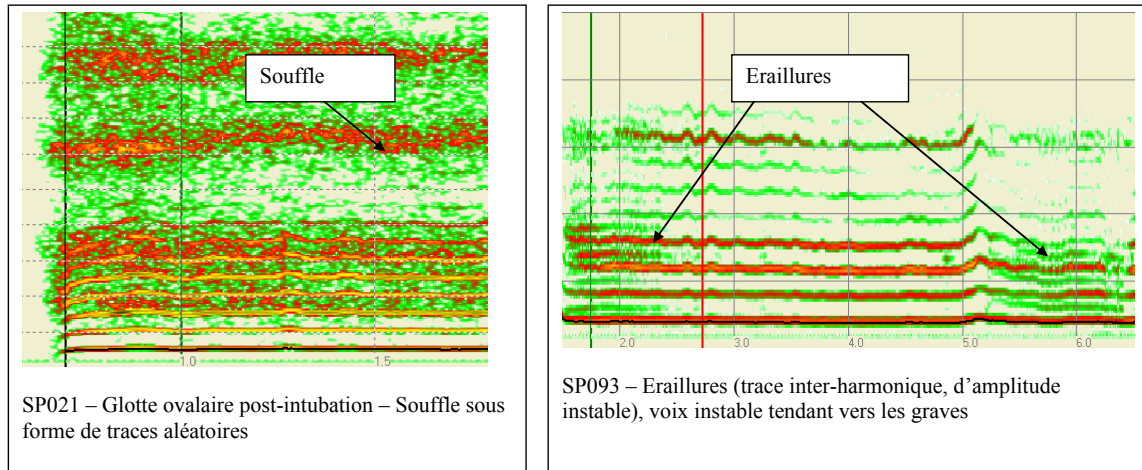


Figure 5. Illustration de deux types de bruits : souffle à gauche et éraillures à droite.

L'évaluation des phénomènes transitoires parasites. Les paramètres DUV (Degree of Voiceless /%/) et DVB (Degree of Voice Breaks /%/) de MDVP mesurent l'incapacité de la voix à soutenir le voisement de manière ininterrompue. Le seuil normatif est 0, car une voyelle en voix normale ne devrait pas avoir de segment sans voix.

Concernant la qualité des transitions lors de l'attaque (mise en vibration) et de la fin de voyelle (désonorisation), les publications s'accordent sur l'importance des informations qu'elles contiennent (Orlikoff et al., 2009 ; Revis et al., 1999). Toutefois, il existe peu de marqueurs validés sur cet aspect. VOCALAB propose une évaluation de l'altération de l'attaque, basée sur la combinaison de l'écart par rapport à une montée en puissance idéale, l'instabilité en hauteur durant la phase d'attaque et enfin le bruit entre les harmoniques du fondamental. Ce marqueur, motivé par l'importance de l'information d'attaque pour l'analyse des voix (Revis et al., 1999), n'existe à notre connaissance dans aucun autre logiciel.

----- CONSTRUCTION D'UNE BASE DE DONNEES -----

Dans le cadre de l'ERU15 du laboratoire LURCO (<http://lurco.unadreo.org>) une de nos actions importantes a été de regrouper environ 200 cas de voix, principalement sur une voyelle /a/ en voix conversationnelle, d'une durée de quelques secondes, dont 150 environ pathologiques, avant et après rééducation orthophonique. Les orthophonistes ont effectué un enregistrement de /a/ et de sirène en bilan, puis en fin de rééducation. Les spectres des voix ont été analysés et commentés, complétés par une description des axes thérapeutiques et des résultats fournis par les indicateurs de VOCALAB et PRAAT.

Les enregistrements de voix ont été mis à disposition sur le serveur de l'INSA de Toulouse (Sicard, Menin-Sicard, 2012) (et à terme sur le serveur du LURCO) à des fins de recherche orthophonique, en conformité avec la loi «Informatique et Libertés» pour protéger la confidentialité sur l'identité des patients. Les voix mises en ligne ont été enregistrées selon des consignes précises, en conditions réelles de cabinet d'orthophonie, avec des microphones de qualités diverses et différents problèmes associés : bruit 50 Hz, sifflements, bruit d'ordinateur,

coupure basse ou haute fréquence, parasites divers. Ces conditions très variées mettent parfois en défaut les logiciels et peuvent significativement altérer la pertinence des calculs, notamment pour des pathologies qui dégradent considérablement la qualité de la voix.

Les noms de fichiers sont codés selon une terminologie précise, afin de faciliter leur exploitation. Les noms de fichier comportent des informations de type <orthophoniste>, <numéro>, <type de son>, <avant/pendant/après rééducation>, <sévérité>, <nomenclature>, <genre>, <âge>. La sévérité est un indicateur du grade de pathologie selon les repères VHI (Jacobson et al., 1997) (40VHI = 40 sur l'échelle Voice-Handicap-Index), auto-évaluation VOCALAB (12AE = 12 sur l'échelle d'auto-évaluation) (Sicard, Menin-Sicard, 2013-a) ou GRBAS (3GR = 3 sur l'échelle Grade, Rough, Breathy, Asthenic, Strained) (Hirano, 1981).

Les pathologies associées à chaque fichier sont décrites au moyen d'une nomenclature inspirée de l'ouvrage de Verdolini et al. (2006). Par exemple "as001-a-avant-12AE7120F44.WAV" fait référence à l'enregistrement par Anne Menin-Sicard du patient 001, pour un son /a/, avant rééducation. L'auto-évaluation est chiffrée à 12 (12AE), le patient souffrant de la pathologie 7120 (paralysie unilatérale du nerf laryngé). Il s'agit d'une femme de 44 ans (F44).

L'exploitation de la base de données de voix par les logiciels PRAAT et VOCALAB, ainsi que la fusion des données et le calcul statistique sont décrits à la figure 6.

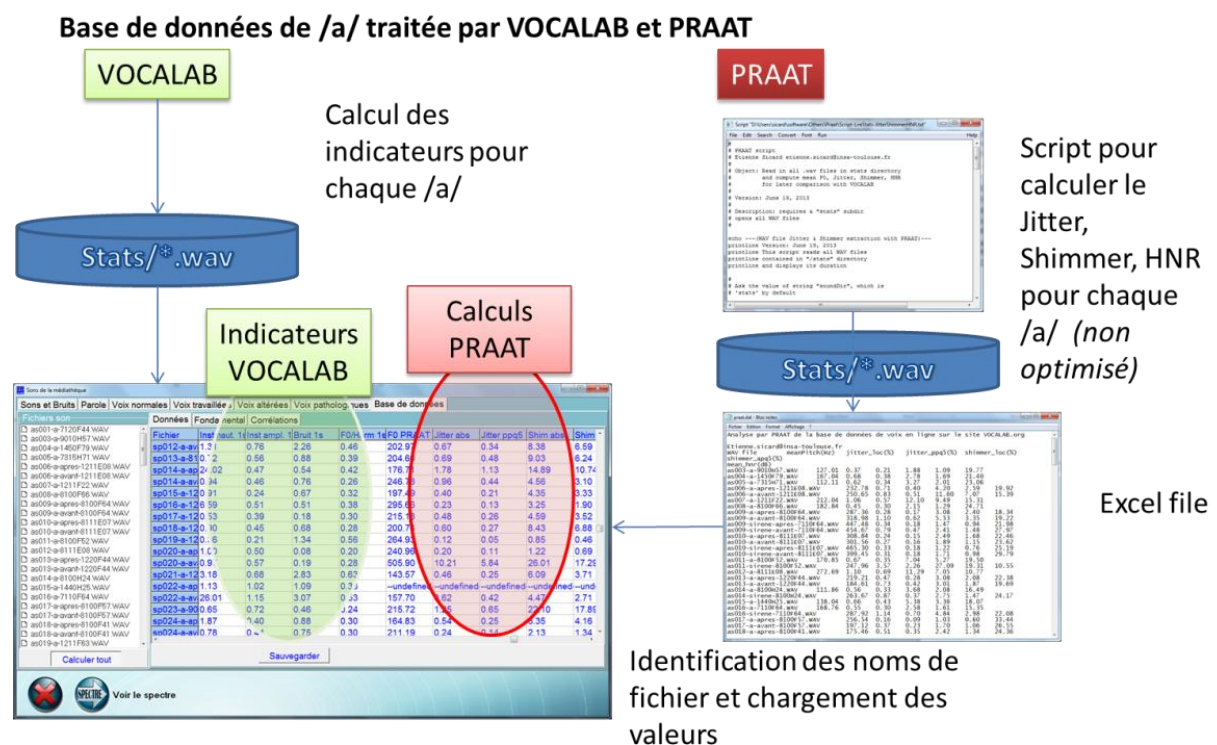


Figure 6. Principe d'exploitation de la base de données de voix avec les logiciels PRAAT et VOCALAB à des fins de comparaisons statistiques.

L'outil de calcul statistique de VOCALAB est décrit à la figure 7. Il a été intégré à la médiathèque et dispose de tous les fichiers de voix avant et après rééducation, ainsi que les voix normales telles qu'elles apparaissent en ligne (Sicard, Menin-Sicard, 2012). L'utilisateur peut choisir parmi un choix d'indicateurs en X et en Y, obtenant ainsi des détails sur le coefficient de corrélation. Chaque point de l'espace X,Y correspond à un fichier. La distribution statistique de la variable choisie à l'axe X (ici l'instabilité en hauteur sur tout le son évaluée par VOCALAB) apparaît au-dessus du nuage de points. Le coefficient de corrélation évolue de 0 (aucune corrélation) à 1 (corrélation totale).

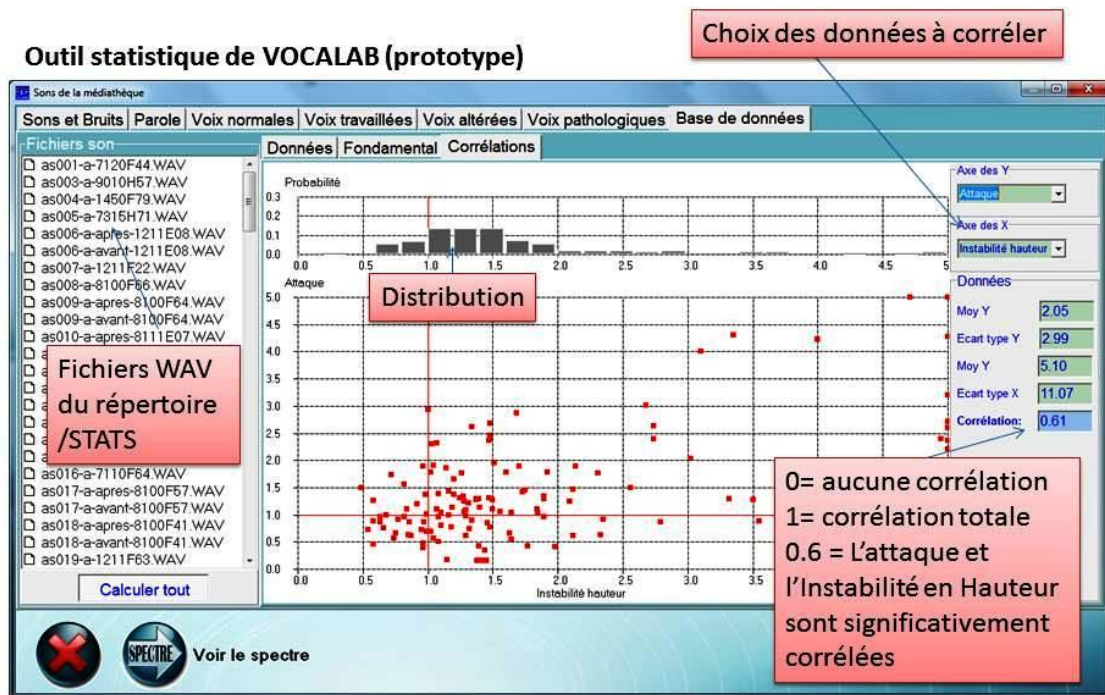


Figure 7. Outil de calcul statistique dans VOCALAB.

----- ANALYSE DES RÉSULTATS -----

Fondamental de la voix

Chaque voix de la base de données fait l'objet d'un calcul du fondamental instantané, puis d'une moyenne sur la durée de l'échantillon. La corrélation entre VOCALAB et PRAAT est excellente (0.9) à l'exception de cas de voix fortement dégradées, où les algorithmes de détection sont mis en défaut. Un tableau gradué en fréquence (axe horizontal) et âge permet de séparer les différents fichiers (figure 8).

La limite de 170 Hz est franchie principalement par les voix d'hommes avec pathologies liées à une chirurgie en cancérologie ORL. Certains cas de voix de femmes, détaillées à la figure 8, présentent un F0 inférieur à 150Hz dans le cadre de certaines pathologies spécifiques, dont

les caractéristiques vont dans le sens des descriptifs compilés dans l'ouvrage de Verdolini et al. (2006). On peut considérer qu'une tendance globale à la réduction de F0 avec l'âge se dessine pour les femmes, comme décrit dans la plupart des ouvrages du domaine (Baken, 2000 ; Kent, Ball, 2000). Cependant, il s'agit ici de voix pathologiques qui élargissent considérablement le nuage de points, et ne permettent pas de dégager de tendance claire.

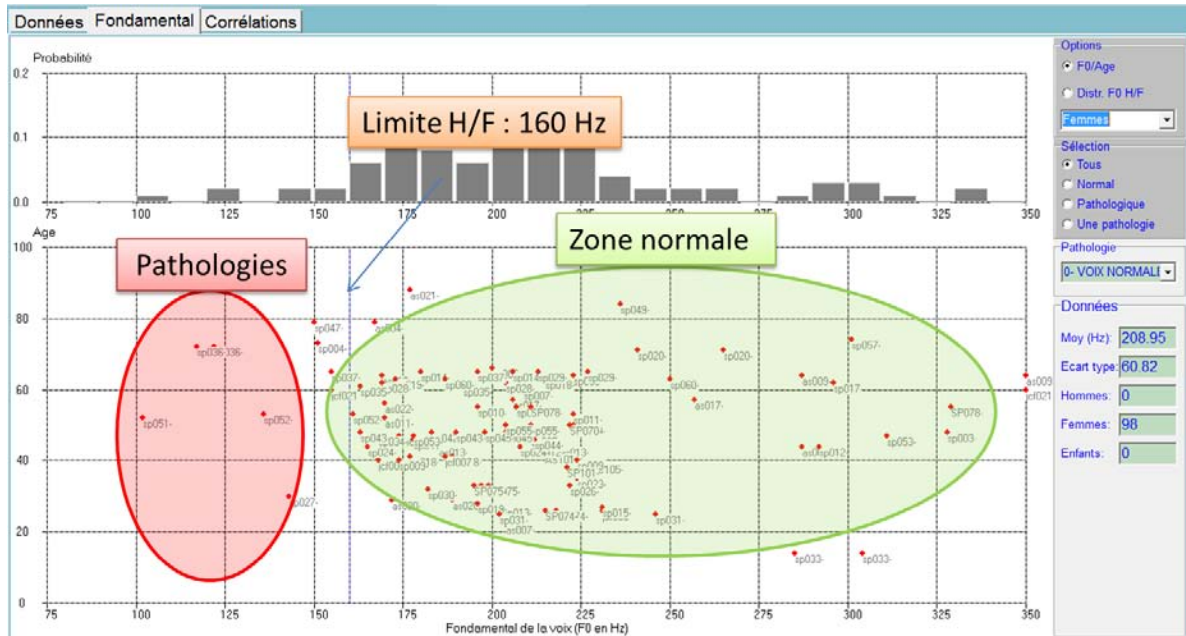


Figure 8. Répartition des voix en fonction de la fréquence du fondamental F0 (Axe des X) et de l'âge (Axe des Y).

Corrélation des marqueurs de VOCALAB pour les voix normales

La première question que nous nous posons est la séparation des indicateurs, autrement dit, le fait que chaque indicateur cible une information différente. Nous utilisons une trentaine de cas de voix normales (12 H, 18 F, âges variant de 15 à 80 ans). Le critère retenu pour considérer une voix comme normale consiste à n'avoir qu'un indicateur sur cinq au-dessus du seuil de pathologie, l'échantillon sonore étant jugé par ailleurs comme normal à l'audition par un évaluateur spécialisé en voix. La synthèse des résultats de corrélation est donnée à la figure 9.

Voix Normales	Attaque	Instabilité hauteur	Instabilité Amplitude	Bruit/Signal	Pauvreté harmonique	Instab haut 1s	Instab ampl 1s	Bruit/Signal 1s	Pauvreté harm 1s
Attaque	1.00	0.21	0.14	0.33	0.05	0.50	0.24	0.14	0.06
Instabilité hauteur		1.00	0.36	-0.02	-0.38	0.72	0.28	-0.01	-0.32
Instabilité Amplitude			1.00	0.35	0.12	0.24	0.68	0.30	0.14
Bruit/Signal				1.00	0.59	-0.01	0.39	0.92	0.58
Pauvreté harmonique					1.00	-0.34	0.22	0.56	0.96
Instab haut 1s						1.00	0.17	-0.00	-0.31
Instab ampl 1s							1.00	0.39	0.23
Bruit/Signal 1s								1.00	0.60
Pauvreté harm 1s									1.00

Figure 9. Matrice de corrélation entre les marqueurs de VOCALAB pour 30 échantillons de voix normales.

Les indicateurs de VOCALAB se révèlent peu corrélés entre eux, à l'exception de la pauvreté harmonique et du bruit/signal. Cette observation nous conforte dans l'idée que chaque indicateur mesure un phénomène bien différencié. Toutefois, l'indicateur de dégradation de l'attaque est assez corrélé avec l'instabilité en hauteur sur la seconde qui suit l'attaque, ce qui est assez logique dans la mesure où l'une des composantes du calcul de l'altération évalue l'instabilité en hauteur. On note moins de corrélation (0.21) entre l'attaque et l'instabilité en hauteur sur le son entier.

Les indicateurs sur le son entier sont assez fortement corrélés au son sur 1 seconde après l'attaque, notamment pour le bruit/signal et la pauvreté harmonique. En d'autres termes, un son de longue durée n'apporte que peu d'informations supplémentaires sur les notions de bruit et de pauvreté harmonique. Par contre, les résultats en instabilité en amplitude et en hauteur peuvent différer selon que l'on fait porter le calcul sur une portion de son après l'attaque, où le patient est en général en zone de confort, ou sur la totalité du son /a/ où divers phénomènes d'instabilité à long terme et de terminaison sonore se cumulent à l'instabilité moyen terme, dégradant presque systématiquement les marqueurs de pathologie.

Corrélations entre l'analyse sur le son complet ou sur 1 seconde après l'attaque

De même que pour les voix normales, nous observons une forte corrélation entre instabilité en hauteur du son entier et instabilité en hauteur d'une seconde après l'attaque pour l'ensemble des 150 voix pathologiques. Ceci signifie qu'un calcul sur l'ensemble du son est pertinent, de même que le calcul sur une seconde, ce dernier n'apporte pas de modification importante des résultats. L'instabilité en hauteur sur l'ensemble du son est en moyenne un peu plus élevée que pour une seconde de son.

Nous observons une corrélation moyenne (autour de 0.65) entre instabilité en amplitude du son entier et celui d'une seconde après l'attaque. Autrement dit, il peut exister d'importantes variations d'amplitude au-delà de la 1^{ère} seconde après l'attaque, plutôt stable et confortable, ce qui confirme la pertinence des calculs de stabilité d'amplitude sur le son complet plutôt que sur la 1^{ère} seconde.

Corrélations entre différents indicateurs

Comme souligné par Werth et al. (2010), le Jitter local et PPQ5 (Jitter moyenné sur 5 cycles) de PRAAT sont redondants, ainsi que le Shimmer local et Shimmer moyenné sur 5 cycles APQ5 (0.99). Evaluer les différentes variantes de Jitter et Shimmer n'apporte donc aucune valeur ajoutée, d'autant que chaque marqueur a une limite de pathologie différente, assez variable d'ailleurs selon les auteurs et le type de population étudiée.

Le HNR et le Shimmer dans PRAAT sont fortement corrélés (0.7) : le bruit induit des variations d'amplitude instantanée, donc du Shimmer, dans l'exemple du cas JCF012 de la base de données, illustré figure 10. Contrairement à PRAAT, VOCALAB analyse l'instabilité de F0, faiblement corrélé au bruit/signal. Les marqueurs sont mieux dissociés dans VOCALAB que dans PRAAT sur ces aspects.

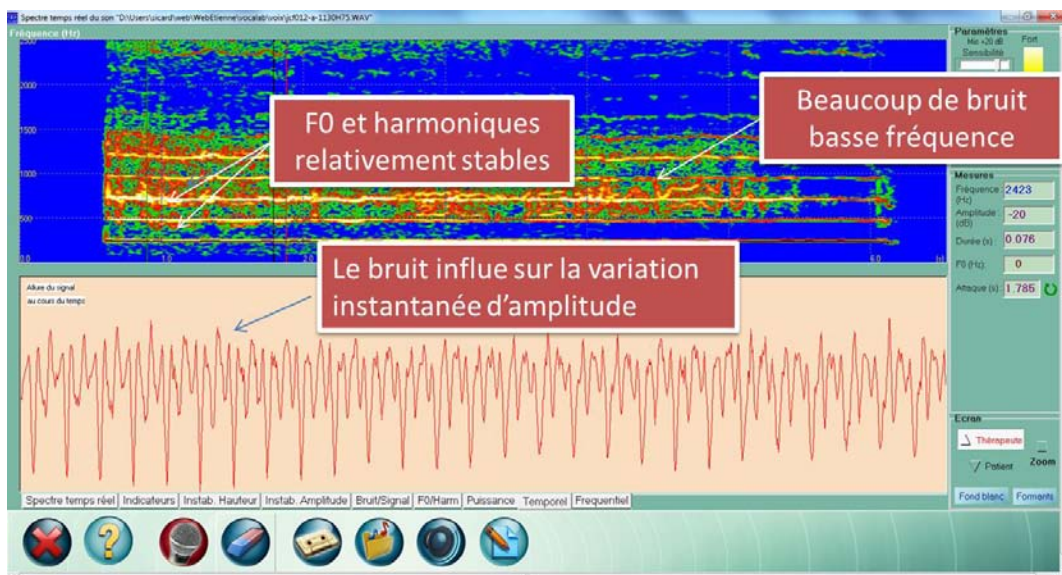


Figure 10. Illustration de l'influence du bruit sur le Shimmer, qui explique en partie la corrélation observée.

Jitter PRAAT et Instabilité en Hauteur de VOCALAB

De manière très surprenante, VOCALAB et PRAAT sont très peu corrélés (de 0.05 à 0.15 selon les algorithmes utilisés) sur la variation de la hauteur. L'analyse des cas extrêmes, les plus éloignés de la ligne pointillée de la figure 11, permet de donner quelques explications sur ce résultat inattendu :

- Dans les cas de voix très pathologiques (voix de substitution ?), PRAAT se trouve souvent en grande difficulté pour détecter F0. VOCALAB est aussi limité, mais pour des cas de voix plus sévères.

- VOCALAB se focalise sur les variations moyen et long terme, PRAAT sur les variations instantanées. Le cas AS009 a un Jitter à 0.3%, donc faible ($< 1\%$), mais une forte instabilité en hauteur sur le moyen terme (Figure 12).
- Le bruit dans PRAAT perturbe la mesure de F0 : HNR/Jitter corrélé à 0.5. Le cas SP012 met en évidence des bruits assez importants à diverses bandes de fréquence. L'allure temporelle complexe met en défaut le calcul de F0 avec PRAAT (Jitter très élevé), alors que l'instabilité n'est pas avérée (F0 jugé stable par VOCALAB, confirmé à l'audition).

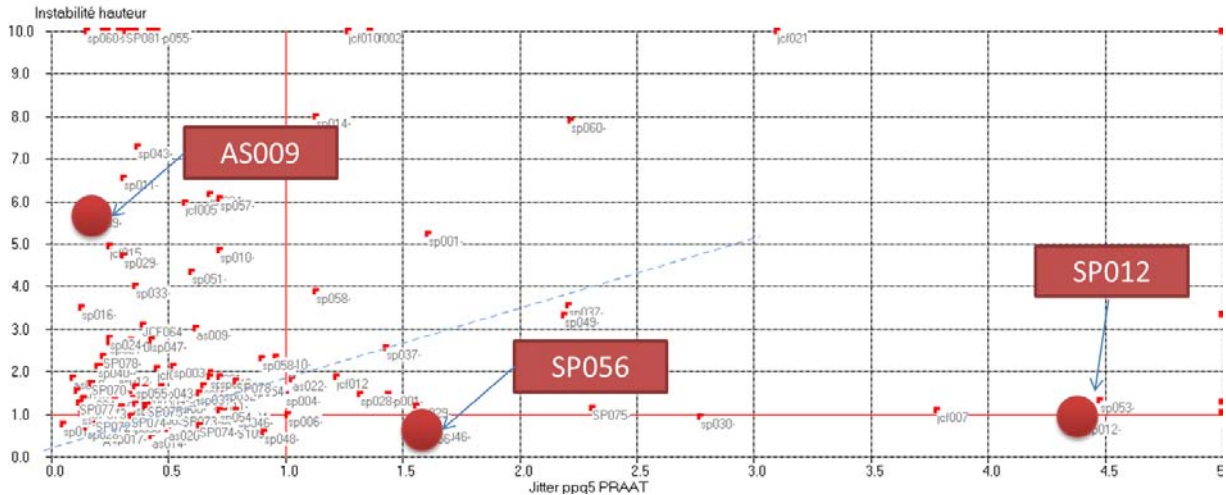


Figure 11. La grande disparité des cas de voix indique une faible corrélation entre le Jitter PPQ5 de PRAAT et l'Instabilité en hauteur de VOCALAB.

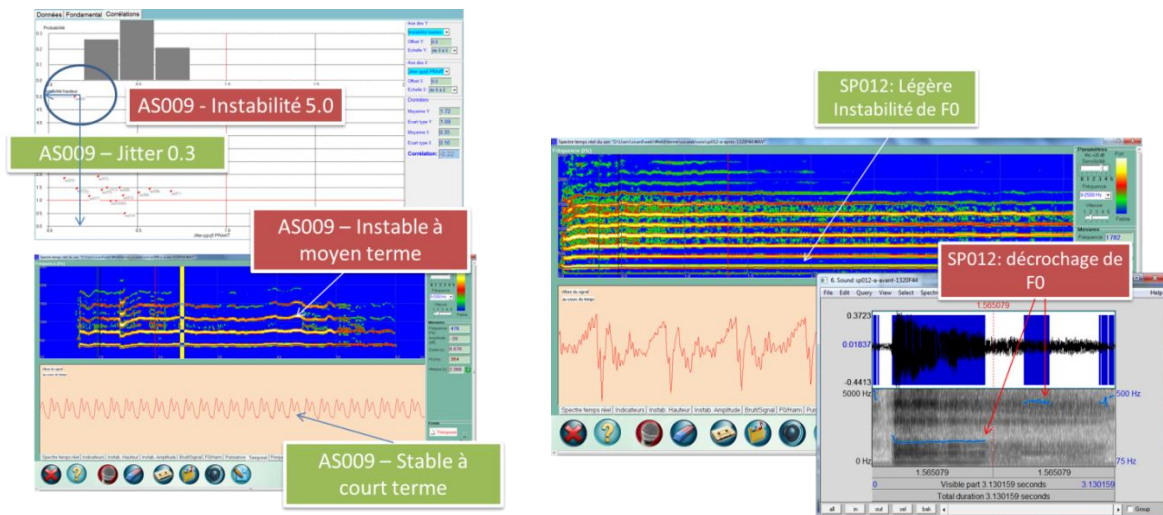


Figure 12. Détails sur les cas AS009 et SP012 avec des résultats divergents entre PRAAT et VOCALAB.

Shimmer et Instabilité en Amplitude

Le Shimmer de PRAAT est peu corrélé à l'instabilité en amplitude de VOCALAB. A l'analyse de différents cas divergents, nous pouvons mettre en évidence le fait que le bruit (souffle, ordinateur..) perturbe le Shimmer dans PRAAT et beaucoup moins dans VOCALAB. De plus, les informations mesurées ne sont pas les mêmes : VOCALAB se focalise sur les variations moyen et long terme, PRAAT sur les variations instantanées. Dans le cas SP053, un bruit BF important est visible proche de F0. Le Shimmer PRAAT est jugé très élevé (17%), bien que F0 soit stable en amplitude. A noter aussi la nécessité d'utiliser une chaîne d'enregistrement calibrée avec PRAAT pour l'estimation précise de la puissance en dB et du Shimmer. VOCALAB utilise l'écran puissance et des repères du type « voix murmurée, normale, projetée » pour affiner le réglage du niveau d'enregistrement.

Signal/Bruit VOCALAB et HNR de PRAAT

Les indicateurs bruit/signal VOCALAB et HNR de PRAAT sont assez bien corrélés (Figure 13). Certains points extrêmes correspondent à des enregistrements de voix très bruitées ou des situations très sensibles au paramétrage PRAAT. Une tendance se dessine sur une limite de pathologie entre 10 et 20 dB, variable selon les auteurs une fois encore (12 dB pour Hirano).

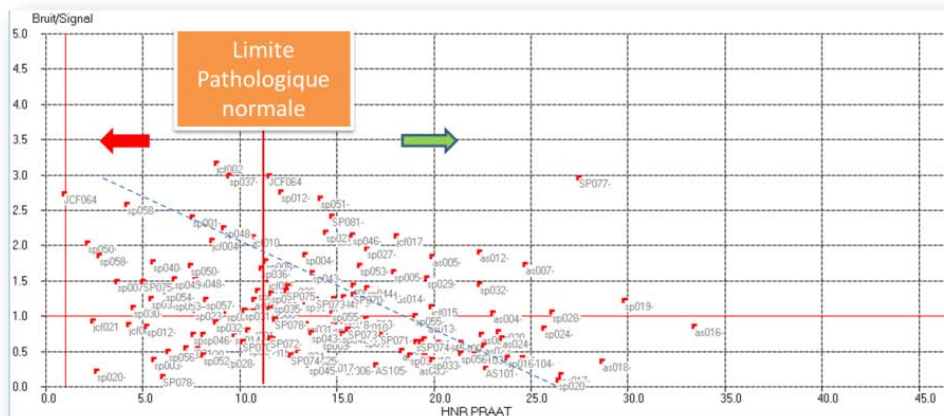


Figure 13. Corrélations entre PRAAT et VOCALAB pour le ratio bruit/signal.

CONCLUSIONS

Une étude sur 150 cas de voix pathologiques a été conduite dans le cadre de l'ERU 15 du laboratoire LURCO, dans le but d'évaluer la pertinence des marqueurs de pathologie de la voix existant dans PRAAT et proposés dans VOCALAB. L'intégralité des échantillons sonores a été mise à disposition des orthophonistes sur un site dédié, à des fins de recherche. Un script PRAAT a permis une extraction automatique des indicateurs de type Jitter, Shimmer et Harmonic/noise ratio. Les indicateurs de VOCALAB ont porté sur l'intégralité du son ou sur une portion de 1 seconde après l'attaque. Nous avons pu confirmer que les indicateurs au sein d'une même famille de Jitter et Shimmer sont très fortement corrélés, mais

aussi que le Shimmer et HNR dans PRAAT sont fortement corrélés, alors qu'ils ne ciblent pas les mêmes phénomènes. Nous avons mis en évidence une sensibilité de l'algorithme de détection d'amplitude au bruit, que nous avons essayé de minimiser dans VOCALAB. De manière surprenante, les indicateurs de hauteur de VOCALAB et Jitter de PRAAT sont peu corrélés, probablement parce que VOCALAB mesure le moyen et long terme, PRAAT le court terme. Une bonne connaissance du paramétrage des outils PRAAT s'avère nécessaire pour améliorer les résultats d'extraction du fondamental pour des pathologies sévères. Une calibration des indicateurs VOCALAB d'après PRAAT, qui avait été envisagée au commencement de l'étude, s'avère peu pertinente, les indicateurs étant difficilement comparables, à l'exception du rapport signal/bruit.

Les choix effectués sur les indicateurs de VOCALAB sont plutôt confortés que remis en question par cette étude de par leur pertinence pour les axes thérapeutiques et leur cohérence avec l'évaluation perceptive. La prise en compte des indicateurs dans les bilans et dans les rééducations est à relativiser cependant, du fait de la variabilité des valeurs obtenues selon les enregistrements, les chaînes d'acquisition, le matériel utilisé ou simplement le mode opératoire. D'autres dimensions telles que l'auto-évaluation sont à intégrer dans la démarche thérapeutique, les évaluations évoquées dans cet article ne devant être considérées que comme des indicateurs et non des marqueurs de la gravité des pathologies prises en charge.

----- BIBLIOGRAPHIE -----

Baken, R.J., Orlikoff, R.F. (2000). *Clinical measurement of speech and voice (2nd ed.)*. San Diego, CA : Singular.

Boersma, P., Weenink, D. (2009). Praat : doing phonetics by computer, version 5.3.56. Consulté le 24.09.2013 de Praat : <http://www.fon.hum.uva.nl/praat/>

Deliyski, D.D. (2012). *Multi-Dimensional Voice Program (MDVP), Model 5105*. Consulté le 24.09.2013 de Kay Pentax Corporation : www.kaypentax.com

Ghio, A., Giovanni, A., Teston, B., Révis, J., Yu, P., Ouaknine, M., Robert, D., Legou, T. (2008). *Bilan et perspectives de quinze ans d'évaluation vocale par méthodes instrumentales et perceptives*. 27èmes Journées d'Etude sur la Parole (JEP). Avignon : AFCP. Consulté le 24.09.2013 de Hyper Articles en Ligne : <http://halshs.archives-ouvertes.fr/docs/00/29/24/01/PDF/JEP-Ghio.pdf>

Hirano, M. (1981). Psycho-acoustic evaluation of voice : GRBAS Scale for evaluating the hoarse voice. In M. Hirano, *Clinical Examination of voice*, pp. 81-84. Wien : Springer Verlag.

Jacobson, B.H., Johnson, A., Grywalski, C., Silbergleit, A., Jacobson, G., Benninger, C.W., Newman, C.W. (1997). The Voice Handicap Index (VHI), development and validation. *American Journal of Speech-Language Pathology*, 6, 66-70.

Kent, R.D., Vorperian, H.K., Kent, J.F., Duffy, J.R. (2003). Voice dysfunction in dysarthria : application of the multi-dimensional voice program. *Journal of Communication Disorders*, 36(4), 281-306. Doi: 10.1016/S0021-9924(03)00016-9

Kent, R.D., Ball, M.J. (2000). *Voice Quality Measurement*. San Diego, CA : Singular Publishing Group.

Maryn, Y., Corthals, P., De Bodt, M., Van Cauwenberge, P., Deliyski, D. (2009). Perturbation measures of voice : a comparative study between multi-dimensional voice program and Praat. *Folia Phoniatica et Logopaedica*, 61(4), 217–226. Doi: 10.1159/000227999

Menin-Sicard, A., Sicard, E. (2009). Utiliser le logiciel Vocalab dans le cadre de la prise en charge des pathologies de la voix. In P. Gatignol (ed), *La voix dans tous ses maux*, pp. 37-45. Isbergues : Ortho-Edition.

Orlikoff, R.F., Deliyski, D., Baken, R.J., Watson, B.C. (2009). Validation of a glottographic measure of vocal attack. *Journal of Voice*, 23(2), 164–168.

Parsa, V., Jamieson, D.G. (2001). Acoustic discrimination of pathological voice : sustained vowels versus continuous speech. *Journal of Speech, Language and Hearing Research*, 44(2), 327–339. Doi: 10.1044/1092-4388(2001/027)

Pützer, M. (2001). Multiparametric description of voice quality for normal male and female voices (German). *Folia Phoniatica et Logopaedica*, 53(2), 73–84. Doi: 10.1159/000052657

Revis, J., Giovanni, A., Wuyts, F., Triglia, J.M. (1999). Comparison of different voice samples for perceptual analysis. *Folia Phoniatica et Logopaedica*, 51(3), 108-116. Doi: 10.1159/000021485

Sicard, E., Menin-Sicard, A. (2012). *Base de données de voix normales et pathologiques*. Consulté le 24.09.2013 de INSA Toulouse : <http://intranet-gei.insa-toulouse.fr/~sicard/vocalab/voix.html>

Sicard, E., Menin-Sicard, A. (2013-a). Implémentation dans VOCALAB d'indicateurs objectifs de la qualité de la voix dans le cadre de l'évaluation de la voix. *Rééducation Orthophonique*, 254, 43-60.

Sicard, E., Menin-Sicard, A. (2013-b). *Etude bibliographique sur les indicateurs de pathologie de la voix*. Consulté le 24.09.2013 de Hyper Articles en Ligne : http://hal.archives-ouvertes.fr/docs/00/83/69/00/PDF/ERU15_-_Etude_biblio_-_v12.pdf

Teston, B. (2004). L'évaluation objective des dysphonies : Etat actuel et perspectives d'évolution. In A. Giovanni (Ed.), *Le bilan d'une dysphonie, Etat actuel et perspectives*, pp.105-169. Marseille : Solal. Consulté le 24.09.2013 de Hyper Articles en Ligne : <http://hal.archives-ouvertes.fr/docs/00/17/35/53/PDF/2075.pdf>

Verdolini, K., Rosen, C.A., Branski, R.C. (2006). *Classification manual for voice disorders–I*. Mahwah, NJ : Lawrence Erlbaum.

Werth, K., Voigt, D., Döllinger, M., Eysholdt, U., Lohscheller, J. (2010). Clinical value of acoustic voice measures : a retrospective study. *European Archives of Oto-Rhino-Laryngology*, 267(8), 1261-1271.

Wolfe, V., Fitch, J.J., Cornell, R. (1995). Acoustic prediction of severity in commonly occurring voice problems. *Journal of Speech and Hearing Research*, 38, 273-279. PMID : 7596093

Xue, S.A., Deliyski, D. (2001). Effects of aging on selected acoustic voice parameters : preliminary normative data and educational implications. *Educational Gerontology*, 27(2), 159–168. Doi: 10.1080/03601270151075561

Yu, P., Ouakine, M., Révis, J., Giovanni, A. (2001). Objective voice analysis for dysphonic patients : a multiparametric protocol including acoustic and aerodynamic measurements. *Journal of Voice*, 15(4), 529-542. Consulté le 24.09.2013 du Laboratoire Parole et Langage : <http://aune.lpl.univ-aix.fr/~ghio/doc/Bib-2001-JVoice-Yu.pdf>

Yumoto, E., Gould, W.J., Baer, T. (1982). Harmonics-to-noise ratio as an index of the degree of hoarseness. *Journal of the Acoustical Society of America*, 71(6), 1544–1550.