

Etude Bibliographique sur les Indicateurs de Pathologie de la Voix

Etienne Sicard, Anne Menin-Sicard

ERU 15 – Laboratoire LURCO www.lurco.unadreo.org

INSA-GEI, 135 Avenue de Rangueil, 31077 Toulouse

Contact: etienne.sicard@insa-toulouse.fr

Résumé :

Le but de cette étude bibliographique est d'inventorier les marqueurs les plus pertinents pour l'analyse objective de la qualité de la voix. Nous étudions tout d'abord les logiciels les plus courants, en tentant de classer les marqueurs utilisés par grandes familles, afin de faire ressortir les points communs et les différences. Nous donnons ensuite une liste des références de publications que nous avons jugées intéressante pour nous guider dans le choix d'indicateurs pertinents pour l'évaluation et la rééducation de la voix, dans le cadre de développements futurs d'outils d'aide à la pratique orthophonique.

Note : de très nombreux sigles sont mentionnés dans cette revue, dont la signification est reportée en fin d'article sous forme d'un glossaire

Mots clé : orthophonie, évaluation de la voix, rééducation de la voix, traitement du signal, transformée de fourier, Shimmer, Jitter

1. Introduction

Dans le cadre du bilan et de la rééducation orthophonique de la voix, les méthodes d'évaluation perceptives faites par les orthophonistes se révèlent limitées de par leur variabilité et leur manque de fiabilité. L'analyse objective à l'aide d'outils dédiés tels qu'EVA et MDVP a longtemps été réservé à des orthophonistes experts, souvent dans le cadre hospitalier. Cette pratique se généralise en France, notamment par le biais des outils moins complets mais plus accessibles comme PRAAT et VOCALAB.

Nous décrivons dans cette synthèse bibliographique les outils les plus communs d'analyse objective de la voix (Section 2), et donnons un aperçu des outils avancés de calcul en lien avec les diverses pathologies de la voix (Section 3). Nous donnons une synthèse des publications les plus pertinentes de ce domaine à la section 4, en langue française et anglaise, puis passons en revue différents points discutés dans ces publications (Section 5). La synthèse et conclusion résume les points les plus significatifs de cette étude.

2. Les outils de base d'analyse objective de la voix

Le bilan et la rééducation orthophonique de la voix utilisent principalement les données synthétisées ci-dessous. Nous détaillons ci-après ces paramètres.

Type de donnée	Instantanée	Très court terme	Court terme	Moyen terme	Long terme	Très long terme
Voyelle /a/	Qualité d'attaque	Variations de fréquence, d'amplitude		Tenue en puissance. Présence de souffle, d'érailures. Aire Dynamique Vocale		Temps Maximum de phonation
Voyelle/Consonne a/s/z						Rapport s/z, a/z
Sirène /ou/				Etendue vocale		
Parole						Fondamental Usuel Moyen ; Interruptions

Figure 1: Principales données objectives utilisées dans le cadre de la rééducation de la voix

Les principaux outils d'analyse de la voix utilisés en France sont MDVP (Deliyski, 2012), PRAAT (Boersma, 2009) EVA (Ghio, 2008) et VOCALAB (Menin-Sicard, 2009). Ces outils ont en commun, à quelques variantes près, la mesure des paramètres listés à la figure xxx.

Le Fondamental usuel moyen (FUM) : celui-ci est calculé de manière statistique en moyennant les valeurs successives du fondamental, lui-même extrait de l'évaluation de chaque période successive, sur un intervalle de temps significatif (typiquement un texte de 20 secondes). Le FUM est dépendant du genre et de l'âge, comme précisé par Baken (1999) et de nombreuses études telles que (Xue 2001).

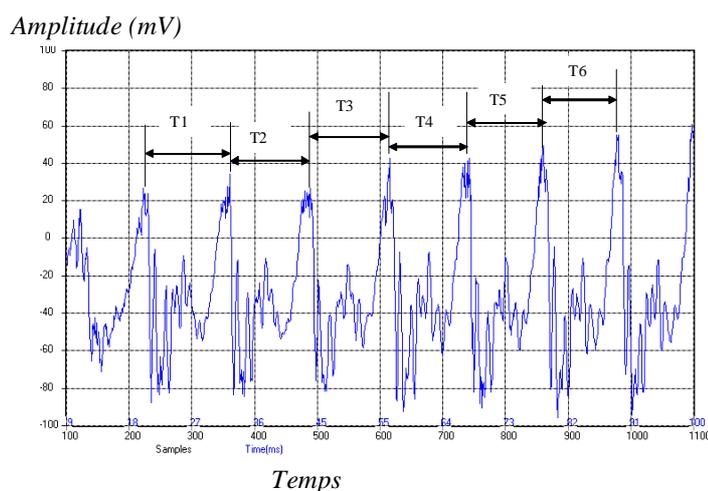


Figure 2: Extraction de chaque période instantanée du signal en vue de l'extraction du fondamental usuel moyen.

L'étendue vocale (EV) : Le patient est par exemple invité à prononcer une voyelle /ou/ de la hauteur la plus grave à la plus aigüe, puis de nouveau vers les graves, afin de déterminer l'étendue vocale, ainsi que la zone de confort maximale. La détection du fondamental et l'analyse des valeurs extrêmes donne une idée de l'étendue vocale.

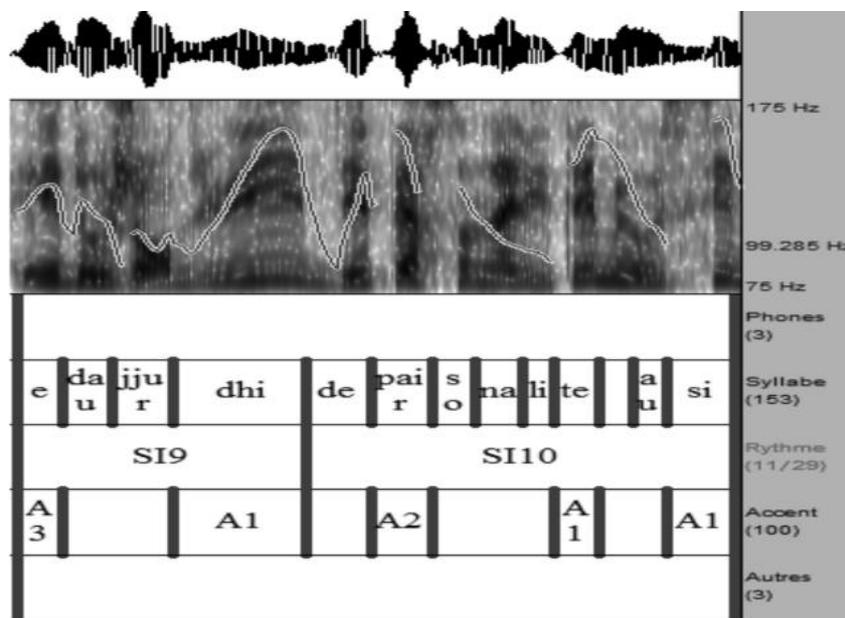


Figure 3 : L'excursion du fondamental dans le spectre permet de déduire l'étendue vocale (PRAAT)

Temps Phonatoire Maximum (TPM) : Le patient prononce la voyelle /a/, les consonnes /s/, /z/ afin de déterminer les rapports s/z et a/z. Les valeurs brutes sont moyennées sur 3 essais.



Figure 4 : Temps phonatoire maximum sous VOCALAB (Sicard, 2009)

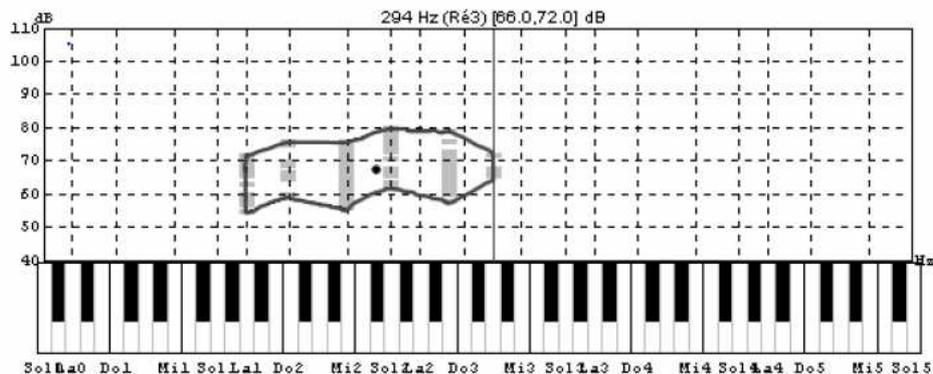


Figure 5 : Aire dynamique vocale avec EVA2 (Teston, 2004)

L'aire dynamique vocale (ADV) : Le patient est invité à prononcer une sirène sur la voyelle /ou/, d'une part avec l'intensité la plus faible, d'autre part avec l'intensité la plus forte.

3. Les outils avancés d'analyse objective de la voix

L'ouvrage de Baken (1999) recense plus de 200 variantes de marqueurs de qualité de la voix, et l'on trouve dans la littérature de très nombreux articles sur ce domaine. Certains marqueurs restent à l'état de recherches, tandis que d'autres sont intégrés dans les outils commerciaux. Nous nous sommes intéressés aux logiciels EVA2, PRAAT et MDVP, en tentant d'identifier leurs similarités et différences.

Les marqueurs de pathologie de la voix utilisés dans ces différents logiciels, ainsi que VOCALAB sont résumés sur le tableau ci-dessous par famille, en s'inspirant des classifications proposées par Putzer (2001). On retrouve dans les logiciels PRAAT, EVA et MDVP 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 conduites (en italique sur la figure) : analyse du débit d'air dans le système EVA, analyse des ruptures et irrégularités dans PRAAT et MDVP.

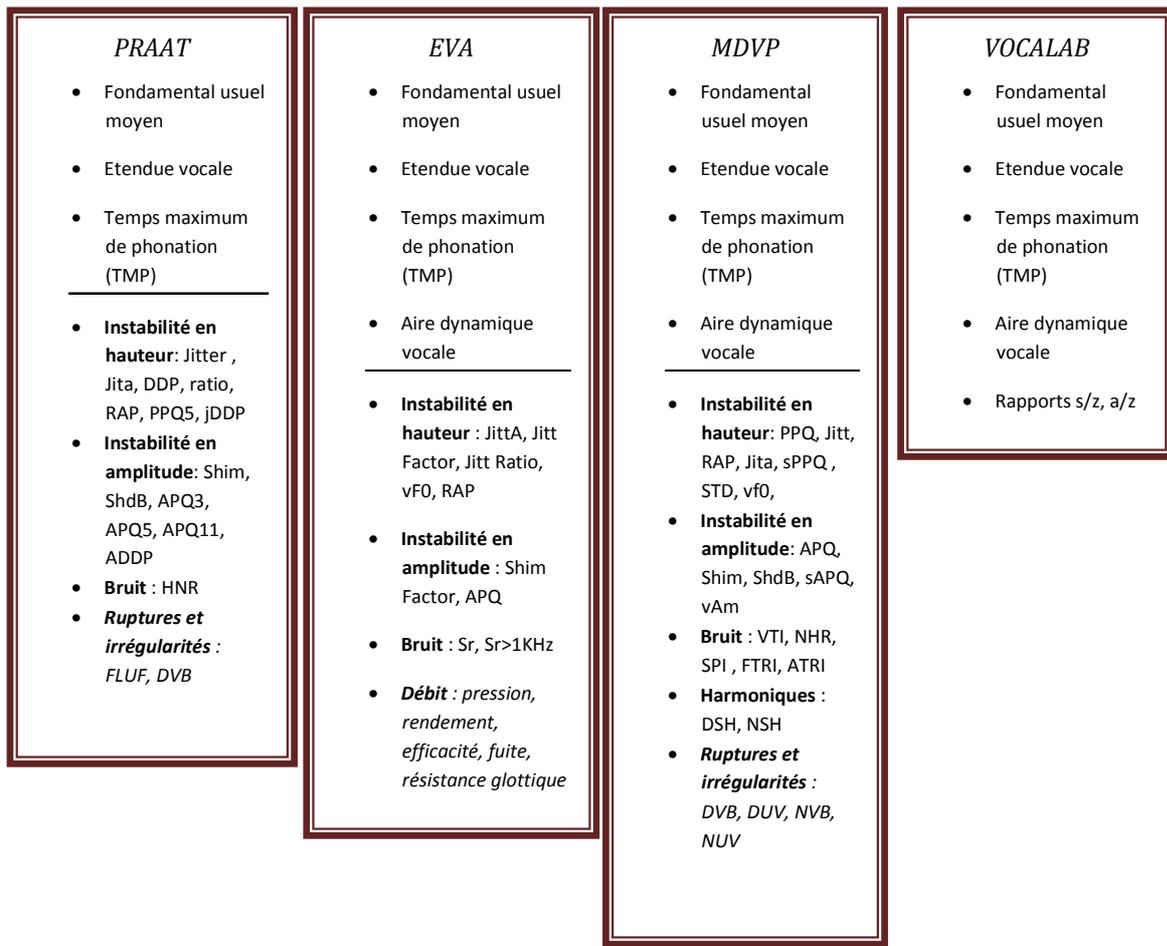


Figure 6 : Analyse des indicateurs de pathologie dans les systèmes EVA, MDVP, PRAAT et VOCALAB.

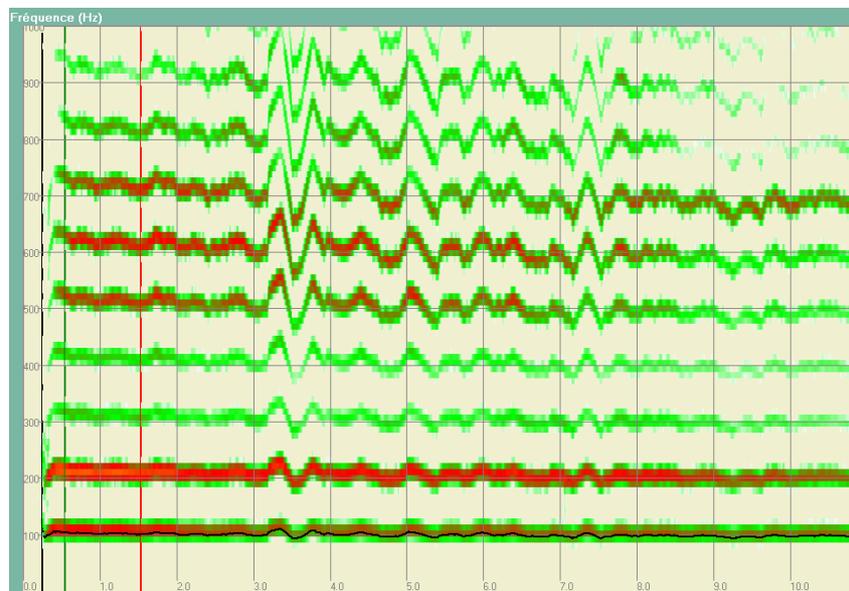


Figure 7: Illustration de l'instabilité de la hauteur de la voix (VOCALAB, cas AS024)

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 ou moyen terme, selon ses variantes. Le coefficient de Lyapunov (Teston 2004) correspond à une instabilité à plus long terme. 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 aux temps 3 et 7 secondes des variations importantes de F0, particulièrement bien visibles sur les harmoniques hautes.

La valeur du Jitter est fortement dépendante de l'algorithme utilisé pour la détection du fondamental de la voix. 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 (2009). Les seuils du Jitter relatif (% Jitt - variabilité de la période par rapport à la moyenne des périodes) communs aux différents logiciels définissent une frontière entre voix normales et pathologiques. Ces seuils sont en général autour de 1.0, mais varient selon les études, les protocoles et les logiciels utilisés (Kent, 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, 1999) pour inclure l'attaque. A noter que les différents indicateurs de Jitter sont fortement corrélés entre eux, comme le prouve Werth (2010) sur 2500 cas. Enfin, selon certains auteurs tels que (Parsa, 2001, Wolfe, 1995), le Jitter est peu corrélé avec le degré de sévérité des dysphonies.

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.

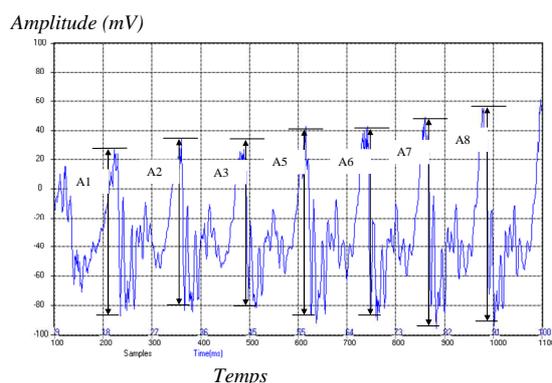


Figure 8: Illustration de la variation de l'amplitude du signal au cours du temps, qui est à la base du calcul du Shimmer.

Du fait que la détection du shimmer n'est pas basée sur l'évaluation de la fréquence fondamentale mais sur la variation d'amplitude de cycle en cycle, on trouve selon les études moins de variabilité sur cet indicateur que pour le Jitter. Cependant, les algorithmes mis en jeu 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).

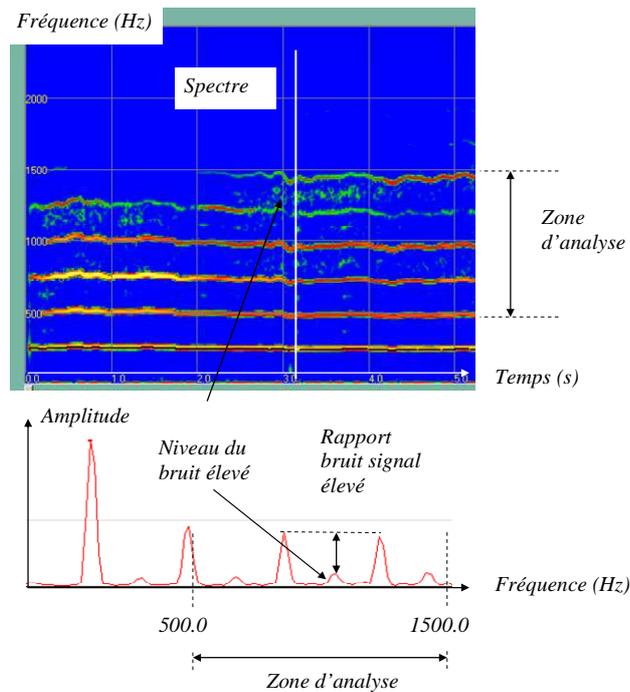


Figure 9: Détection du bruit et de son amplitude par rapport au signal.

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 (2001) utilise dans EVA deux marqueurs d'évaluation du rapport Signal-Bruit (S_r , en %, et $S_r > 1\text{KHz}$ en %), relativement peu corrélées entre elles. Le marqueur S_r décroît de 60% à 20% avec la sévérité, $S_r > 1\text{KHz}$ de 10 à 5%, avec peu de différence Homme/Femme.

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 2001). Werth (2010) analyse un très grand nombre de cas de pathologies de voix à l'aide du logiciel Dr. Speech. Les valeurs typiques de HNR 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 HNR (-0.67) passe de -8 dB à -6dB pour ces mêmes cas. En résumé, on constate une certaine variété de conditions de calculs, de normes et de résultats selon les outils utilisés.

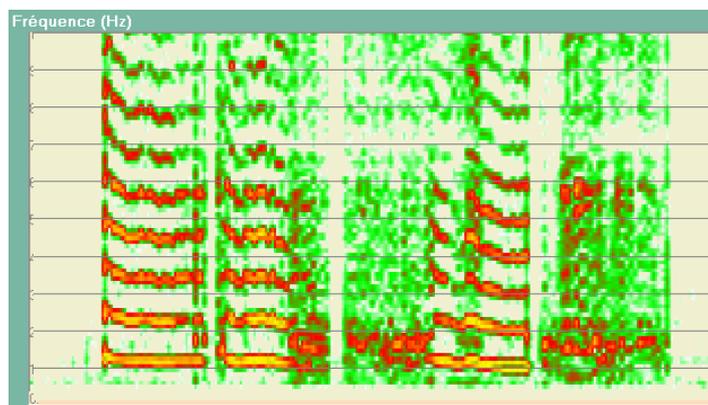


Figure 10: Exemple de voix pathologique (voyelle /a/) mettant en évidence des interruptions dans la production du son.

L'évaluation des phénomènes transitoires parasites. Les paramètres DUV (Degree of Voiceless /%/) et DVB (Degree of Voice Breaks /%/) de MDVP mesure 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. DUV et DVB évaluent les pauses entre les zones voisées. Dans PRAAT les indicateurs « Fraction of locally unvoiced frames » (FLUV) et DVB sont aussi proposés, fortement corrélés à MDVP.

4. Synthèse des publications pertinentes

Nous donnons ci-après un tableau récapitulatif des références que nous avons analysées, en précisant la population étudiée, les marqueurs ciblés, et quelques éléments qui ressortent de l'étude. A noter qu'un certain nombre de mémoires d'orthophonie traitent aussi de la voix et des indicateurs de pathologie (Site FNEO).

Référence	Population étudiée	Marqueurs pertinents	Commentaires
Parsa (2001)	53 sujets normaux et 175 sujets pathologiques	Jitter	Le jitter détecte 68% des voix pathologiques
Yumoto (1982)	40 normaux, 41 pathologiques	Harmonics-to-Noise Ratio (HNR)	Différence entre l'énergie du signal et l'énergie moyennée (50 périodes). H/N normal : 12 dB, décroît et devient négatif pour des pathologies sévères.
Piccirello (1998)	97 sujets dysphoniques et 35 sujets témoins	Temps maximum de phonation (TMP) et l'étendue vocale.	
Wuyts (2000)	319 sujets présentant une pathologie variée	13 paramètres acoustiques et aérodynamiques.	4 paramètres pertinents : fréquence la plus élevée (Fo-High in Hz), intensité la plus basse (I-Low in dB), temps maximum de phonation (TMP), et jitter (%).
Yu (2001)	63 dysphonies et 21 sujets témoins	Paramètres EVA2	jitter, coefficient de Lyapunov (CL), rapport signal sur bruit (SRf>0 et SRf>1kHz), étendue vocale, TMP
An Xue, (2001)	100 H/F	14 marqueurs de Kay Elemetrics	APQ %, Jita ms, Jitt %, PFR, PPQ %, RAP %, sPPQ %, STD Hz, vFo %, ShdB dB, Shim %, SPI, NHR, and VTI
Campisi, (2002)	100 boys and girls, with vocal fold nodule group	33 variables	PPQ, SPPQ, Jitt, RAP strongly correlated
Baken (1999)		250 marqueurs de qualité de la voix	Revue très complètes des algorithmes publiés
Robert (2011).	40 dont 10 témoins		High Freq Power Ratio, VOT, Burst, Formant F3
Wolfe (1995)	20 normaux + 60 patients atteints de dysphonies de sévérité diverses	Jitter, Shimmer, Signal-Noise Ratio	Jitter et Shimmer peu corrélée avec la sévérité de dysphonie (r=0.5 max) pour les cas de nodules, paralysie laryngée unilatérale et dysphonie dysfonctionnelle

Lalot (2008)	12 normaux + 12 voix soufflées	Jitter, Shimmer, HNR,	HNR intéressant. Jitter et le shimmer peu corrélés aux voix soufflées
Werth (2010)	580 healthy and 1,700 pathologic voices	All types of jitter and Shimmer	Très forte corrélation entre les Jitter entre eux et Shimmer entre eux. Etude de HNR et NNE. Outil : Doctor Speech, Tiger Electronics Inc., USA
Koreman (1997)	20 divisés en 4 groupes, dont 5 normaux et pathologies de type paralysie unilatérale, cordectomie		Paramètre différentiateur : étendue vocale. Dans certains cas, Absolute Jitter (Jita) et PPQ sont plus élevés pour les paralysies, mais uniquement pour les femmes
Demenko (2004).	100 patients	Utilisation échelle GRBAS	Lien entre l'échelle GRBAS et les paramètres Jitter, Shimmer, VFO, NHR (noise to harmonic Ratio)
Kent (2003)	Parkinson disease (PD), cerebellar disease, amyotrophic lateral sclerosis (ALS), traumatic brain injury (TBI), unilateral hemispheric stroke, and essential tremor	MDVP de KayElemetrics	Recommandations dans le choix des échantillons. Fortes variabilités sur les « norme » (Jitter =1%, 2%) selon les études.
Putzer (2001)			Regroupe les MDVP en quatre groupes
Revis (2004)	Etude 80 patients : /a/ avec ou sans attaque	Rôle de l'attaque ;	Jitter plus important dans les transitoires ; durée de l'attaque augmente avec le degré de dysphonie ; importance de l'attaque du /a/ pour juger de la sévérité de la dysphonie
De Krom (1995)	57 voice patients + 21 healthy speakers ; Importance de l'attaque dans l'évaluation de la dysphonie	F0, HNR, applied on vowel onset (VO), post-onset (PO), and whole vowel fragments (WV).	Corrélation 0.79 sans attaque, 0.89 avec attaque. Acoustic disturbances at VOT : associated with roughness. Vbreathy sensation. Découpage b0, 60 to 400 Hz; b1, 400 to 2000 Hz; b2, 2000 to 5000 Hz; b3, 5000 to 8000 Hz
Ghio (2008)	450 voix. Voyelle /a/, analyse sur partie stable	Jitter, F0, SNR, TMP, Etendue Vocale, Lyapounov	Jitter augmente de manière exponentielle avec la sévérité de la dysphonie, TMP, SNR diminue de manière exp ; nécessité de séparer H/F ; limite F0 autour de 175 Hz
Shoji (1992)		HPR (< et > à 6000Hz)	Comparaison énergie sous et au-dessus de 6000 Hz différencie une voix normale d'une voix soufflée

5. Discussion

Classements

Putzer (2001) regroupe 4 types de marqueurs entre eux, de part leur nature très fortement corrélée :

- Jitter factor: PPQ, Jitt, RAP, Jita, sPPQ
- Shimmer factor: APQ, Shim, ShdB, sAPQ
- f0-shimmer factor: STD, vf0, vAm
- Spectral factor: VTI, NHR, SPI

Lierde (1996) classe plutôt les paramètres par objectifs de mesures :

- Fundamental frequency parameters: f0, FHI, STD, FLO, PFR
- Frequency perturbation parameters: Jita, Jitter, RAP, PPQ, sPPQ, vf0
- Amplitude parameters: ShdB, Shim, APQ, sAPQ, vAm
- Voice irregularity parameters: DUV, NUV
- Tremor analysis parameters; FTRI, ATRI, FFTR, FATR
- Voice break parameters: DVB, NVB
- Subharmonic parameters: DSH, NSH
- Noise-related parameters: NHR, VTI, SPI

Shimmer, Jitter

Une illustration de la corrélation entre les différents Shimmer et Jitter est donnée par Werth (2010) sur 2500 cas avec Dr Speech, en se basant sur le Jitter « PPQ ».

Table 5 Correlation values for the different jitter and shimmer measures for males and females

	Jitter (male)	Jitter (female)
Jitter [abs]	0.939**	0.913**
Jitter [percent]	0.987**	0.985**
Jitter [RAP]	0.969**	0.966**
Jitter [PPQ]	1.0**	1.0**
Jitter [11PA]	0.983**	0.989**
	Shimmer (male)	Shimmer (female)
Shimmer [percent]	0.968**	0.970**
Shimmer [3PA]	0.941**	0.941**
Shimmer [APQ]	0.967**	0.976**
Shimmer [11PA]	1.0**	1.0**

Figure 11: Corrélations entre Jitter et Shimmer (Werth, 2010).

Un des problèmes les plus fréquemment soulevés est la difficulté de comparer les résultats sur des marqueurs identiques, mais obtenus avec des outils différents. Les algorithmes mis en jeu sont montrés du doigt (Maryn, 2009). Ainsi, la trace de F0 est très différente entre MDVP (probablement erroné, vu la nature du signal) et PRAAT (semble plus conforme), ce qui expliquerait les valeurs globalement 2 fois plus faibles de PRAAT. L'échantillon de son est assez fortement bruité, ce qui se traduit par un défaut de détection dans MDVP.

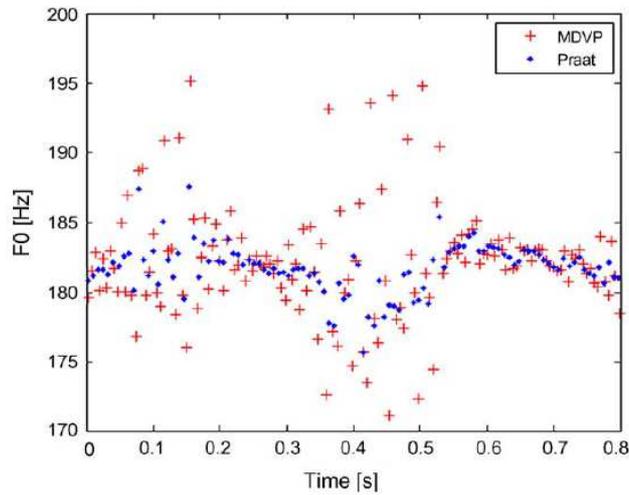


Figure 12 : la détection instantanée du Fondamental F0 de la voix est incorrecte avec MDVP, et correcte avec PRAAT (Maryn, 2009).

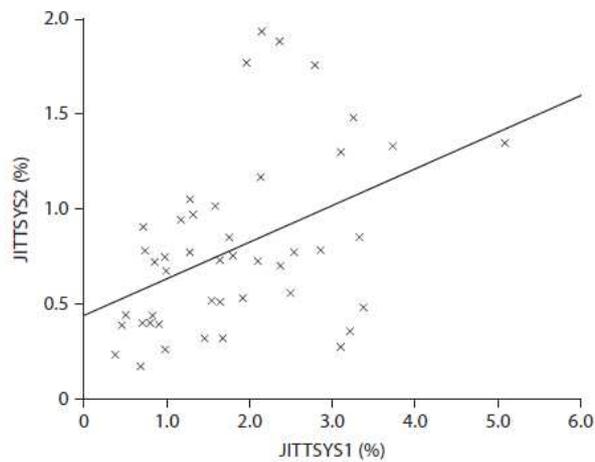


Figure 13 : Comparaison entre les calculs de Jitter MDVP et PRAAT, démontrant une corrélation assez faible (Maryn, 2009).

Maryn (2009) constate une corrélation $r = 0.44$ (idéalement 1.0) entre MDVP (SYS1) et PRAAT (SYS2), avec un nuage de point éloigné de la droite idéale. Deux zones sont particulièrement problématiques : au-delà des frontières de normalité (1.0 pour MDVP, 0.5 pour PRAAT environ), de nombreux cas sont détectés par l'un des logiciels comme fortement pathologiques, l'autre non. Maryn cite Smits (2005) qui trouve une corrélation encore plus basse : 0.31 entre Dr. Speech et CSL de KayElementrics. Dans les différentes études, le shimmer moyen terme est mieux corrélé : APQ MDVP (PRO1) et APQ PRAAT (PRO2) sont proches avec $r=0.87$ (idéalement 1).

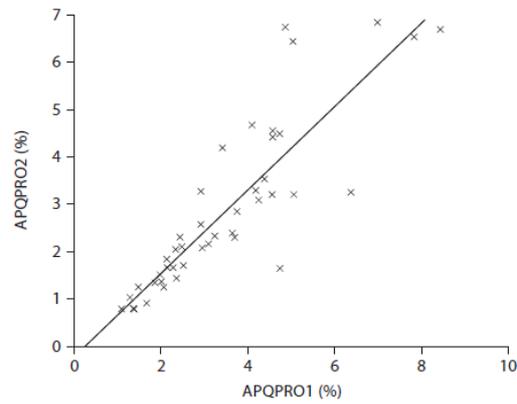


Figure 14 : le calcul du Shimmer (APQ) est mieux corrélé entre MDVP et PRAAT (Maryn, 2009).

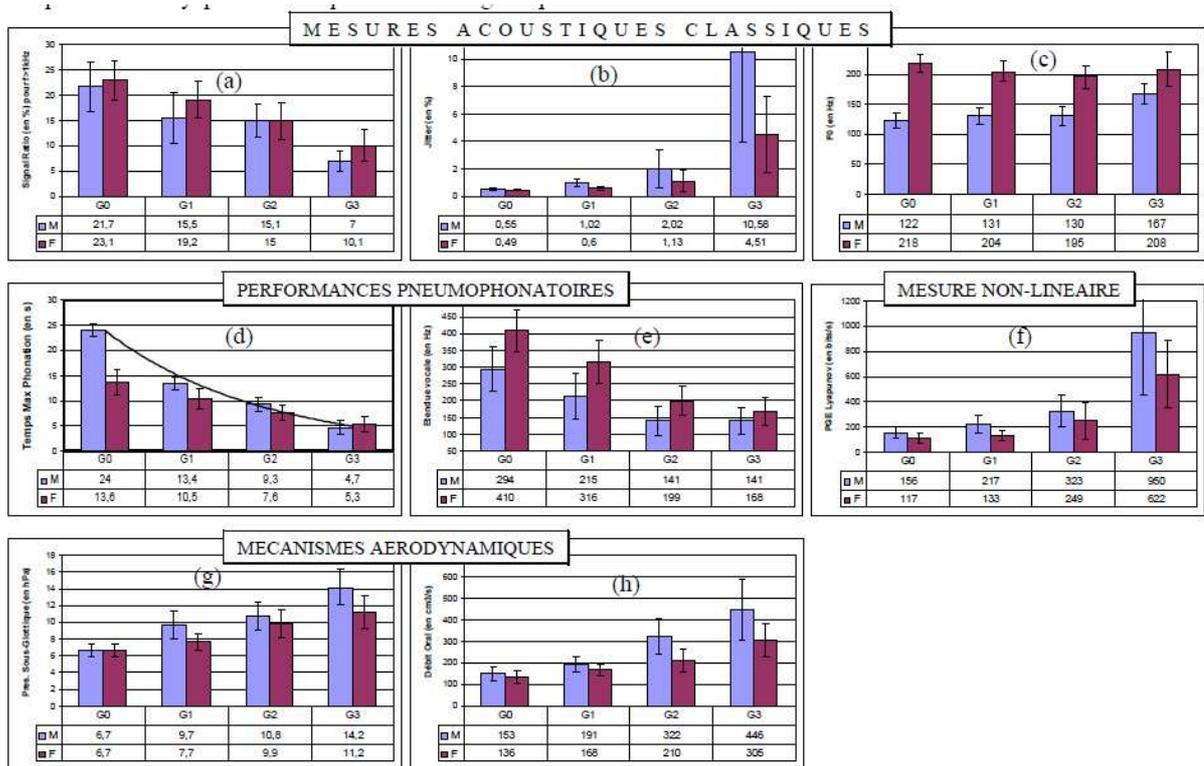


Figure 15: Corrélations entre les principaux indicateurs et le degré de sévérité des dysphonies (Ghio, 2008).

Ghio (2008) montre le lien entre les différentes métriques (jitter, F0, TPM, Lyapounov, etc) et le degré de sévérité, montrant des tendances décroissantes pour SNR, TMP, EV et croissantes pour les autres. A noter des valeurs significativement différentes entre H/F. Plus la dysphonie est jugée sévère, plus le temps maximal de phonation (TMP) diminue, plus l'étendue vocale (Voice Range ou VR), plus le jitter augmente (instabilité laryngée, plus le taux d'énergie harmonique (Sr) diminue (pauvreté harmonique et présence de bruit, plus le plus grand exposant de Lyapunov (PGE) augmente (comportement chaotique du vibrateur laryngé).

Paramètre	Unités	Frontière normalité	Limite supérieure pathologique	Relation avec l'état pathologique
Débit d'air oral	dm ³ /s	0.2	0.5	++++
Pression sous glottique	hPa	7	18	++++
Efficacité glottique	dB/hPa	12.5	5	----
Fuite glottique	cm ³ /s/dB	4	8	++++
Rendement glottique	dB/(hPa.dm ³ /s)	60	10	----
Résistance glottique	hPa/dm ³ /s	40	100	++++
Jitter absolu	Hz	0.3	4	++++
Jitter Factor	o/o	0.99	5	++++
Jitter Ratio	o/oo	8	15	++++
C. V. Fo	o/o	1.5	4	++++
RAP	----	0.003	0.01	++++
Shimmer	dB	0.4	1.5	++++
Shimmer Factor	o/o	0.99	5	++++
APQ	----	0.1	0.3	++++
Signal Ratio	o/o	70	0	----
H/N Ratio	dB	7.4	-10	----
NNE	dB	-6	0	++++

Figure 16: Frontières de normalité pour les marqueurs EVA et tendance à la hausse (++++) ou à la baisse (----) avec la sévérité de la pathologie (Ghio, 2008).

TABLE 1 Means and Standard Deviations of Norms of 15 Selected Acoustic Parameters From the Sustained /a/ Productions of Elderly and Young and Middle-Aged Speakers

voice parameter	All elderly (n = 44)		All young (n = 53)		Elderly men (n = 21)		Young men (n = 20)		Elderly women (n = 23)		Young women (n = 33)	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Fo Hz	159.02	47.09	204.85	54.82	127.62	29.18	145.22	23.41	187.70	42.15	243.97	27.46
APQ %	4.04	2.78	1.63	0.71	4.20	2.28	1.99	0.81	3.89	3.22	1.40	0.53
Jita µs	143.71	130.24	32.77	27.33	170.11	127.78	41.66	36.48	119.61	130.53	26.93	16.65
Jitt %	2.06	1.81	0.62	0.43	2.10	1.55	0.59	0.54	2.02	2.03	0.63	0.33
NHR	0.19	0.10	0.12	0.01	0.18	0.08	0.12	0.01	0.20	0.11	0.11	0.01
PFR st	4.11	3.19	2.19	1.07	3.52	1.63	2.10	1.07	4.65	4.11	2.25	1.06
PPQ %	1.25	1.17	0.36	0.24	1.24	0.98	0.34	0.29	1.26	1.35	0.37	0.21
RAP %	1.22	1.07	0.37	0.27	1.24	0.92	0.35	0.33	1.20	1.21	0.38	0.21
ShdB dB	0.48	0.37	0.19	0.08	0.49	0.31	0.22	0.09	0.48	0.42	0.18	0.07
Shim %	5.43	4.02	2.21	0.92	5.54	3.51	2.52	1.00	5.34	4.51	2.00	0.79
SPI	14.50	10.50	7.23	4.02	19.24	12.47	6.77	3.78	10.18	5.76	7.53	4.13
sPPQ %	2.11	2.71	0.54	0.26	1.74	0.98	0.56	0.30	2.45	3.64	0.53	0.22
STD Hz	5.44	8.27	2.18	1.83	3.12	1.56	1.35	0.68	7.57	11.02	2.72	2.12
vFO %	3.53	5.59	1.07	0.83	2.45	1.20	0.94	0.43	4.51	7.59	1.15	1.01
VTI	0.08	0.07	0.05	0.01	0.08	0.07	0.05	0.02	0.08	0.07	0.05	0.01

Figure 17: Marqueurs de voix normales et évolution avec l'âge, pour une centaine de cas de voix (Xue, 2001).

Normes

Un des problèmes soulevé par Kent (2004) est ce que l'on appelle un « groupe normal ». Selon les populations « normales » étudiées (âge, pays..), les valeurs moyennes varient très fortement. Le Jitter à 1% (norme KeyPentax) passe à 2% selon les auteurs. Ghio (2008) donne pour EVA2 les frontières de normalité pour les différents marqueurs.

An Xue (2001) analyse les voix d'une centaine d'hommes et de femmes, et dresse un tableau comparatif de 15 marqueurs de MDVP, permettant d'avoir une idée des normes dans le cas de voix normales.

Selon Ghio (2010=), les études sur le dysfonctionnement de la voix et de la parole souffrent d'une dispersion et d'une hétérogénéité des données, ainsi que de la perte quasi systématique des méta-données (exemples : pathologie, durée de la maladie, âge, contexte d'enregistrement...), ce qui explique souvent la difficulté de faire émerger des résultats clairs.

La construction d'une base de données permet de décrire et évaluer les dysfonctionnements de la voix et de la parole, tout en respectant la protection de l'identité du patient (CNIL, 2006). Une expérimentation de ce type est conduite dans la région PACA (Ghio, 2010), incluant la collecte de données sociolinguistiques, médicales, symptomatiques et contextuelles.

Liens avec l'évaluation subjective

En lien avec le GRBAS, Demenko (2004) démontre un lien entre l'estimation subjective (GRBAS) et certains paramètres objectifs tels que Jitt, Shim, vFO et NHR extraits avec MDVP.

MEAN	Hoarseness		Roughness		Breathiness		NORM	
	G3	G2	G1	R2	R1	B2		B1
JITT%	6,31	5,53	2,49	6,41	2,98	4,82	5,55	1,44
SHIM%	9,87	7,62	3,08	9,19	3,79	6,90	9,08	4,15
vFO %	7,31	6,46	3,26	7,78	3,11	5,76	6,38	1,57
NHR	0,23	0,19	0,12	0,15	0,11	0,27	0,21	0,15

Figure 18: Corrélations entre Jitter, Shimmer, NHR (Demenko, 2004).

Analyse du bruit

Yumoto (1982) expose une technique simple de calcul du rapport Harmonique/Bruit (H/N en Anglais). Ce rapport H/N est basée sur la différence entre l'énergie instantanée du signal (qui contient le bruit) et l'énergie moyennée de 50 périodes de son (la moyenne sur un nombre suffisant de périodes filtre le bruit du fait de sa nature aléatoire). Pour le groupe normal, H/N est centré sur 12 dB, et peut devenir négatif pour des pathologies sévères.

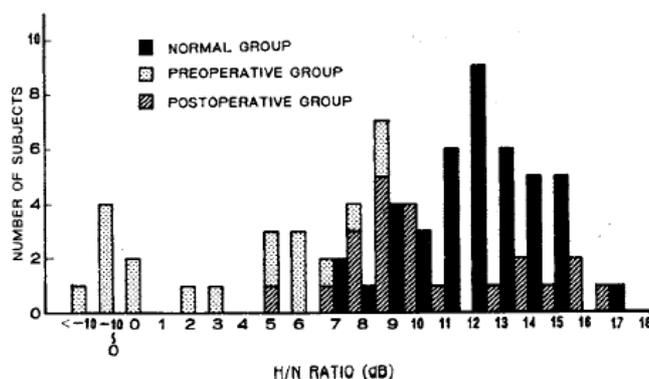


Figure 19: Exemple de H/N ratio pour des groupes normaux et pathologiques, avant et après opération (Yumoto, 1982)

Yu (2001) utilise dans EVA deux marqueurs d'évaluation du rapport Signal-Bruit (Sr, en %, et Sr>1KHz en %), peu corrélées entre elles. Le marqueur Sr décroît de 60% à 20% avec la sévérité, Sr>1KHz de 10 à 5%. Ghio (2008) observe peu de différence Homme/Femme.

L'outil MDVP propose un marqueur Noise-to-harmonic ratio (NHR) qui est la moyenne de 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 l'âge autour de 0.2 (Xue 2001).

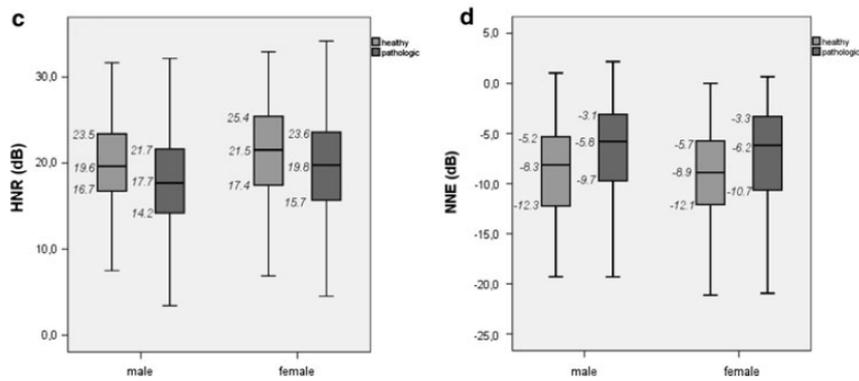


Figure 20: comparaison entre marqueurs HNR et NNE pour les groupes normaux et pathologiques sur plus de 1500 cas (Werth 2010),

Werth (2010) analyse un très grand nombre de cas de pathologies de voix (1700 cas) à l'aide Dr. Speech. Les valeurs typiques de HNR sont autour de 20 dB pour les cas normaux, et ne baissent que de 2 dB pour les voix pathologiques. Le facteur NNE, corrélé à HNR (-0.67) passe de -8 dB à -6dB pour ces mêmes cas. A noter la corrélation entre HNR et Shimmer (-0.72), ainsi que HNR et Jitter (-0.6).

Attaque

J. Revis (1999, 2004) estime que l'attaque du /a/ est porteuse d'information dysphonique qui manque à l'évaluateur pour une portion stable de la voyelle. Plus la dysphonie est importante, plus le /a/ tronqué sous-estime le degré de sévérité, par rapport à un jugement sur la parole ou sur le /a/ complet. L'attaque est aussi mentionnée par Henrich (2008).

G. De Krom (1995) indique que la perturbation acoustique juste avant la mise en vibration de la voyelle est associée à "roughness", tandis que la fin de voyelle correspond à la sensation de soufflé "breathy". Il indique aussi que l'information d'attaque augmente la corrélation avec la pathologie (de .79 sans information d'attaque à .89 avec attaque).

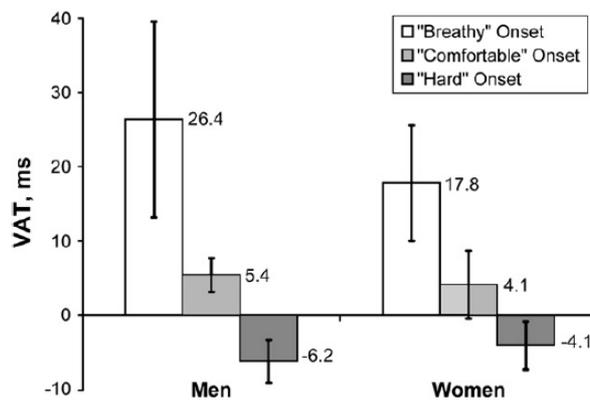


Figure 21: Indicateur de durée d'attaque (Orlikoff, 2009) différenciant les attaques soufflées, normales et dures.

Dans son article de 2009, Orlikoff propose une métrique « Voice Attack » qui caractérise la durée de mise en vibration de la voyelle, et constitue un indicateur significatif de dysfonctionnement neuronal central ou périphérique. L'attaque confortable est située à 5 ms, tandis qu'une attaque soufflée (breathy) est autour de 20 ms, une attaque dure donnant un VAT négatif. Cependant, l'étude ne porte que sur dixaine d'individus.

6. Synthèse et Conclusion

De cette étude bibliographique, il ressort une très grande diversité de calculs possibles en vue d'une évaluation objective de la voix, dans un contexte d'évaluation et de rééducation orthophonique. On peut toutefois regrouper les marqueurs par familles, du fait de leur lien souvent étroit, et sélectionner ainsi un nombre réduit d'indicateurs qui couvrent un domaine assez large de caractéristiques, tout en restant pertinents pour la majorité des cas de voix traités par l'orthophoniste.

On peut considérer 4 indicateurs principaux : la variation de la hauteur de la voix, la variation de la puissance sonore, la qualité des transitions et la présence de bruit.

Sur l'aspect hauteur de la voix, les mesures instantanées peu visibles et peu audibles, semblent moins intéressantes que les mesures à moyen terme (vibrato) et long terme (instabilité). La plus grande vigilance est de mise concernant la robustesse de l'algorithme de détection du fondamental, sur lequel se base le calcul de l'instabilité en hauteur et ses dérivés. La comparaison avec PRAAT est à privilégier au détriment de MDVP (mise en évidence de défauts). Aucune étude comparative n'est à ce jour disponible concernant EVA.

De même, sur **l'aspect puissance**, la mesure moyen terme (à-coup) ou long terme (tonus) est à privilégier, plutôt que la variation instantanée. Une moindre vigilance est requise quant à la robustesse des algorithmes de détection, PRAAT et MDVP donnant des résultats moins disparates que pour l'instabilité en hauteur.

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 de l'information qu'elles contiennent. Toutefois il existe peu de marqueurs validés sur cet aspect. Un travail d'investigation est donc à prévoir, ainsi qu'une validation portant sur un grand nombre de cas. De manière plus générale, le logiciel MDVP est le seul à fournir une palette assez riche d'indicateurs sur les phénomènes transitoires. Toutefois, certains phénomènes tels que l'éraillure sont peu étudiés car difficiles à détecter, et ne sont en outre pas assimilable simplement à du bruit.

Enfin, le **niveau de bruit** par rapport au signal semble, quel que soit l'algorithme choisi, un indicateur pertinent. Différentes limites en fréquences sont proposées selon les auteurs et les pathologies adressées. Un choix devra être fait selon la représentativité des pathologies en cabinet d'orthophonie.

Il existe encore peu de passerelles entre les outils, peu de précisions non plus sur les protocoles de passation des tests et d'exploitation des fichiers de voix, permettant de limiter la variabilité des mesures et garder une certaine fiabilité dans les interprétations. Une évaluation simple et factuelle des performances des outils est encore difficile à ce jour. De même, la capitalisation du savoir-faire, les explications sur les incohérences de calculs, limites des outils, conditions pour tirer le meilleur parti des évaluations objectives représente encore un vaste chantier, surtout à l'échelle des orthophonistes.

La validation passe par une vérification des calculs sur un nombre conséquent de cas de voix normales et pathologiques, en variant les conditions d'enregistrement, les logiciels étant utilisés en cabinet d'orthophonie dont l'ambiance sonore et la chaîne d'acquisition sont mal maîtrisés. L'utilisation d'une **base de données** de voix normales et pathologiques est à privilégier pour la validation, avec une ouverture aux orthophonistes permettant d'effectuer toutes les comparaisons

possibles. Aucune université numérique n'ayant encore vu le jour dans le domaine paramédical, le site du LURCO est à privilégier pour concentrer ce type de ressources, en attente de projets structurants à l'échelle francophone.

Les éléments scientifiques regroupés dans cette étude donnent un cadre méthodologique au développement de nouveaux indicateurs, dans le but de donner aux orthophonistes les moyens supplémentaires pour une rééducation efficace dans le cadre de pratiques probantes. Une validation soignée des outils passe par la construction d'une base de données ouverte de cas de voix normales et pathologiques, permettant les comparaisons entre outils, un accès partagé à l'information, mais aussi la capitalisation du savoir-faire dans l'utilisation des indicateurs selon les pathologies prises en charge.

Références

- BAKEN, R., J., ORLIKOFF, R. (1999). *Clinical measurement of speech and voice. San Diego : Singular.*
- BOERSMA, P., WEENINK, D., (2009) Praat: doing phonetics by computer, version 5.1.05
<http://www.fon.hum.uva.nl/praat/>
- CAMPISI, P., TEWFIK, T. L. MANOUKIAN, J. J., SCHLOSS, M. D., PELLAND-BLAIS, E., SADEGHI N. (2002) Computer-Assisted Voice Analysis, Establishing a Pediatric Database, Arch Otolaryngol Head Neck Surg/Vol 128
- CNIL, (2006). « Méthodologie de référence pour les traitements de données personnelles opérés dans le cadre des recherches biomédicales », MR-001, www.cnil.fr
- DE KROM, G. (1993). "A cepstrum based technique for determining a harmonics-to-noise ratio in speech signals," J. Speech Hear. Res., vol. 36, pp. 254–266.
- DEMENKO, G., (2004). Suprasegmental Analysis for Complex Quality Assessment in Pathological Voice, Speech Prosody, Nara, <http://www.isca-speech.org/archive>
- DELIYSKI, D.D. (2012). Multi-Dimensional Voice Program, Model 5105. *Software Instruction Manual. Kay Pentax Corporation. www.kaypentax.com*
- FNEO. <http://www.fneo.fr/index.php?page=8>, Annuaire des mémoires en ligne de la Fédération Nationale des Étudiants en Orthophonie
- GHIO, A. (2008) Bilan et perspectives de quinze ans d'évaluation vocale par méthodes instrumentales et perceptives. 27^{èmes} Journées d'Etude sur la Parole. Avignon: AFCP.
- GHIO, A. (2010), « Expériences et recommandations pour la structuration des données sonores, physiologiques et cliniques dans le cas des dysfonctionnements de la parole », XXVIIIèmes Journées d'Etude sur la Parole, Mons, 25 - 28 mai 2010. <http://aune.lpl.univ-aix.fr/~fulltext/4529.pdf>
- KASUYA, H., OGAWA, S., MASHIMA, K., EBHARA, S. (1988) "Normalized noise energy as an acoustic measure to evaluate pathologic voice," J. Acoust. Soc. Amer., vol. 80, pp. 1329–1334
- KENT, R.D. (2003). Voice dysfunction in dysarthria: Application of the Multi-Dimensional Voice Program. *Journal of Communication Disorders, 36 : 281–306.*
- KOREMAN, J., PITZER, M. (1997) Finding Correlates Of Vocal Fold Adduction Deficiencies, Institute of Phonetics, University of the Saarland, 1997, Phonus, N°3, pp. 155-178
- LALOT, A., REYT, S., (2008) "Production des voyelles /i a u/ : comparaison entre voix normales et pathologiques (voix soufflés de paralysies récurrentielles, papillomatose laryngée ...). Lien entre phoniatrie et production des voyelles." Maître(s) de mémoire : Jalal AL-TAMIMI, UCB, Lyon
- MARYN, Y. (2009). Perturbation Measures of Voice: A Comparative Study between Multi-Dimensional Voice Program and Praat. *Folia Phoniatri Logop, 61:217–226.*
- MENIN-SICARD, A., SICARD, E. (2009). Utiliser le logiciel VOCALAB dans le cadre de la prise en charge des pathologies de la voix. *La voix dans tous ses maux. Ortho-Editions.*
- ORLIKOFF, R. F., DELIYSKI, D., BAKEN, R.J., WA, B. C. (2009). Validation of a Glottographic Measure of Vocal Attack", Volume 23, Issue 2, pp. 164–168.
- PARSA, V., JAMIESON, D. G. (2001). Acoustic discrimination of pathological voice: Sustained vowels versus continuous speech. *J. Speech, Language Hear. Res., 44 : 327–339*
- PICCIRELLO J, COLIN P, DENNIS F, FREDERICKSON J. (1998) Multivariate analysis of objective vocal function. *Ann Otol Rhinol Laryngol.;107:107-112*
- PUTZER, M. (2001). Multiparametric description of voice quality for normal male and female voices (German). *Folia Phoniatri et Logopaedica, 53, 73–84.*

- QI, Y., HILLMAN, R. E., (1997). Temporal and spectral estimations of harmonics-to-noise ratio in human voice signals, *J. Acoust. Soc. Am.* 102 (1)
- REVIS, J. (1999). Comparison of different voice samples for perceptual analysis. *Folia Phoniatr Logop* 51 (3) : 108-116.
- ROBERT, M. (2011), Analyse acoustique des troubles articulatoires chez les patients atteints de paralysie faciale périphérique (VOT), Directeur de mémoire : Peggy Gatignol, Année : 2011
- SMITS, I., CEUPPENS, P., DE BODT, M. S. (2005), Comparative study of acoustic voice measurements by means of Dr. Speech and Computerized, *Speech Lab. J Voice* 2005; 19: 187–196.
- TESTON, B. (2004). « L'évaluation objective des dysphonies : Etat actuel et perspectives d'évolution », <http://hal.archives-ouvertes.fr>
- 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.
- WUYTS, F.L., et AL. (2000), The dysphonia severity index: An objective measure of vocal quality based on a multiparameter approach. *J Speech Hear Res.* 43:796-809
- XUE, S. A., DELIYSKI, D. (2001). Effects of aging on selected acoustic voice Parameters: preliminary normative data and Educational implications. *Educational Gerontology*, 27:159–168.
- 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*, 15:529-542
- YUMOTO, E., GOULD, W., AND BAER, T. (1982). Harmonics-to-noise ratio as an index of the degree of hoarseness. *J. Acoust. Soc. Am.* 71, 1544–1550.

Glossaire

Définition des indicateurs	Symbole	Utilisé dans
Aire Dynamique Vocale	ADV	EVA2, VOCALAB
Amplitude tremor intensity index	ATRI	MDVP
Amplitude perturbation quotient	APQ	MDVP, EVA2, PRAAT
Amplitude perturbation quotient 5 cycles	APQ5	PRAAT
Amplitude perturbation quotient 11 cycles	APQ11	PRAAT
Jitter percent	DDP	MDVP, EVA2
Degree of voice breaks	DVB	MDVP
Degree of subharmonic segments	DSH	MDVP
Degree of voiceless	DUV	MDVP
Etendue Vocale	EV	VOCALAB
Amplitude tremor frequency	FATR	MDVP
Average fundamental frequency	F0	MDVP, VOCALAB, PRAAT
Formant F1..F5	F1..F5	PRAAT
Fondamental Usuel Moyen	FUM	VOCALAB
Highest fundamental frequency	FHI	MDVP
Lowest fundamental frequency	FLO	MDVP
Fraction of locally unvoiced frames	FLUV	PRAAT
f0-tremor intensity index	FTRI	MDVP
Harmonic Noise Ration	HNR	EVA2
Absolute jitter	Jita	MDVP, PRAAT
Jitter percent	Jitt	MDVP, EVA2
Jitter ratio	Jitt	EVA2
Number of subharmonic segments	NSH	MDVP
Number of unvoiced segments	NUV	MDVP
Number of voice breaks	NVB	MDVP
Normalized Noise Energy	NNE	EVA2
Noise-to-harmonics ratio	NHR	MDVP
Pitch perturbation quotient	PPQ	MDVP, PRAAT
Pitch perturbation quotient 5 cycles	PPQ5	PRAAT
Pitch perturbation quotient 11 cycles	PPQ11	PRAAT
Phonatory f0-range in semitones	PFR	MDVP
Relative average perturbation	RAP	MDVP, EVA2, PRAAT
Standard deviation of f0	STD	MDVP
Smoothed amplitude perturbation quotient	sAPQ	MDVP
Shimmer in dB	ShdB	MDVP, PRAAT
Smoothed pitch perturbation quotient	sPPQ	MDVP
Soft phonation index	SPI	MDVP
Shimmer absolute	Shima	PRAAT
Shimmer percent	Shim	MDVP
Temps Phonatoire Maximum	TPM	VOCALAB
Temps Maximum de Phonation	TMP	EVA2
Average pitch period To	T0	MDVP
Length of analyzed sample	Tsam	MDVP
Fundamental frequency variation	vf0	MDVP
Peak-amplitude variation	vAm	MDVP
Voice turbulence index	VTI	MDVP