

Détection de la dysphonie parkinsonienne : pertinence de la mesure d'harmonicité de la voix

Sahar HAFSI^{1,2} Gaël MAHÉ² Linda MARRAKCHI-KACEM¹ Farouk MHAMDI¹ Nicolas LOMÉNIÉ² Sonia DJAZIRI-LARBI¹

¹Laboratoire Signaux & Smart Systèmes, École Nationale d'Ingénieurs de Tunis, Université de Tunis El Manar, Tunisie

²Laboratoire d'informatique Paris Descartes, Université Paris Cité, Paris, France

Résumé – La dysphonie est un symptôme vocal courant chez les patients atteints de la maladie de Parkinson. Elle est caractérisée par des anomalies de fonctionnement du vibrateur laryngé qui peuvent toucher la hauteur, l'intensité et le timbre de la voix. Plusieurs mesures acoustiques ont été utilisées pour la détection automatique de ces troubles vocaux, à l'instar du ratio harmoniques sur bruit (Harmonics to Noise Ratio, HNR), pour évaluer la qualité de la voix en comparant la quantité de bruit présent dans le signal vocal à celle de sa composante harmonique. Bien que le HNR soit largement utilisé dans ce contexte, la mesure de platitude spectrale (Spectral Flatness Measure, SFM) est une technique alternative qui capture la même information d'harmonicité. Toutefois sa pertinence clinique reste une piste de recherche. Dans cet article, nous comparons le pouvoir discriminant du HNR et de la SFM pour détecter automatiquement la dysphonie parkinsonienne, tout en les combinant avec d'autres descripteurs vocaux extraits à partir de différentes voyelles. Nous montrons que la SFM est un descripteur plus pertinent que le HNR pour détecter la dysphonie.

Abstract – Dysphonia is a common vocal symptom in patients with Parkinson's disease. It is characterized by abnormalities in the functioning of the laryngeal vibrator which can affect the pitch, intensity and timbre of the voice. Several acoustic measures have been used for the automatic detection of these voice disorders, such as the Harmonics to Noise Ratio (HNR) to assess voice quality by comparing the amount of noise present in the voice signal to its harmonic component. Although HNR is widely used in this context, Spectral Flatness Measure (SFM) is an alternative technique that captures the same harmonicity information. However, its clinical relevance remains an area of research. In this article, we compare the discriminating power of HNR and SFM to automatically detect Parkinsonian dysphonia, while combining them with other vocal descriptors extracted from different vowels. We show that the SFM is a more relevant descriptor than the HNR to detect dysphonia.

1 Introduction

La maladie de Parkinson (MP) est la deuxième maladie neuro-dégénérative la plus fréquente dans le monde, qui affecte le système nerveux central [1]. Son diagnostic est généralement basé sur l'observation médicale de signes cliniques spécifiques principalement moteurs. Cependant, certains symptômes comme l'altération de la voix se manifestent à un stade précoce et de manière subtile, ce qui rend leur observation clinique difficile [2]. Des études [3] rapportent qu'environ 90 % des patients atteints de la MP ont une forme de déficience vocale. Ainsi, pour aider les cliniciens à poser un diagnostic précoce, les approches d'apprentissage automatique (AA) ont été largement appliquées à différents signaux physiologiques [4], dont la voix.

L'intérêt croissant pour la voix en tant qu'un biomarqueur de la MP s'explique par sa mesure non invasive, et la disponibilité de plusieurs outils puissants de traitement du signal pour en extraire des informations pertinentes. Des études récentes sur la classification de la voix parkinsonienne ont été faites en combinant plusieurs types de données vocales (texte, syllabes, chiffres), diverses familles de descripteurs (de fréquence, d'amplitude), tout en utilisant des techniques de sélection de données (SD) pour réaliser un modèle d'AA robuste. Plusieurs mesures acoustiques de la voix ont été étudiées, notamment les perturbations d'amplitude et de fréquence (Jitter et Shimmer) et le rapport harmoniques sur bruit (HNR) [5]. Selon [6], le HNR est significativement réduit chez les patients atteints

de la MP par rapport aux sujets témoins, ce qui suggère une augmentation du bruit dans la voix parkinsonienne. De plus, le HNR diffère en fonction de la voyelle utilisée, il est plus élevé pour la voyelle /i/ par rapport à la voyelle /a/ [7], et il varie en fonction du genre et de l'âge [8]. Une information similaire à celle portée par le HNR peut être mesurée par la platitude spectrale (SFM), utilisée classiquement en Music Information Retrieval et en codage audio [9], plus récemment et plus rarement en détection de la voix pathologique [10, 11].

La principale contribution de cet article est la comparaison entre le pouvoir discriminant du HNR et celui de la SFM. En effet, traditionnellement on utilise un de ces deux descripteurs -le plus souvent le HNR- combiné à d'autres paramètres qui mesurent les variations en termes de fréquence et d'amplitude pour détecter la dysphonie parkinsonienne. Mais, à notre connaissance, une telle comparaison entre la SFM et le HNR n'a pas été effectuée par des études antérieures. Ainsi, nous comparons leur pouvoir discriminant dans un contexte de classification de la voix parkinsonienne : d'abord indépendamment d'autres mesures acoustiques, puis en combinaison avec d'autres descripteurs de base et en appliquant une SD.

L'article est organisé comme suit : l'ensemble des données, les descripteurs vocaux, la méthode de SD et la classification sont introduits dans la section 2. La section 3 présente les résultats statistiques, les performances de classification et la discussion des résultats.

2 Méthodologie

2.1 La base de données utilisée

Notre intérêt est la détection de la dysphonie dans la voix parkinsonienne via l'analyse des descripteurs d'harmonicité. Ainsi, nous avons choisi d'utiliser une base de données (BD) vocale italienne [12], qui contient différents types d'enregistrements de la parole : la lecture d'un texte, des syllabes et la phonation soutenue des voyelles /a/ et /i/, avec une fréquence d'échantillonnage de 16 kHz. Elle comprend des enregistrements de 19 contrôles sains (CS) et 28 patients parkinsoniens (PP). Nous avons sélectionné pour chaque individu un seul enregistrement par voyelle, donnant une BD de taille 47 échantillons pour chaque voyelle. Cette étude s'intéresse à la dysphonie, nous nous sommes donc limités aux enregistrements des voyelles soutenues, auxquels un prétraitement a été appliqué : suppression du silence au début et à la fin des enregistrements et sous-échantillonnage à 8 kHz.

2.2 Descripteurs vocaux

Bien qu'il existe une longue liste de descripteurs vocaux dans les travaux de détection de la voix pathologique, nous avons choisi de limiter l'étude aux descripteurs de base résumés dans le tableau 1. Les différentes mesures de Jitter (local, absolu...) et celles de Shimmer et leurs méthodes de calcul sont expliquées dans [5]. Ces mesures ont été extraites à l'aide de la bibliothèque Python Parselmouth [13].

2.2.1 Rapport harmoniques sur bruit (HNR)

Le HNR est une évaluation du rapport entre les composantes périodiques et les composants non périodiques d'un segment de parole voisée [14], exprimé en dB. La première composante provient de la vibration des cordes vocales et la seconde du bruit glottal. Il est aussi utilisé comme mesure de la qualité de la parole. Dans ce travail, nous avons utilisé le HNR basé sur le calcul d'autocorrélation [15]. Selon la littérature, cette mesure est très sensible à la taille du segment vocal, au type de fenêtre d'analyse, au nombre de cycles du pitch, voire au protocole d'acquisition de la voix [15] [16].

2.2.2 Mesure de la platitude spectrale (SFM)

La SFM est le rapport de la moyenne géométrique à la moyenne arithmétique du spectre d'amplitude du signal [17]. Ce ratio est d'autant plus faible que le signal est tonal (spectre composé de raies) : il quantifie donc le caractère harmonique (SFM proche de zéro) ou bruité (SFM proche de 1) du signal. La SFM est une mesure couramment utilisée dans la caractérisation des signaux audio ; néanmoins elle est moins répandue dans la détection des pathologies vocales : hormis quelques études, comme dans [10, 11] par exemple, on lui préfère généralement le HNR. D'où l'intérêt de comparer leur efficacité dans un contexte de voix pathologique.

2.3 Sélection des descripteurs

La Sélection des descripteurs (SD) est un processus d'identification des descripteurs les plus pertinents pour le modèle

TABLE 1 : Les descripteurs vocaux des voyelles soutenues

Descripteurs	Paramètres
F0 (MeanF0, StdevF0)	Pitch
HNR, SFM	Harmonicité
Jitter (Local, absolute, rap, ppq5, ddp)	Fréquence
Shimmer (Local, LocaldB, apq3, apq5, apq11, dda)	Amplitude

prédicatif, afin de réduire sa complexité et d'améliorer sa performance, puisqu'en pratique, l'utilisation d'un grand nombre de paramètres pour une BD de taille réduite surentraîne le modèle et peut générer une baisse des performances [18]. Dans ce papier, nous avons utilisé le test statistique de Mann Whitney (MW) pour éliminer les descripteurs redondants, et identifier ceux pour lesquels il existe une différence significative entre les deux groupes CS et PP.

2.4 Classification

Nous avons réalisé différentes expériences pour la classification de la voix parkinsonienne à partir des deux voyelles /a/ et /i/. D'abord, une classification basée sur : soit les descripteurs sélectionnés par MW (Σ MW), soit le HNR, soit la SFM dans le but de comparer leur pertinence. Par la suite, nous avons combiné séparément le HNR et la SFM avec (Σ MW). Deux classifieurs ont été choisis : k-plus proches voisins (k-Nearest Neighbors ou KNN) et machine à vecteurs de support (Support Vector Machine ou SVM). 70% de la BD a été utilisée comme un ensemble d'entraînement, et 30% en tant qu'ensemble de test. Les descripteurs ont été normalisés afin de les ramener à une échelle commune. La recherche des hyperparamètres de chaque modèle est effectuée avec Grid-Search (GS), et dans cette étape d'optimisation, nous avons choisi la validation croisée de type Leave One Out (LOO). Pour l'évaluation des performances, la précision et le score-F1 ont été choisis comme métriques. En suivant cette démarche, on peut s'assurer que nos données de test restent non vues par les modèles d'AA et éviter le risque de biaiser les résultats, comme expliqué dans la figure 1.

Base de données voix

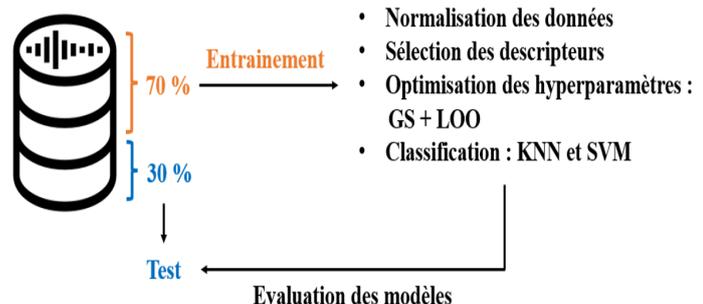


FIGURE 1 : Approche globale de la classification

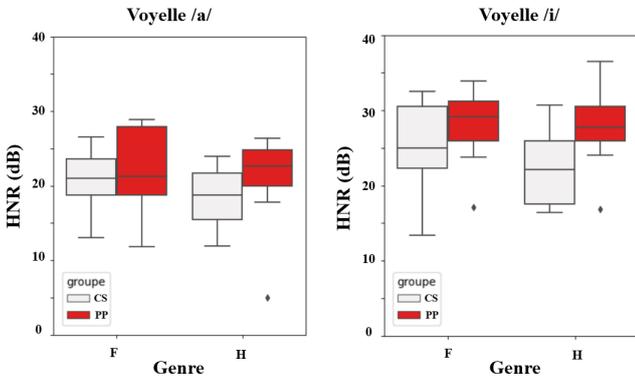


FIGURE 2 : HNR (dB) en fonction du groupe et du genre.

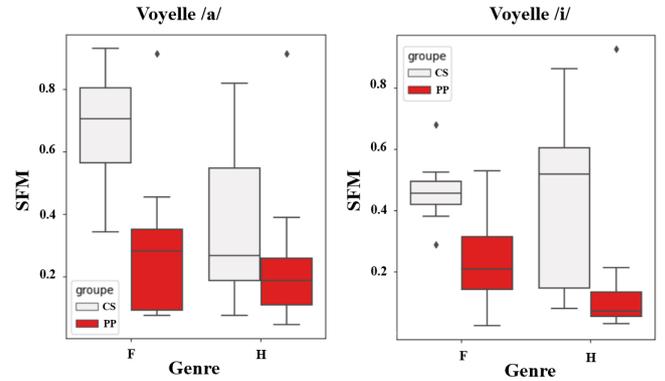


FIGURE 3 : SFM en fonction du groupe et du genre.

3 Résultats

3.1 Analyse descriptive

Nous avons utilisé les boîtes à moustaches (boxplots) pour visualiser les variations entre la classe à prédire (CS et PP) et les deux descripteurs d'intérêt (SFM et HNR) en fonction de la voyelle, de l'état et du genre des participants. D'après la figure 2, le HNR est plus élevé chez les PP, avec une différence plus significative pour la voyelle /i/. Il varie aussi en fonction du genre : le HNR des PP féminins est plus élevé que celui des PP masculins. Ces observations montrent une contradiction entre les variations du HNR extrait à partir de la BD vocale [12] et celles dans la littérature, où le HNR des PP est généralement plus faible que celui des CS. Ceci pourrait s'expliquer par la sensibilité de ce paramètre aux différentes conditions d'acquisition et de prétraitement des signaux vocaux, comme évoqué dans [15]. D'après la figure 3, la SFM des CS est plus élevée que celle des PP pour les deux voyelles /a/ et /i/. Elle dépend aussi du genre et de la voyelle : la SFM des femmes est plus élevée que celle des hommes pour la voyelle /a/. Cependant, elle est plus élevée chez les CS masculins par rapport aux CS féminins pour la voyelle /i/. Au terme de cette analyse descriptive, la SFM semble un descripteur plus pertinent que le HNR pour la détection de la dysphonie parkinsonienne.

TABLE 2 : Les p-values des descripteurs les plus significatifs

Descripteurs	Voyelle /a/	Voyelle /i/
HNR	0.01925	0.01132
SFM	0.00810	0.00041
Σ MW = descripteurs sélectionnés par MW		
LocalShimmer	0.00096	0.00214
LocaldBShimmer	0.00096	0.00188
Apq3Shimmer	0.00313	0.00509
Apq5Shimmer	0.00055	0.004
Apq11Shimmer	0.00014	0.00096
DdaShimmer	0.00313	0.00041

3.2 Analyse statistique

Pour comparer le pouvoir discriminant du HNR et de la SFM pour la classification de la voix parkinsonienne, nous appliquons à ces paramètres le test de Mann Whitney. Ce test montre une différence statistique (à un risque d'erreur de 5%) entre les groupes CS et PP pour les descripteurs du tableau 2. La SFM est plus pertinente que le HNR, car elle présente des p-values plus significatives pour les deux voyelles. En outre, parmi les descripteurs du tableau 1, on trouve que ceux de la famille Shimmer présentent les valeurs les plus significatives, avec une différence entre les p-values en fonction de la voyelle. Cela peut s'expliquer par les différences acoustiques des voyelles soutenues utilisées, principalement leur contenu spectral.

3.3 Performances de classification

Pour évaluer et comparer la pertinence du HNR et de la SFM, nous avons réalisé 5 expériences de classification par KNN définies chacune par un ensemble spécifique de descripteurs, tel qu'indiqué dans la première colonne du tableau 3. Ces 5 expériences ont été répliquées avec une classification par SVM (tableau 3). Que ce soit seule ou combinée aux descripteurs issus du test de MW, la SFM permet de meilleurs scores de classification que le HNR quel que soit le classifieur utilisé. La meilleure performance a été obtenue par SVM sur la voyelle /i/ avec la combinaison Σ MW + SFM (précision de 88.27% et F1 score de 90.16%). Nous pouvons conclure de ces expériences que la SFM est plus pertinente que le HNR pour la détection de la dysphonie parkinsonienne.

4 Conclusion

Notre principale découverte est que la SFM est un descripteur vocal plus pertinent que le HNR pour l'analyse de la dysphonie parkinsonienne. En effet, la SFM présente des valeurs statistiques plus significatives que celles du HNR. Elle dépend de la voyelle utilisée : elle est plus discriminante pour la détection de la dysphonie en utilisant la voyelle /i/. L'une des limites de ce travail de recherche est le manque de bases de données vocales de la MP ouvertes et en libre accès. Toutefois, malgré la taille réduite de la BD utilisée, nous avons pu choisir deux modèles de classification adaptés à ces dimensions avec les métriques les plus appropriées. Dans les futures recherches, nous explorerons des méthodes alternatives de prétraitement

TABLE 3 : Performances de classification avec KNN et SVM (%)

	KNN				SVM			
	Voyelle /a/		Voyelle /i/		Voyelle /a/		Voyelle /i/	
	Précision	Score-F1	Précision	Score-F1	Précision	Score-F1	Précision	Score-F1
Σ MW	58.76	54.63	59.12	57.32	59.15	56.13	62.64	59.87
HNR	63.15	68.79	71.42	79.76	64.28	66.46	72.38	78.56
SFM	78.57	84.21	82.35	87.95	79.71	85.49	84.35	89.12
Σ MW et HNR	68.32	73.36	78.94	83.36	69.80	73.89	79.87	83.92
Σ MW et SFM	82.64	86.43	87.96	89.73	84.12	87.55	88.27	90.16

du signal de la parole ainsi que d'autres tâches vocales pour améliorer la précision globale de la classification de la voix parkinsonienne. Il serait également important d'étudier l'effet d'autres descripteurs acoustiques notamment les MFCC. Nous pourrions aussi explorer d'autres modèles de classification, utilisant plus de descripteurs, pour mieux identifier ce qui caractérise acoustiquement la dysphonie parkinsonienne, ce qui suppose de disposer d'une BD plus importante.

Cette étude a bénéficié du soutien financier du partenariat Hubert Curien « Utique » du Ministère de l'Europe et des Affaires Étrangères français et du Ministère de l'Enseignement Supérieur et de la Recherche Scientifique tunisien.

Références

- [1] L. Ali, S. U. Khan, M. Arshad, S. Ali, and M. Anwar. A Multi-model Framework for Evaluating Type of Speech Samples having Complementary Information about Parkinson's Disease. *ICECCE*, 2019.
- [2] J. Mei, C. Desrosiers, and J. Frasnelli. Machine learning for the diagnosis of Parkinson's disease : a review of literature. *Frontiers in aging neuroscience*, 13, 2021.
- [3] A. K. Ho, R. Ianseck, C. Marigliani, J. L. Bradshaw, and S. Gates. Speech impairment in a large sample of patients with Parkinson's disease. *Behavioural Neurology*, 11, 1998.
- [4] S. Bhat, U. R. Acharya, Y. Hagiwara, N. Dadmehr, and H. Adeli. Parkinson's disease : Cause factors, measurable indicators, and early diagnosis. *Computers in Biology and Medicine*, 102, 2018.
- [5] J. P. Teixeira, C. Oliveira, and C. Lopes. Vocal Acoustic Analysis : Jitter, Shimmer and HNR Parameters. *Procedia Technology*, 9, 2013.
- [6] J. Ruzs, R. Cmejla, H. Ruzickova, and E. Ruzicka. Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease. *The Journal of the Acoustical Society of America*, 129, 2011.
- [7] N. D. Pah, M. A. Motin, and D. K. Kumar. Phonemes based detection of parkinson's disease for telehealth applications. *Scientific Reports*, 12, 2022.
- [8] C. T. Ferrand. Harmonics-to-Noise Ratio : An Index of Vocal Aging. *Journal of Voice*, 16, 2002.
- [9] E. Allamanche, J. Herre, O. Hellmuth, B. Fröba, T. Kastner, and M. Cremer. Content-based identification of audio material using mpeg-7 low level description. In *ISMIR*, 2001.
- [10] E. M. Zdancewicz, A. Potulska-Chromik, M. Nojszewska, and A. K. Pruszczyk. Parametric analysis of pilot voice signals in Parkinson's disease diagnostics. *Journal of Automation, Electronics and Electrical Engineering*, 4, 2022.
- [11] E. Vaiciukynas, A. Verikas, A. Gelzinis, and M. Bacauskiene. Detecting Parkinson's disease from sustained phonation and speech signals. *PLOS ONE*, 2017.
- [12] G. Dimauro, V. Di Nicola, V. Bevilacqua, D. Caivano, and F. Girardi. Assessment of speech intelligibility in parkinson's disease using a speech-to-text system. *IEEE Access*, 5, 2017.
- [13] Y. Jadoul, B. Thompson, and B. de Boer. Introducing Parselmouth : A Python interface to Praat. *Journal of Phonetics*, 2018.
- [14] P. J. Murphy and O. O. Akande. Cepstrum-Based Estimation of the Harmonics-to-Noise Ratio for Synthesized and Human Voice Signals. 2005.
- [15] P. Boersma. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proceedings of the institute of phonetic sciences*, 1993.
- [16] J. Fernandes, F. Teixeira, V. Guedes, A. Junior, and J. Paulo. Harmonic to Noise Ratio Measurement - Selection of Window and Length. *Procedia Computer Science*, 2018.
- [17] N. Madhu. Note on measures for spectral flatness. *Electronics Letters*, 2009.
- [18] I. Guyon and A. Elisseeff. An introduction to variable and feature selection. *Journal of machine learning research*, 3, 2003.