Analyse des performances des algorithmes d'estimation de la fréquence fondamentale dans le cadre de la voix pathologique

Robin VAYSSE[1,2], Corine ASTÉSANO[2], Jérôme FARINAS[1]

[1] Institut de Recherche en Informatique de Toulouse, CNRS, Université Toulouse III, France [2] Laboratoire Octogone Lordat, Université Toulouse II, France

Introduction

La mesure de la **fréquence fondamentale (F0)** est un élément essentiel du traitement automatique de la parole

De nombreux **algorithmes d'estimation de la F0** existent

Leurs performances sur la parole saine sont bonnes, mais **certaines erreurs arrivent parfois**

- Une mauvaise **détection du voisement**
- Une **multiplication ou division par 2** de la F0

Les performances de ces algorithmes sur la **parole pathologique** a très peu été étudiée

- La pertinence des résultats est pourtant cruciale car la F0 est très utilisée pour étudier ces voix

Objectifs

Mesurer les performances d'une dizaine d'algorithmes d'estimation de la F0 sur des enregistrements de **personnes atteintes de cancers des voies aérodigestives** supérieures (VADS) ainsi que des personnes avec **Maladie de Parkinson** afin de connaître les algorithmes les plus adaptés pour de futures études sur ces pathologies

Corpus de données

24 enregistrements

- 8 personnes **saines**
- 8 personnes atteintes de **cancers VADS**
- 8 personnes atteintes de la **maladie de Parkinson**
- **Genres équilibrés** pour chaque groupe

La **F0 de référence** a été **obtenue manuellement**

- Estimation automatique via l'algorithme d'autocorrélation intégré à Praat (ACF)
- Correction de **l'alignement des pics glottaux** manuellement

Méthodologie

Le choix des algorithmes testés est basé sur une étude comparative d'algorithmes de F0 sur la parole bruité [7]

- De nouveaux algorithmes basés sur des **réseaux de neurones profonds (DNN)** ont été intégrés
- Une méthode basée sur la **valeur médiane de 5 algorithmes** a été testée

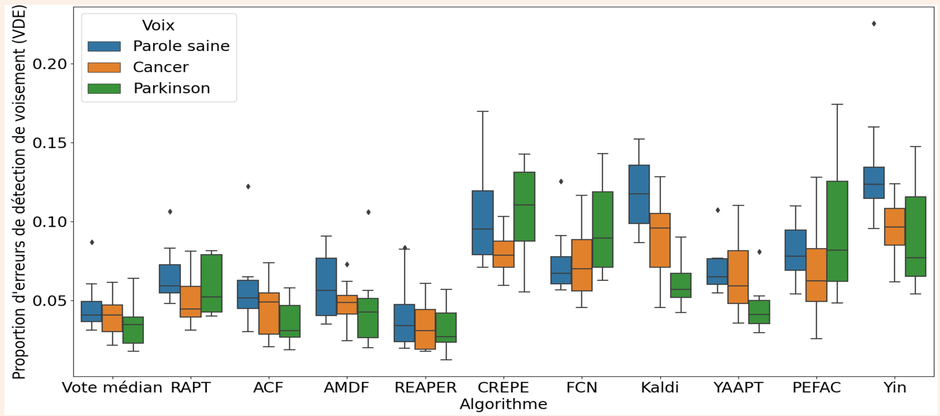
Les algorithme sont exécutés avec leurs **paramètres par défaut** en générant une estimation de la F0 sur des **fenêtres de 10 ms**

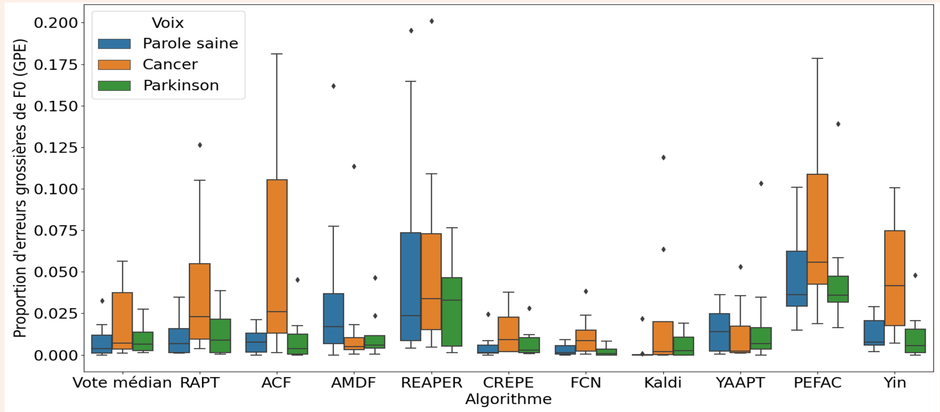
Métriques d'évaluation choisies :

- *Voicing Detection Error* (**VDE**) mesure la proportion de fenêtres de 10 ms avec une **erreur de détection de voisement**
- *Gross Pitch Error* (**GPE**) mesure la proportion de fenêtres contenant une **erreur d'estimation éloignée** de plus de 20 % de la valeur de référence
- *F0 Frame Error* (**FFE**) mesure la proportion de fenêtres contenant une erreur

Algorithm	Temporel	Spectral	Réseau de neurones
ACF [2]	X		
*AMDF [3]	X		
*REAPER	X		
RAPT [4]	X		
*Enhanced RAPT [5]	X		
Yin [6]	X		
NDF [7]	X	X	
*YAAPT [8]	X	X	
SWIPE [9]		X	
PEFAC [10]		X	
CREPE [11]			X
*FCN-F0 [12]			X
Vote médian (*)	-	-	-

Résultats





Discussion

Les **algorithmes basés sur le domaine temporel** du signal (ACF, AMDF, REAPER)

- Sont les meilleurs pour la détection de zones voisées avec environ **5 % de fenêtres contenant une erreur de voisement**
- Génèrent plus d'erreurs dans l'estimation des valeurs de F0 sur **la parole cancer**

Les algorithmes basés sur des **DNN**

- Génèrent de très **bonnes estimations de la valeur de F0** quel que soit le type de voix **(≈1 % d'erreurs)**

Le vote médian est un bon compromis au niveau de performances mais demande davantage de temps de calcul

Références

[1] JOUVET, D., and LAPRIE, Y. Performance analysis of several pitch detection algorithms on simulated and real noisy speech data, 2017, pp. 1614–1618
[2] BOERSMA, P. 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, 2000
[3] ROSS, M., SHAFFER, H., COHEN, A., FREUDBERG, R. and MANLEY, H., Average magnitude difference function pitch extractor, in IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 22, no. 5, pp. 353-362, 1974
[4] D. Talkin, "A Robust Algorithm for Pitch Tracking (RAPT)" in "Speech Coding & Synthesis", W B Kleijn, K K Paliwal eds, Elsevier ISBN 0444821694, 1995.
[5] GHAHREMANI, P., BABAALI, B., POVEY, D., RIEDHAMMER, K., TRMAL, J., and KHUDANPUR, S., A pitch extraction algorithm tuned for automatic speech recognition, pp. 2494–2902498, 2014
[6] DE CHEVEIGNÉ, A., KAWAHARA, H., YIN, a fundamental frequency estimator for speech and music. The Journal of the Acoustical Society of America, vol. 111,4 (2002), 1917-30.
[7] KAWAHARA, H., CHEVEIGNÉ, A., BANNO, H., TAKAHASHI, T., and IRINO, T. Nearly defect-free f0 trajectory extraction for expressive speech modifications based on straight, Interspeech 2005, pp. 537–540
[8] KASI, K., and ZAHORIAN, S., Yet another algorithm for pitch tracking, 2002, Vol. 1, pp.I–361
[9] CAMACHO, A., HARRIS, J., A sawtooth waveform inspired pitch estimator for speech and music, The Journal of the Acoustical Society of America, 124, 1638–52 (2008)
[10] GONZALEZ, S., and BROOKES, M., Pefac - a pitch estimation algorithm robust to high levels of noise, IEEE/ACM Transactions on Audio, Speech, and Language Processing 22(2), 518–530, 2014
[11] KIM, J. W., J. Salamon, P. Li and J. P. Bello, "Crepe: A Convolutional Representation for Pitch Estimation," 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, 2018, pp. 161-165
[12] ARDAILLON, L., ROEBEL, A., Fully-Convolutional Network for Pitch Estimation of Speech Signals, Proc. Interspeech 2019, doi:10.21437/Interspeech.2019-2815