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

Robin VAYSSE[1, 2]
Corine ASTESANO[2]
Jérôme FARINAS[1]
[1] Institut de Recherche en Informatique de Toulouse, CNRS, Université Paul Sabatier Toulouse III, France
[2] Laboratoire Octogone-Lordat, Toulouse, France

La mesure de la fréquence fondamentale (F0) est un élément essentiel du traitement automatique de la parole, notamment dans le cadre de l'étude de la prosodie. Il est crucial d'avoir une bonne estimation de ce paramètre. De nombreux algorithmes d'estimation de la F0 fournissent de bonnes approximations sur ~~de~~ la parole saine, cependant les performances de ces algorithmes ne sont pas connues dans le cadre de la parole pathologique. L'objectif ici est de tester plusieurs algorithmes sur des enregistrements de personnes atteintes de cancers des voies aérodigestives supérieures (VADS) ainsi que de la Maladie de Parkinson afin de savoir quels algorithmes sont les plus aptes à être utilisés pour de futures études sur ces pathologies.

Nous avons retenu 12 algorithmes de détection de F0 en se basant en partie sur une récente étude [9] ayant comparé ces algorithmes pour la parole bruitée. Nous avons ajouté plusieurs algorithmes basés sur des réseaux de neurones profonds ainsi qu'un vote médian entre plusieurs algorithmes (cf. tableau 1).—Les enregistrements sont issus du projet RUGBI, contenant des patients atteints de cancer VADS [15] et des patients atteints de la Maladie de Parkinson [8]. Nous avons sélectionné 24 enregistrements (8 sains, 8 cancers, et 8 Parkinson) correspondant à une tâche de lecture. Les enregistrements présentant les plus grosses déficiences au niveau de la F0 ont été choisis en se basant sur des annotations d'experts. La F0 de référence a été obtenue via une correction manuelle de

l'alignement des pics glottaux automatiquement annotés par le logiciel Praat [3] comme illustré sur la Figure 1.

Chaque algorithme a été évalué selon sa capacité à déterminer si une zone de parole est voisée ou non ainsi que selon sa capacité à calculer une estimation proche de la F0 de référence [9].

Les résultats obtenus sont décrits dans le Tableau 1. Les algorithmes se basant sur le domaine temporel du signal proposent de bons résultats sur la détection de voisement : ACF [2], AMDF [13] et REAPER (score aux alentours de 5% d'erreurs que ce soit pour la parole pathologique ou saine). Concernant la précision des estimations de la F0, ce sont les algorithmes basés sur des réseaux neuronaux qui procurent les meilleurs résultats avec environ 1% d'erreurs grossières sur la parole cancer pour FCN-F0 et moins de 0.5% sur la parole saine et Parkinsonienne. Le vote médian est le meilleur compromis entre détection de voisement et estimation de la F0.

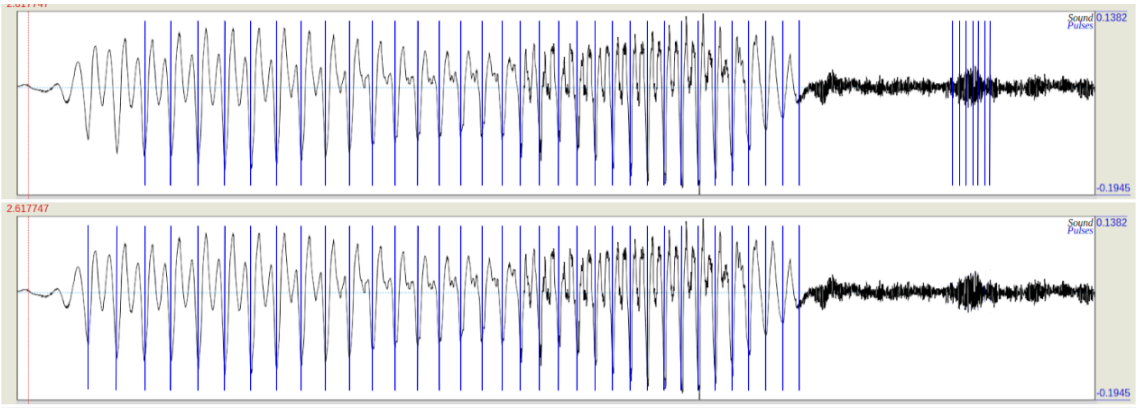

Figure 1. Exemple d'annotation de la F0 sur un extrait de signal d'une personne saine. La figure du haut correspond à l'annotation automatique des pics glottaux obtenus via l'algorithme d'autocorrélation (ACF) du logiciel Praat, celle du bas montre la correction manuelle.

Tableau 1. Liste des algorithmes d'estimation de la fréquence fondamentale testés ainsi que les résultats obtenus pour les différentes métriques obtenues en fonction des groupes

Algorithme	**Domaine du signal**	**Voicing Detection Errors (%)**			**Gross Pitch Errors (%)**		
		Saine	**Cancer**	**Park**	**Saine**	**Cancer**	**Park**
ACF [2]	**Temporel**	5.8	4.5	3.6	0.9	6	1
AMDF [13]	**Temporel**	5.9	4.7	4.6	3.9	2	1.2
REAPER	**Temporel**	**4.2**	**3.4**	**3.3**	5.9	6	3.1
RAPT [14]	**Temporel**	6.6	5	5.9	1.1	4.2	1.3
Enhanced RAPT [6]	**Temporel**	11.8	8.9	6	**0.3**	2.4	0.6
Yin [4]	**Fréquentiel**	13.5	9.6	9	1.2	4.6	1.2
NDF [11]	**Temporel et fréquentiel**	35.2	30.2	33.5	21	14	15.8
YAAPT [10]	**Temporel et fréquentiel**	7	6.4	4.6	1.5	1.3	2.1
SWIPE [5]	**Fréquentiel**	9	8.1	6.2	33.3	31.3	40.1
PEFAC [7]	**Fréquentiel**	8.2	6.8	9.5	4.6	7.7	4.9
CREPE [12]	**Réseau de neurones**	10.4	7.9	10.6	0.5	1.4	0.7
FCN-F0 [1]	**Réseau de neurones**	7.5	7.4	9.6	**0.3**	**1.2**	**0.2**
Vote médian (AMDF, Kaldi, YAAPT, FCN, REAPER)		4.7	4	3.5	0.8	2	0.9

Références bibliographiques

[1] ARDAILLON, L., ROEBEL, A., Fully-Convolutional Network for Pitch Estimation of Speech Signals, Proc. Interspeech 2019, doi:10.21437/Interspeech.2019-2815

[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] BOERSMA, P., WEENINK, D. Praat: Doing phonetics by computer (version2726.1.16), 2020

[4] 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. doi:10.1121/1.1458024

[5] 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), https://doi.org/10.1121/1.2951592

[6] 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, doi:10.1109/ICASSP.2014.6854049.

[7] 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

[8] JANKOWSKI, L., PURSON, A., TESTON, B., VIALLET, F., Effets de la L-DOPA sur la dysprosodie et le fonctionnement laryngien de patients parkinsoniens, Journées d'Etude sur la Parole 2004 (JEP), pp. 285-288

[9] JOUVET, D., and LAPRIE, Y. Performance analysis of several pitch detection algo-304rithms on simulated and real noisy speech data, 2017, pp. 1614–1618, doi:10.23919/EUSIPCO.3052017.8081482

[10] KASI, K., and ZAHORIAN, S., Yet another algorithm for pitch tracking, 2002, Vol. 1, pp.I–361, doi:10.1109/ICASSP.2002.5743729

[11] 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

[12] 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, doi: 10.1109/ICASSP.2018.8461329.

[13] 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, doi: 10.1109/TASSP.1974.1162598.

[14] TALKIN, D., A Robust Algorithm for Pitch Tracking (RAPT), in Speech Coding and Synthesis, W. B. Kleijn and K. K. Palatal, pages 497-518, Elsevier Science B.V., 1995

[15] WOISARD, V., ASTÉSANO, C., BALAGUER, M. et al. C2SI corpus: a database of speech disorder productions to assess intelligibility and quality of life in head and neck cancers. Lang Resources & Evaluation, 2020, https://doi.org/10.1007/s10579-020-09496-3