ACOUSTIC DIAGNOSIS OF VOCAL TREMOR

[1,2]Huawei Colin Li, [3]Hilmi R. Dajani, [1,2]Willy Wong and [1]Pascal van Lieshout
[1]Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, ON
[2]Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON
[3]School of Information Technology and Engineering, University of Ottawa, Ottawa, Canada

1. INTRODUCTION

Tremor is an unintentional muscle movement involving oscillations of one or more parts of the body. Vocal tremor is known to be associated with neurological disorders such as Parkinson's disease, essential tremor, and cerebellar ataxia. These vocal tremors are associated with a low-frequency modulation of the vocal folds. Characterizing the frequency, level and intensity of these modulations may help in diagnosing and treating a disorder The primary non-invasive quantification of vocal tremor has been through acoustic analysis, which involves measuring the variation in the pitch contour or the fundamental frequency in a sustained voice signal.

While much work towards measuring the variation of the pitch contour using time domain techniques has been done [1-3], these methods are not able to classify abnormal phonations accurately. There are two important limitations associated with the current methods: (a) they cannot extract the fundamental frequency in an accurate and precise manner; (b) they have not provided a fine structural display of the variation in the acoustic energy across frequency domain with respect to time. To overcome these limitations, a method involving time-frequency analysis using the Fine Structure Spectrogram (FSS) [4] is proposed as a potential analysis tool for abnormal phonations, in particularly, Parkinson's disease. Results between normal and abnormal phonations are compared.

2. METHOD

Vocal tremor characteristics are best analyzed in terms of the modulations in the fundamental frequency (F0), i.e. the pitch contour. As shown in Figure 1, F0 of the acoustic signal of the sustained /a/ from pathological and normal voices can be first extracted using our custom-built speech processing algorithm: FSS. The design of FSS was inspired by the human auditory system and can provide a detailed picture of the fine-structure found in the time-frequency display of acoustic energy in a speech signal. The fundamental frequency is the component with lowest frequency in the display. More details about this approach can be found in [4]. In our implementation here, we first preprocessed the raw acoustic signal by performing half-wave rectification, then down-sampling to 1kHz to reduce computational time. The average F0 was then extracted and a band-pass filter with cut-off frequencies 15Hz above and below the average F0 is applied to eliminate undesired noise. In the next stage, we apply FSS to extract F0. We used a range of 25Hz below the average F0 to 25Hz above for FSS to extract the pitch contour. The fundamental frequency was then further processed with the following steps: (a) subtraction of the mean (b) removal of linear trend (c) smoothing through low-pass filtering (cut-off at 20Hz).

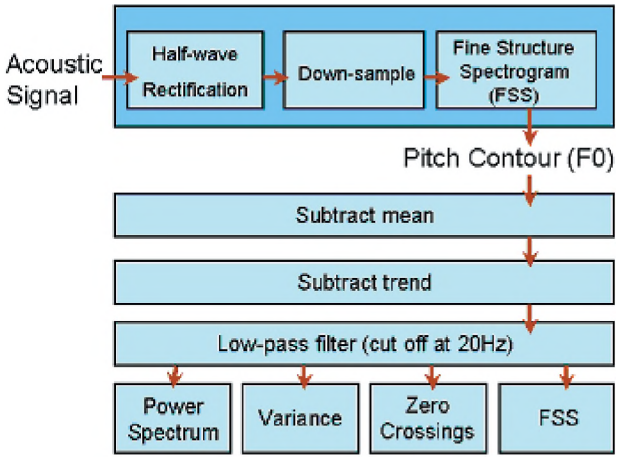

Fig. 1. Overview of the stages in vocal tremor analysis

Four analytical approaches were then applied to this output: (a) Power Spectrum (b) Variance of F0 across frequency (c) Zero-crossing technique (d) Spectrographic display of F0 using FSS. Power spectrum provides a display of the spectral density of F0. F0 variance is a single number quantitatively measuring the overall variability of the fundamental frequency. Zero-crossings measure is also a single number, which acts as an estimation of the most dominating modulating frequency of F0. Lastly, FSS provides a high-resolution time-frequency display of the acoustic energy in F0.

3. RESULTS AND DISCUSSION

The results show that modulations in F0 is the primary difference between normal and pathological speech. From the FSS output, consistent low frequency modulations in F0 over time were observed in the pathological phonations. Variance of the modulations was much higher in the pathological phonations than in the normal. Shown in Figure 2 is an example of the comparison between the two.

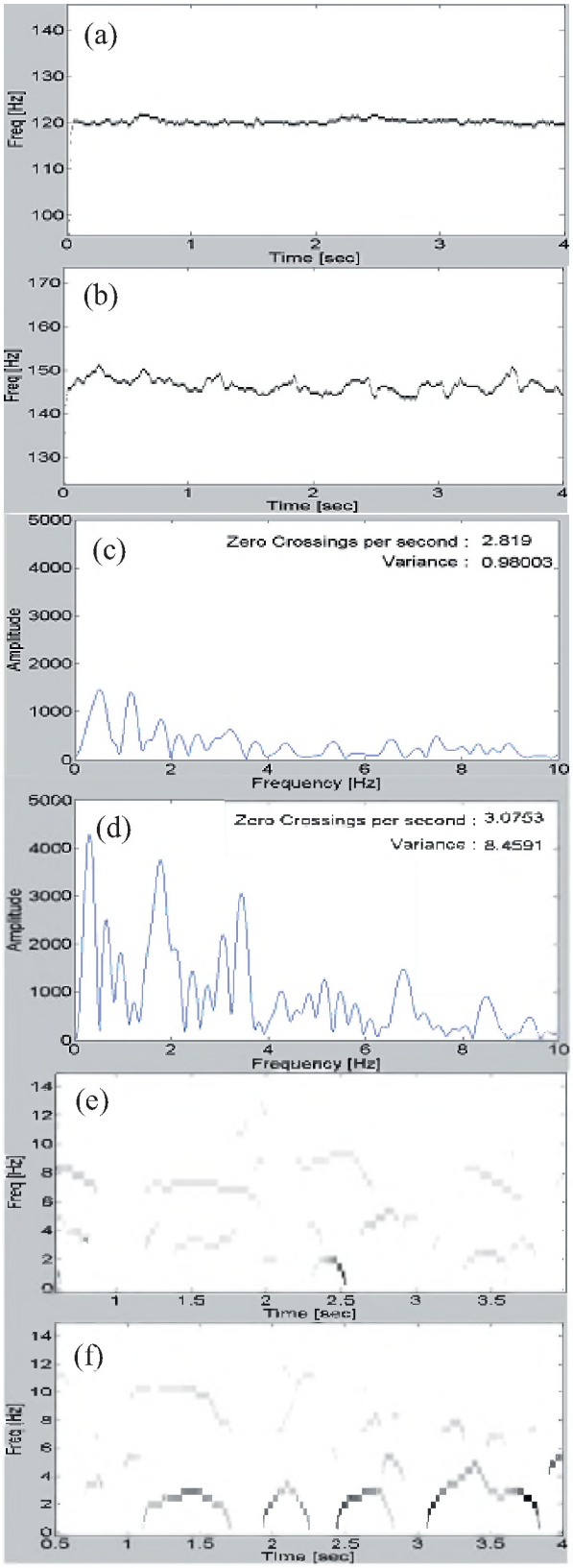

Fig. 2. (a) F0 of a normal phonation of /a/. (b) F0 of a pathological phonation of /a/. (c) Analysis of F0 of the normal phonation. (d) Analysis of F0 of the pathological. (e) FSS display of F0 of the normal. (f) FSS display of F0 of the pathological

Figure 2(a) and (b) show the F0 of a normal phonation and a pathological phonation respectively. It is easy to observe the low frequency modulation in F0. Figure 2(c) and (d) provide a more in-depth analysis showing spectral density, the dominant frequency and the variance in F0. The visible difference in the spectral density and variance offers a qualitative and quantitative cue in separating between the two types of speech. The zero crossings display also shows information, which may potentially be used to distinguish between various types of vocal tremors. Shown in Figure 2(e) and (f) are FSS display of the F0 from the normal and pathological phonations respectively. Through FSS, a relatively consistent modulating frequency is observed in the pathological phonation, in contrast to the normal case.

4. Conclusion

The method proposed here involving the Fine Structure Spectrogram is capable of extracting the fundamental frequency or pitch track contour in an accurate and precise manner. The technique shows the capacity of discriminating between two populations of speech (normal versus pathological) based on modulations found in F0. It is promising in future work to develop a classifier in distinguishing between normal and pathological phonations for a clinical environment.

REFERENCES

1. William S. Winholtz and Lorraine Olson Ramig (1992). "Vocal Tremor Analysis With the Vocal Demodulator" *American Speech-Language-Hearing Association* Vol. 35 562-573

2. Eugene H. Buder and Edythe A. Strand (2003). "Quantitative and Graphic Acoustic Analysis of Phonatory Modulations: The Modulogram" *Journal of Speech, Language, and Hearing Research* Vol. 46, 475-490

3. Laurence Cnockaert, Francis Grenez, Jean Schoentgen (2005). "Fundamental Frequency Estimation and Vocal Termor Analysis by Means of Morlet Wavelet Transforms" *ICASSP'05*

4. Dajani, H., Willy Wong and Hans Kunov (2005). "Fine structure spectrogram and its application in speech" J. *Acoust Soc. Am.* Vol. 117, No. 6, 3902-3918

ACKNOWLEDGEMENTS

This work was carried out under the support of OGS and NSERC.