FUNDAMENTAL FREQUENCY ESTIMATION AND VOCAL TREMOR ANALYSIS BY MEANS OF MORLET WAVELET TRANSFORMS

Laurence Cnockaert[*]*, Francis Grenez*

Université Libre de Bruxelles
Department Waves and Signals
Brussels, Belgium

Jean Schoentgen[†]

Université Libre de Bruxelles
Laboratory of Experimental Phonetics
Brussels, Belgium

ABSTRACT

A vocal frequency estimation method and an analysis of vocal tremor are proposed. Vocal tremor is a narrow-band low-frequency perturbation of the vocal frequency. The vocal frequency estimate is the instantaneous frequency calculated in an automatically selected frequency-band of a wavelet transform of the speech signal. The vocal frequency estimation method is compared to an event-based method and a Hilbert-transform method for speech signals. The tremor frequency and amplitude are obtained by means of a continuous wavelet transform applied to the instantaneous frequency trace. The proposed analysis has been applied to parkinsonian and normophonic speakers. The results suggest that the vocal tremor features differ for both groups.

1. INTRODUCTION

Measuring the fundamental frequency F_0 accurately is still considered a difficult task in speech processing. However, it is necessary for applications like the study of vocal tremor. Vocal tremor is a narrow-band low-frequency perturbation of the fundamental frequency. For this application, it is necessary to be able to track small perturbations of the F_0, with an amplitude of a few percents and a variation frequency up to 20Hz. This implies that the time-resolution of the F_0 trace should be at least one value per vocal cycle.

The proposed analysis of vocal tremor has two stages. First, the F_0 trace is obtained by means of a continuous wavelet transform. Second, vocal tremor features are extracted by means of a second continuous wavelet transform.

The F_0 extraction method is compared to an event-based and to a Hilbert-transform-based method. It is shown that the proposed method gives a F_0 value for each time sample of the speech signal, does not need any prior estimation of the speakers average F_0 and is more adaptive than the Hilbert-transform-based method.

[*]The first author is a fellow with the FRIA (Belgium).

[†]The third author is a Senior Research Associate with the Fonds National de la Recherche Scientifique (Belgium).

The proposed analysis of vocal tremor has been carried out on normophonic and parkinsonian speakers. Results on the difference between both groups are presented.

2. FUNDAMENTAL FREQUENCY EXTRACTION

The instantaneous frequency $\omega(t)$ of a band-pass signal $s(t)$ may be defined using its Hilbert transform $H[s(t)]$ [1].

$$\omega(t) = \frac{d\Phi(t)}{dt} \quad (1)$$

$$\Phi(t) = arg[s(t) + jH[s(t)]] \quad (2)$$

The instantaneous frequency can also be defined using a continuous wavelet transform (CWT) [2]. The continuous wavelet transform of a signal $x(t)$ is defined as

$$CWT(\lambda, t) = \int_{-\infty}^{+\infty} x(u) \frac{1}{\sqrt{\lambda}} \psi^* \left(\frac{u-t}{\lambda} \right) du \quad (3)$$

where $\psi(t)$ is the wavelet function and $CWT(\lambda, t)$ is the wavelet transform coefficient for a scale factor λ, at time t.

For a CWT using analytical wavelets, the amplitude and phase of the CWT coefficients represent the envelope and instantaneous phase of the spectral components of the signal in the frequency band centered on the central frequency f_c of the wavelet [3]. The time-derivative of the phase of the complex CWT coefficients is therefore an estimate of the instantaneous frequency of the signal in that frequency band. The evolution of the instantaneous frequency in different frequency-bands of the signal can thus be studied by means of the CWT coefficients.

In the proposed F_0 extraction method, the complex Morlet wavelet is used:

$$\psi_{\omega_c}(t) = C\, e^{i\omega_c t} \left[e^{-\frac{t^2}{2\sigma_t^2}} - \sqrt{2} e^{-\frac{\omega_c^2 \sigma_t^2}{4}} e^{-\frac{t^2}{\sigma_t^2}} \right] \quad (4)$$

The scale of the wavelet is determined by the central frequency $f_c = \frac{\omega_c}{2\pi}$, which fixes the frequency of oscillation of the wavelet. The parameter σ_t fixes its decay. The

product $\omega_c \sigma_t$ must be constant for a wavelet family and is chosen equal to 5.The normalization factor C is chosen so that $\int_{-\infty}^{+\infty} |\psi_{\omega_c}(t)|^2 \, dt$ equals 1.

In the neighborhood of the wavelet frequencies that fit best the cyclicity of the signal, the amplitude of the CWT coefficients presents a maximum and the instantaneous frequencies obtained by means of the CWT phase coefficients are very close to the cyclicity of the signal [4]. For speech signals, the CWT coefficients will thus be maximum around the fundamental frequency.

Using the instantaneous frequency gives a better frequency resolution than the frequency step of the CWT evaluation [5]. Indeed, for each time sample, in the (wavelet central frequency - instantaneous frequency) plane, the values of the instantaneous frequency are very close to the fundamental frequency in a large frequency-band around the actual value. Fig. 1 shows the (wavelet central frequency - instantaneous frequency) plane for a given time sample, for a synthetic speech signal. The plain line is the instantaneous frequency, the dashed line is the bisecting line and the dotted line shows the wavelet central frequency for which the amplitude is maximal. Fig. 2 shows the fundamental frequency traces obtained by the maximum amplitude trace of the CWT and by the corresponding instantaneous frequency obtained for a sinusoidal signal which frequency is modulated by a sinus and for a wavelet frequency step of 5Hz. The reference is also plotted. The difference between the F_0 trace based on the instantaneous frequency and the reference F_0 trace is too small to be seen on this plot.

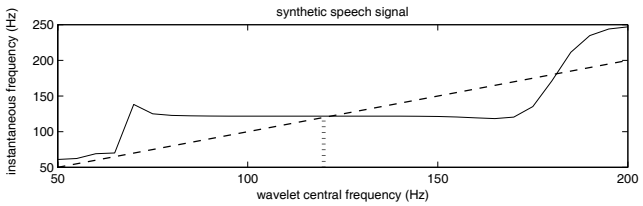

Fig. 1. The plain line is the instantaneous frequency, the dashed line is the bisecting line and the dotted line shows the wavelet central frequency where the amplitude is maximal.

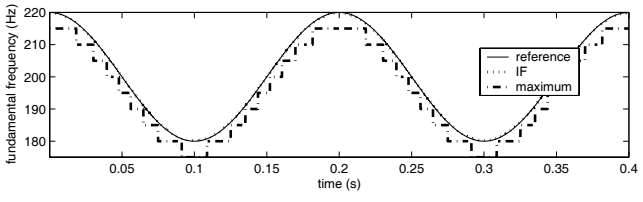

Fig. 2. Comparison of the frequency resolution of the F_0 traces based on the instantaneous frequency and on the maximal amplitude of the CWT, with the reference F_0 trace.

In the proposed method, the F_0 is estimated by the instantaneous frequency based on the phase of the CWT coefficients whose amplitudes are at a maximum in the interval from 50 Hz to 500 Hz.

3. FEATURE EXTRACTION

A second CWT is performed on the F_0 trace extracted in the first stage, to analyze the F_0 features. The tremor frequency and tremor amplitude are determined.

3.1. Tremor frequency

The perturbation of the vocal frequency usually presents more than one frequency component. To take all the frequency components into account, the tremor frequency is obtained by the weighted sum of all instantaneous frequencies higher than 1Hz, for which the amplitude of the CWT energy is higher than a threshold. The weight is given by the corresponding wavelet transform energy. An instantaneous tremor frequency can thus be obtained:

$$TF(t) = \frac{\sum_{f_c>1Hz}[CWT^2(2\pi f_c, t) IF(2\pi f_c, t)]}{\sum_{f_c>1Hz}[CWT^2(2\pi f_c, t)]} \qquad (5)$$

where $IF(2\pi f_c, t)$ is the instantaneous frequency based on the phase of the CWT coefficients.

3.2. Tremor amplitude

In the literature, the tremor amplitude is defined as the maximal or the standard deviation of the F0 trace, normalized by the average F_0 [6]. A definition of the tremor amplitude is proposed, based on the wavelet transform coefficients. An instantaneous energy value can be obtained by the sum of the energy density of the wavelet transform coefficients for each time sample. The square root of this instantaneous energy, normalized by the average F_0 is an estimation of the instantaneous tremor amplitude:

$$TA(t) = \frac{\sqrt{\sum_{f_c>1Hz} CWT^2(2\pi f_c, t)}}{\bar{F}_0} \qquad (6)$$

where $\bar{F}_0$ is the average F_0.

4. RESULTS

4.1. Fundamental frequency extraction

The precision of the proposed F_0 extraction method has been tested on sinusoidal signals which are modulated in frequency. Fig. 3 shows the maximum error of the proposed F_0 extraction method, as a function of the modulation amplitude, for different average values of F_0, for a sinusoidal signal modulated in frequency by a 5Hz sinusoid.

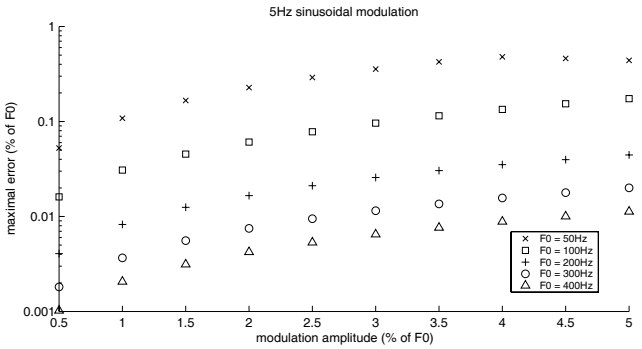

Fig. 3. Maximum error of the proposed F_0 extraction method, as a function of the modulation amplitude, for different average values of F_0 (50 to 400Hz).

The proposed F_0 extraction method has been compared to event-based and Hilbert-Transform-based F_0 extraction methods. Event-based methods extract cycle length time series from the positions of the positive-going zero-crossings that precede the main vocal cycle peak [6]. For the Hilbert-transform method [7], the speech signal is band-pass filtered around the fundamental frequency, the characteristic value of which must be estimated first. The instantaneous frequency trace obtained from the associated analytical signal is an estimate of the time-evolving F_0. The analyzed speech signals are sustained vowel segments [a], sampled at 25kHz.

Fig. 4 illustrates the wavelet-based method, an event-based and Hilbert-Transform-based method for a disordered speech signal. For clarity, the vertical scale of the lower figure has been dilated.

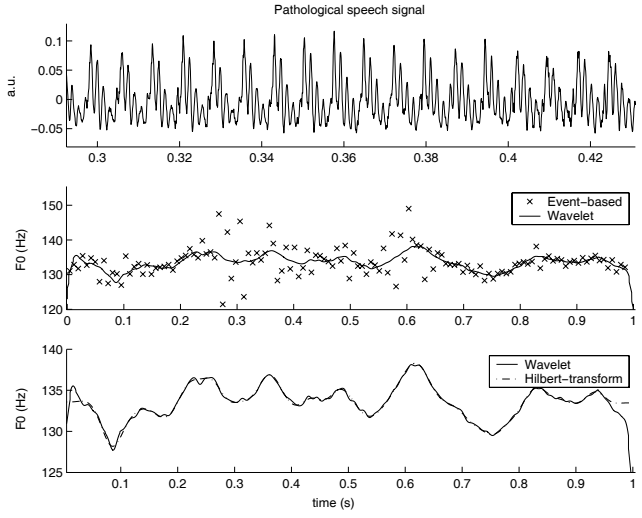

Fig. 4. Speech signal and fundamental frequency obtained by the event-based, wavelet-based and Hilbert-transform-based methods for a parkinsonian speaker.

4.2. Tremor features

The proposed analysis has been carried out on a corpus of 26 normophonic and 25 parkinsonian speakers (all male). The speech signals are 5-sec-long stable segments of sustained vowel [a], sampled at 25kHz.

Fig. 5 and Fig. 6 show the F_0 trace, the CWT^2 coefficient, the tremor frequency and the tremor amplitude for a parkinsonian and for a normal speaker.

The spectral energy distributions of the vocal frequency traces have been analyzed. Differences have been observed between the two speaker groups. To emphasize this difference, a ratio R of the spectral energy of the F_0 trace in the frequency-bands (1-5Hz) and (5-20Hz) has been calculated.

$$R = \sum_t \frac{\sum_{f_c=1Hz}^{5Hz} CWT^2(2\pi f_c, t)}{\sum_{f_c=5Hz}^{20Hz} CWT^2(2\pi f_c, t)} \qquad (7)$$

Fig. 7 shows the spectral energy ratio, as a function of the average fundamental frequency of the speaker.

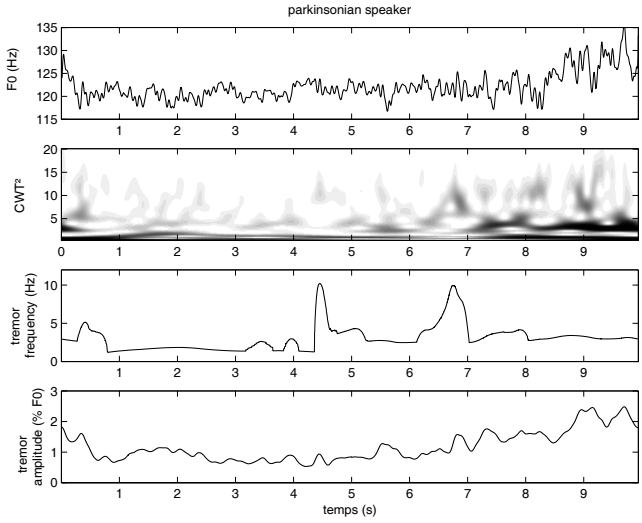

Fig. 5. Fundamental frequency trace, CWT^2 coefficients (high amplitude of the wavelet coefficients are represented in black, low amplitudes in white), tremor frequency and tremor amplitude for a parkinsonian speaker.

5. DISCUSSION

5.1. Fundamental frequency extraction

The comparison of the proposed F_0 extraction method with the event-based method shows that the F_0 traces obtained by both methods are close for normal speech signals. When comparing both methods, one must indeed take into account that the event-based method reports vocal jitter (wide-band perturbations), while the wavelet-based method does not.

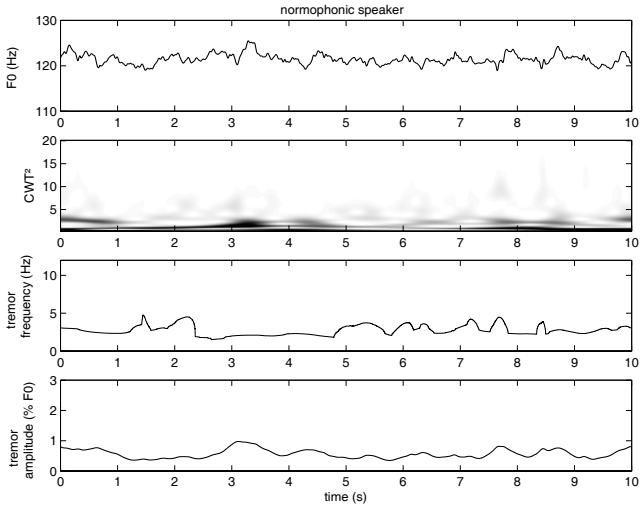

Fig. 6. Fundamental frequency trace, CWT^2 coefficients (high amplitude of the wavelet coefficients are represented in black, low amplitudes in white), tremor frequency and tremor amplitude for a normophonic speaker.

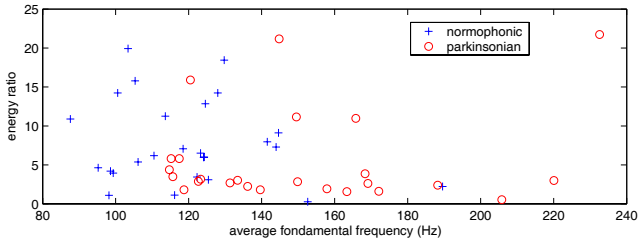

Fig. 7. Spectral energy ratio as a function of the average fundamental frequency.

However, in the case of the disordered speech signal (Fig. 4), vocal jitter does not explain all observed differences. Some discrepancies are due to the lack of robustness of the event-based method. For event-based methods, the time series of the cycle lengths are obtained by means of event markers that are not equidistant. The advantage of wavelet-based and Hilbert-Transform-based methods is that they enable the F_0 values to be computed for each time sample, i.e. at a constant step that agrees with the sampling step of the speech signal. Further spectral analysis can thus be carried out directly without intermediate processing. The F_0 traces obtained by the Hilbert-transform and wavelet methods are quasi-identical. However, the wavelet method has the advantage that the filter is adapted for each sample. It can track large variations of the F_0 and does not require any prior estimation of the speakers typical F_0. The wavelet-based method has been tested on disordered speech signals uttered by parkinsonian speakers. It has been able to track the perturbed vocal frequencies of these speakers.

5.2. Vocal tremor features

Differences have been shown in the spectral energy distributions of the fundamental frequency traces for normophonic and parkinsonian speakers. A ratio of the spectral energies of the F_0 trace in different frequency-bands has been calculated. Together with the average fundamental frequency, this spectral energy ratio seems to separate both groups. Further investigations are being made on that topic.

6. CONCLUSION

An analysis method of vocal tremor is proposed. The vocal frequency trace and the vocal tremor features are obtained by means of continuous wavelet transforms. The application of this analysis to parkinsonian and normophonic speakers suggests differences between both groups.

7. ACKNOWLEDGMENTS

The authors would like to thank Dr. C. Ozsancak and Dr. P. Auzou from the EA 2683, CHRU Lille, France and Dr. M. Jan from the CHU Rouen, France for providing the corpus.

8. REFERENCES

[1] B. Boashash, "Estimation and interpreting the instantaneous frequency of a signal - part i : Fundamentals," *Proceedings of the IEEE*, vol. 80, no. 4, pp. 520 – 539, 1992.

[2] T. Le-Tien, "Some issues of wavelet functions for instantaneous frequency extraction in speech signals," *Proc. IEEE Tencon 1997*, pp. 31–34, 1997.

[3] St. Mallat, *A Wavelet Tour of Signal Processing*, San Diego: Academic Press, 2nd edition, 1999.

[4] R. Carmona, W. Hwang, and B. Torresani, "Characterization of signals by the ridges of their wavelet transform," *IEEE Trans. on Signal Processing*, vol. 45, no. 10, pp. 2586 – 2590, 1997.

[5] H. Kawahara, H. Katayose, A. de Cheveigne, and R.D. Patterson, "Fixed point analysis of frequency to instantaneous frequency mapping for accurate estimation of f0 and periodicity," *Proc. Eurospeech*, pp. 2781–2784, 1999.

[6] J. Schoentgen, "Modulation frequency and modulation level owing to vocal microtremor," *J. Acoust. Soc. Am.*, vol. 112, no. 2, pp. 690 –700, 2002.

[7] W.S. Winholtz and L.O. Ramig, "Vocal tremor analysis with the vocal demodulator," *J. Speech Hear. Res.*, vol. 35, pp. 562–573, 1992.