

Analysis of vocal tremor by means of a complex wavelet transform

Laurence Cnockaert(1), Francis Grenez(1), Jean Schoentgen(2),
Canan Ozsancak (3) & Pascal Auzou (4)

(1) Department Signals and Waves, (2) Laboratory of Experimental Phonetics
Université Libre de Bruxelles, Brussels, Belgium

(3) Service de Neurologie A et EA 2683, CHRU de Lille, France

(4) Service d'explorations fonctionnelles neurologiques, Groupe HOPALE, Berk sur Mer et EA
2683, CHRU de Lille, France

lcnockae@ulb.ac.be

Abstract

A vocal frequency estimation method based on an analytical continuous wavelet transform is proposed, with a view to the study of vocal tremor. Vocal tremor designates a low-frequency narrow-band 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 analysis method is compared to an event-based method and a Hilbert-transform method for speech signals uttered by normal and Parkinsonian speakers. The results suggest that the ratio of the spectral energy of the vocal frequency trace in the intervals (1 – 5 Hz) and (5 – 20 Hz) differ for normophonic and Parkinsonian speakers.

1. Introduction

Measuring phonatory frequency accurately and capturing its small and rapid variations is still considered a difficult task in speech processing. In modal voice, two types of vocal frequency perturbations must be distinguished: wide-band perturbations designated as vocal jitter, and narrow-band perturbations designated as vocal tremor. Different origins are attributed to these perturbations, which therefore must be studied separately.

To study vocal tremor, it is desirable to be able to calculate the vocal frequency values for frames that are as short as possible. The duration of a frame should be about the length of one speech cycle. The cycle lengths or instantaneous frequency values are therefore obtained via event-based algorithms or via Hilbert-transforms of low-pass filtered speech [1].

A problem with event-based methods is that they may not be reliable when the speaker is severely hoarse. Also, to study vocal tremor frequency and thus the spectrum of cycle length time series, the lengths must be arbitrarily assigned to an instant in time: beginning, middle or end of cycle and the length time series must be interpolated and re-sampled to obtain a constant sampling step before its Fourier spectrum can be calculated by conventional methods.

The Hilbert-Transform can only be performed meaningfully for speech signals that have been band-pass filtered around the frequency of interest, that is, the speaker's typical phonatory frequency. Also, the cut-off frequencies of the band-pass filter must be fixed for a given analysis interval.

Here, a vocal frequency extraction algorithm is proposed that is based on a continuous wavelet transform. The reasons are the following. An estimate of the vocal frequency is

obtained for each sample of the speech signal. Also, the wavelet-based method is more robust than event-based or Hilbert-Transform-based methods and does not require any prior estimation of the speaker's characteristic vocal frequency.

The objective of the presentation is first to demonstrate that the obtained vocal frequency estimate is precise and suitable to study vocal tremor. The wavelet-based analysis is therefore compared to Hilbert-Transform-based and event-based methods. The comparisons are carried out on sustained vowels [a]. A second objective is the presentation of results regarding the low-frequency behaviour of the vocal frequency trace for normophonic and Parkinsonian speakers.

2. Wavelet Transform Method

In the proposed method, the complex Morlet wavelet is used (Figure 1). This wavelet has two parameters: the central frequency f_c which fixes the frequency of oscillation of the wavelet and the parameter σ_t which fixes its decay and its bandwidth and is related to f_c for a given wavelet family.

The instantaneous frequency of a signal may be defined as the derivative of the phase of its associated analytical signal [2]. The complex Morlet wavelet is an analytical wavelet and has the following properties [3]. The amplitude and phase of the wavelet transform coefficients represent the envelope and instantaneous phase of the spectral components of the signal in the frequency-band centred on frequency f_c , which is the centre frequency of the wavelet. The time-derivative of the phase of the complex wavelet coefficient is therefore an estimate of the instantaneous frequency of the signal in that frequency band. This enables the study of the time-evolution of the instantaneous frequency trace in different frequency-bands by tracking the time-evolving complex wavelet coefficients.

Consequently, the evolving instantaneous frequency of the complex wavelet coefficients whose amplitudes are at a maximum in the interval from 50 Hz to 500 Hz are assigned to the vocal frequency trace. The reason is that the time-evolving amplitude of the wavelet transform coefficient is maximal for those wavelets whose central frequency f_c fits best the cyclicity of the speech signal.

3. Results

3.1. Vocal Frequency Extraction

The wavelet-based analysis has been compared to event-based and Hilbert-Transform-based vocal frequency extraction

methods. Event-based methods extract cycle length time series from the positions of the positive-going zero-crossings that precede the main speech cycle peak [4]. For the Hilbert-transform method, the speech signal is band-pass filtered around the phonatory 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 vocal frequency. The speech signals that have been analyzed are sustained vowel segments [a], sampled at 20kHz or 25kHz.

Figure 2 illustrates the wavelet-based analysis and an event-based analysis for a normophonic speaker. The event markers are placed at the beginning of each cycle.

Figure 3 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.

3.2. Features

The spectral energy distributions of the vocal frequency traces have been analyzed for 10 Parkinsonian speakers and 8 normal speakers (all male). The speech signals are 10-second stable segments of sustained vowel [a], sampled at 25kHz. In Figure 4, the ratio of the spectral energy in the frequency band (1-5Hz) and the spectral energy in the frequency band (5-20Hz) are displayed together with the average vocal frequency for each speaker.

4. Discussion

A comparison of the proposed wavelet method with the event-based method shows that the vocal frequency traces obtained by both methods are close for the normal speech signal. When comparing both methods, one must indeed take into account that the event-based method reports vocal jitter, while the wavelet-based method does not. However, in the case of the disordered speech signal, vocal jitter does not explain all observed differences between traces. Some discrepancies are due to the lack of robustness of the event-based method, which is affected by additive noise owing to turbulence, for instance.

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 or Hilbert-Transform-based methods is that they enable the vocal frequency values to be computed for each time sample, i.e. at a constant step that agrees with the sampling step of the original speech signal. Spectral analysis can thus be carried out directly without prior processing.

The vocal frequency trace obtained by the Hilbert-transform and wavelet methods are quasi-identical. However, the wavelet method has the advantage that the best filter is obtained for each sample. This method is thus adaptive. It can track large variations of the vocal frequency and does not require any prior estimation of the typical phonatory frequency of the speaker. The wavelet-based method has been tested on disordered speech signals uttered by Parkinsonian speakers. It has been able to track the time-evolving perturbed vocal frequencies of these speakers.

5. Conclusions

A vocal frequency estimation method based on an analytical continuous wavelet transform has been presented. The low-

frequency spectral energy distribution of the vocal frequency traces thus obtained, related to the average vocal frequency, suggests differences between Parkinsonian and normophonic speakers that should be investigated further.

6. Acknowledgements

Laurence Cnockaert is a Fellow with the FRIA (Belgium). Jean Schoentgen is a Senior Research Associate with the National Fund for Scientific Research, Belgium.

7. References

- [1] Winholtz, W. S.; and Ramig, L. O., 1992. *Vocal Tremor Analysis with the vocal Demodulator*. *J. Speech Hear. Res.*, 35: 562--573.
- [2] Boashash, B., 1992. *Estimating and Interpreting the Instantaneous Frequency of a Signal - Part 1: Fundamentals, Proceedings of the IEEE*, vol. 80, n°4: 520--539.
- [3] Mallat, St., 1999. *A Wavelet Tour of Signal Processing*. San Diego: Academic Press, 2nd Ed.
- [4] Schoentgen, J.; and De Guchteneere, R., 1991. *An algorithm for the measurement of jitter*. *Speech Commun.*, 10:533--538.

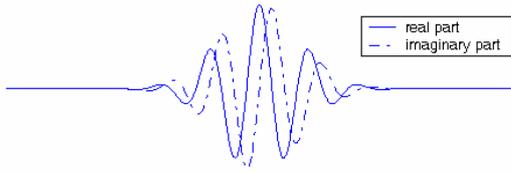


Figure 1: Complex Morlet wavelet for $(f_c \sigma_t)/(2\pi) = 5$.

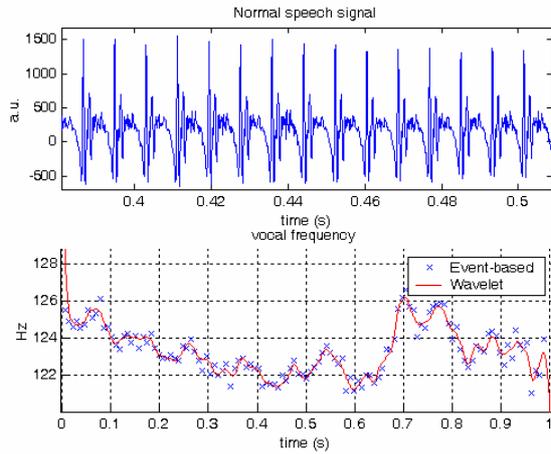


Figure 2: Vocal frequency trace of a normal sustained vowel.

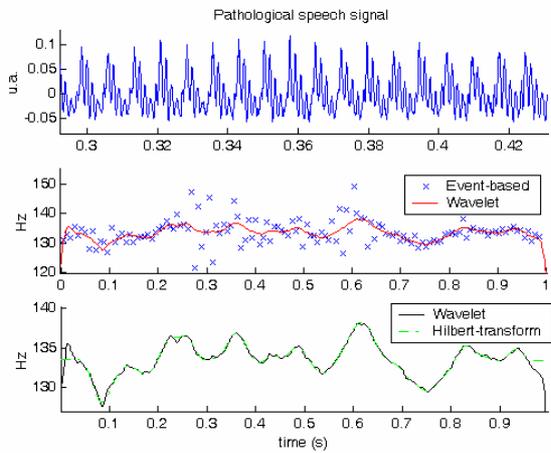


Figure 3: Vocal frequency trace of a disordered vowel signal sustained by a Parkinsonian speaker.

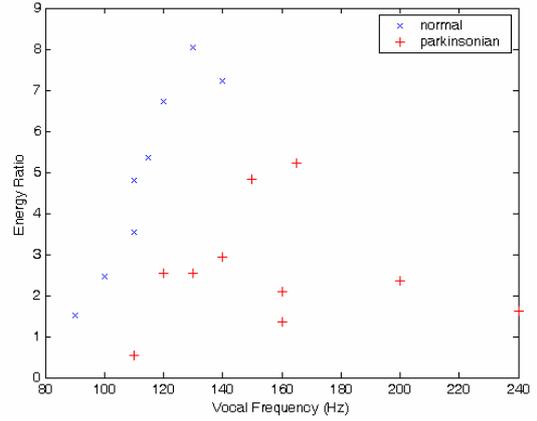


Figure 4: Spectral energy ratio as a function of the average vocal frequency. The spectral components of the vocal frequency, which are involved in the spectral energy ratio are comprised in the intervals (0-5Hz) and (5-20Hz) respectively.