IEEE Transactions on Information Theory VOL.38, NO. 2 March 1992 p917-924

Application of the wavelet transform for pitch detection of speech signals

Shubha Kadambe and G. Faye Boudreaux-Bartels

8913520 林建光

Background

- Wireless communication systems often use waveform coding before 1992. ex: take ADPCM as its speech codec.
- After 1992,systems change to use synthesis/analysis coding (parametric coding). Ex: take CELP、RELP-LTP、VSELP ect. as its speech codec.

Introduction

Pitch period information is used in various applications

- Speaker identification
- Pitch synchronous speech analysis and synthesis
- Linguistic and phonetic knowledge acquisition
- Voice disease diagnostics

Definition of problem

- The pitch detectors can be broadly classified into either **event detection pitch detectors** or **nonevent pitch detectors**
- These nonevent pitch detectors are computationally simple; However,they assume that pitch period is stationary within each segment and each segment contains at least two full pitch period

Definition of problem

Hence,the disadvantages of these pitch detectors are that

- Insensitive to nonstationary variations in the pitch period
- Unsuitable for both low pitched and high pitched speakers

Definition of problem

- Event detection pitch detectors estimate the pitch period by locating the instant at which the glottis closes (call an event).
- Some of Event detection pitch detectors are unsuitable for all vowels and for nonstationary pitch periods, some are applicable for "clean"data and certain vowels and some are not suitable for high pitched speakers.

Key to the solutions

- We describe an event detection pitch detector which is robust to noise and is suitable for for a wide range of pitch periods and for different speakers .
- We apply the **dyadic wavelet transform** (D_yWT) to the task of locating glottal closure.

dyadic wavelet transform

The D_yWT of a signal x(t)

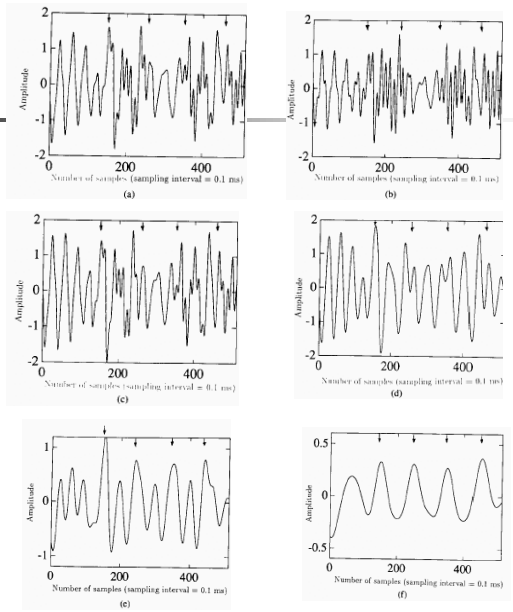
$$D_yWT_x(b,2^j) = \frac{1}{2^j}\int_{-\infty}^{\infty} x(t)g^*\left(\frac{t-b}{2^j}\right)dt$$
$$= x(t) \otimes g^*_{2^j}(t)$$

g(t) is wavelet function that satisfies some conditions

$\otimes$ Represents the convolution operator

Property of dyadic wavelet transform

- The D_yWT is linear and shift invariant
- If a signal x(t) or its derivatives have discontinuities, then the modulus of the DyWT of x(t), $|D_yWT_x(b,2^j)|$,exhabits local maxima around the points of discontinuity
- The local maxima of D_yWT indicate the sharp variations in the signal whereas the local minima indicate the slow variations. Hence, the local maxima of the D_yWT should be useful for detecting the abrupt changes in a speech signal caused by the glottal closure.



Concrete results

- Comparison of Performance
  - The accuracy with which the pitch periods can be estimated
  - The robustness of the algorithms to noise
  - The computational complexity of the algorithms
  - The problems associated with segmentation of a speech signal

Comparison of Performance

- We compare the performance of the DyWT event based pitch detector with standard nonevent based pitch detectors based on **cepstrum** and **autocorrelation** of a given signal

Comparison of Performance

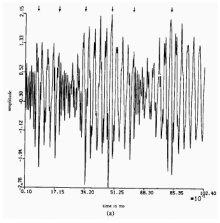
- The cepstrum of a signal x(t) is defined as

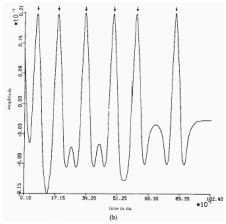
$$C_x(t) = \left[\int_0^{\infty} \log \left| X(w) \right|^2 \cos(wt)\, dw \right]^2$$

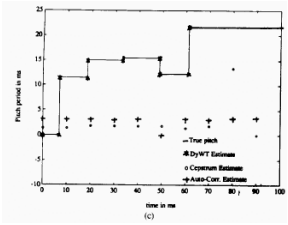
- The autocorrelation function Rx(τ) of a signal x(t) is defined as

$$R_x(\tau) = \int_{-\infty}^{\infty} x^*(t)\, x(t+\tau)\, dt$$

Accuracy







Robustness to Noise

- Define relative error =

  |true pitch period – estimated pitch period| / true pitch period

Percentage of Relative Error				
SNR	0 dB	-6 dB	-12 dB	-18 dB
D_yWT-based Pitch Detector	0.44%	0.44%	0.92%	1.1%
Cepstrum-based Pitch Detector	17%	20%	43%	76.5%
Autocottelation-based Pitch Detector	52%	54%	60.5%	80.3%

Computational Complexity

- The cepstrum method involves computing the FT,the logarithm of both the power spectrum and its inverse FT. The computation of both sampled autocorrelation and the DyWT methods involve only a summation of products.

Computational Complexity			
	Number of adds	Number of mults	Number of log ops
D_yWT Pitch Detector	2min(M-1,L-1)	2min(M,L)L	0
Cepstrum Pitch Detector	$O(2L\log_2 L)$	$2(O(L\log_2 L)+L)$	L
Autocorrelation Pitch Detector	(L-1)P	LP	0

O(x) means the order of x
L=segment length

M=length of the wavelet
P=number of correlation coefficients

Segmentation

- For the cepstrum and the autocorrelation methods,the choice of the segmentation length is very important. Both methods estimate the **average pitch period** of a length L signal and hence they need at least two pitch periods within a chosen segment.
- In the case of D_yWT,the choice of the segmentation is not very crucial,since we estimate the pitch period by locating the instant at which glottis closes.

Conclusion

In this paper,We have described an event based pitch detector using D_yWT and compared its performace. The main advantages of the proposed D_yWT method in comparison with the existing pitch detectors are the following :

- Does not assume stationarity or quasi-stationarity within the analysis window
- Estimates the pitch period very accurately
- Is suitable for a wide range of pitch periods

Conclusion

- Can detect the beginning of a pitch period and the number of pitch periods present in a given segment of a speech signal and, hence,can be used for pitch or event synchronous modeling applications
- Is computationally simple since we need to compute the D_yWT at only two or three scales
- Exhibits superior performance as compared to the autocorrelation and the cesptrum-based pitch detectors