Robust pitch detection for normal and pathologic voice

B. Boyanov†* , G. Chollet*
† Bulgarian Academy of Sciences, CLBA, Acad. "G. Bonchev" str. BLOC 105
Sofia 1113, Bulgaria
*TELECOM-Paris, CNRS URA-820
46, rue Barrault, 75 634 PARIS CEDEX 13, FRANCE

session **Speech, Hearing and Communication**

Introduction

It is known that most of the laryngeal pathologies produce a change in the vocal quality of the patient. The pitch period (To) is significantly affected by these diseases. In most of the pathological voices there are present:

a) large deviations of To and in magnitudes of the peaks of the pitch;

b) deformation of the shape of pitch impulses;

c) abrupt changes in To and the magnitude of the peaks of the pitch;

d) interruptions of pitch generation during sustained vowel phonation - voice breaks;

c) noisy components having a significant amplitude.

In order to overcome these difficulties a method is proposed for calculating To by analysis of different domains of the signal.

METHOD

The speech signal is analyzed by means of the following procedure:

Preprocessing of segments.

The signal is divided into segments with a duration of 30ms. In order to minimize errors caused by low level signals [1] a verification of the signal's level is carried out:

a) search for at least 3 peaks: $A_m(t1)$, $A_m t2)$, $A_m(t3)$, (where: t3 > t2 > t1) fulfilling the following conditions:

$A_m(t1) > TR$, $A_m(t2) > TR$, $A_m(t3) > TR$, (1)

where: TR is 50% from the maximum possible value of the signal.

and

$$t2 - t1 > T_{hp} \text{ and } t3 - t2 > T_{hp} \quad (2)$$

The distances between these peaks have to be more than the highest To possible (T_{hp}) for the pathologic voice.

b) the signal in the segment is classified as a normal level and is processed if at least 3 peaks fulfilling these conditions are found. Otherwise the segment is rejected and the next one is processed.

Pitch Period Evaluation.

The calculation of To is realized in parallel in 3 different domains:

1. PITCH EVALUATION IN TIME DOMAIN

1) To is calculated in the time domain using the autocorrelation function $R(\tau)$ [2,3]. The $R(\tau)$ is calculated over the center-clipped signal, allowing robust To detection from noisy speech [3]. However this method may give erroneous results due to [3] :

a) strong harmonics coinciding with the first formant;

b) strong harmonic and formant structure;

c) presence of several peaks in $R(\tau)$.

On the basis of the fact that $R(\tau)$t of a periodic signal is periodic the following procedure is used to minimize the above-mentioned errors:

1.1) Voiced-unvoiced detection by means of the algorithm described in[2];

1.2) In voiced segments the largest peak($R_{MAX}(\tau_{max})$) of the autocorrelation function in the range of To is found;

1.3) A threshold TRτ is calculated:

$$TR\tau = 0.6R_{MAX}. \quad (3)$$

This threshold is used because it was found [1] that for some pathological voices the peak in $R(\tau)$, corresponding to To is with reduced amplitude (nearly $0.6R_{MAX}$);

1.4) Location of all the peaks ($R_p(\tau j)$) of $R(\tau)$ in the range of To greater than TRτ;

1.5) Calculation of the differences (distances) between the lags of these peaks :

$$T(j) = \tau_{j+1} - \tau_j, \quad (4)$$

where: $\tau_0, \tau_1, \ldots \tau_J$ –successive lags of $R_p(\tau j)$,

j=0,1,...J,

J - number of the peaks,

τ_0=0.

1.6) Calculation of the maximal difference (δT) found between T(j);

1.7) Calculation of the mean T(j) - $\overline{T}$;

1.8) The value of the pitch is obtained in the time domain as To_{time} in the following cases:

1.8.1) If only one peak in $R(\tau)$ is found:

$$To_{time} = \tau_{max}. \quad (5)$$

1.8.2) The autocorrelation function is periodic i. e. the values of T(j) nearly constant:

$$To_{time} = \bar{T} \quad \text{if} \quad \delta T < 0.2\,To \qquad (6)$$

When no decision about To_{time} it taken then all T(j) are saved as possible To_{time} .

2. PITCH EVALUATION IN SPECTRAL DOMAIN

2.1) Calculation of the cepstrum (c(t));

2.2) Calculation of the smoothed spectrum by means of the group delay function (GDF) (the negative first derivative of the phase spectrum). The GDF is used because it was found [7] that it represents well the low and high energy spectral regions.

2.3) Coding the spectral components on the base of different thresholds for the low and high energy spectral regions:

a) For every spectral region are found the three largest peaks (X1(f1), X2(f2), X3(f3)), having a distance between them greater than the lowest fundamental frequency F_{low} for pathological voices;

$$f2 - f1 > F_{low} \quad \text{and} \quad f3 - f2 > F_{low} \qquad (7)$$

b) Calculation of a threshold for the region:

$$TR_{spec} = lev[X1(f1) + X2(f2) + X3(f3)]/3,$$

where: lev=0.7. (8)

c) Coding the spectral components on the base of the different TR_{spec} .

2.5) Calculation of a spectral autocorrelation function over the coded spectral components and evaluation of To using the procedure already described in the previous stage "analysis in time domain". If the segment is classified as voiced To is evaluated as To_{spect} or no decision about To is taken and P possible values for To_{spect} (where P-number of peaks in spectral autocorrelation function) are obtained

3. PITCH EVALUATION IN CEPSTRAL DOMAIN

To is evaluated in the cepstral domain using the robust method described in [2, 3]. The cepstral analysis is performed in order to compensate for inconveniances "b)" and "c)" of $R(\tau)$ [p. 405 in 3]. If the segment is classified as voiced the value of To is obtained in the cepstral domain - To_{ceps}.

4. OBTAINING THE PITCH PERIOD ESTIMATE

The calculated values of To are analysed for determination of To by means of the following procedure:

The segment is classified as unvoiced in the following cases:

a) In two domains it is classified as unvoiced;

b) In two domains there are no decision for the pitch period and in cepstral domain it is classified as unvoiced.

The difference between To_{spect} and To_{ceps} is calculated:

$$T11 = To_{ceps} - To_{spect} \qquad (9)$$

The segment is classified as unknown and is eliminated from future analysis if:

a) $T11 > 0.3.To_{spect}$ and $T11 > 0.3.To_{ceps}$,.

Here the results (To_{ceps} and To_{spect}) from the most robust pitch detectors are.used.

b) In two domains no decision for the pitch period is obtained;

c) In one domain it is classified as unvoiced and in one domain no decision for the pitch period. is obtained

In all the other cases To is calculated as:

$$To = [To_{ceps} + To_{spect} + To_{time}]/3 \qquad (10)$$

As a result the erroneous values of To are eliminated almost in all the possible cases.

Experimental research and results

The vowel "a" and a control phrase pronounced by 45 patients (laryngeal pathology) and 28 normal speakers are analysed. Normal and pathologic voice signals were passed through a low-pass filter with a cutoff frequency of 5kHz and sampled at 16 kHz with 16 bits directly into the computer's memory. No large errors in the values of calculated To and no wrong classification of unvoiced segments as voiced were found. However 2% of voiced segments are classified as unvoiced and nearly 8% of the segments are rejected as no decision for To was obtained.

REFERENCES

1.Boyanov B. , Ivanov T. ,"Analysis the speech of patients with laryngeal diseases", Report 15/90, Bulgarian Academy of Sciences (In bulgarian).

2.Rabiner L., Shaffer R.,"Digital Processing of speech signals," Prentice Hall, NY, 1978.

3.Hess W.,"Pitch determination of speech signals", Springer Verlag. N.Y. 1983.

4.Lahat M., Niederjohn R.,Krubsack D.,"A spectral autocorrelation method for measurement of the fundamental frequency of noise-corrupted speech",IEEE Tr. Acoust. Speech, Signal Proc., ASSP-35, pp. 741-750,1987.

5.Laver J., Hiller S., Hanson R.,"Comparative performance of pitch detection algorithms of dysphonic voices," IEEE Proc. of ICASSP, pp.192-195,1982.

6.B. Boyanov, G. Chollet Pathological Voice Analysis using Cepstra, Bispectra and Group Delay Functions, Proc. Int. Conference on Spoken Language Processing, Banff, Canada, October, 1992. .