THE METHOD OF PITCH FREQUENCY DETECTION ON THE BASE OF TUNING TO ITS HARMONICS*)

V.Sercov, A. Petrovsky
Computer Science Department,
University of Technology Bialystok, POLAND
Computer Engineering Department,
Belarusian State University of Informatics and Radioelectronics
6, P.Brovky Str., 220027, Minsk, BELARUS
Tel: +375 172 312910; fax: ++375 172 312910
e-mail: palex@org.by

ABSTRACT

The major part of speech compression algorithms from the first simple formant vocoders to widely used at presence LP- and MBE-vocoders in spite of different principals of coding use by the analysis-synthesis procedures information about the frequency of pitch [1-3]. There is a number of methods for its detection [4-6]: on the base of auto-correlation processing, cepstrum analysis, after linear prediction remaining signal. The methods provide good results when analysing parts of speech with pronounced vocalisation by insignificant noise level. When there is a strong noise component in the input signal the number of errors and accuracy of pitch frequency detection is greatly increased [7]. To eliminate the errors they use generally either method of linear smoothing of pitch contour that additionally decrease the accuracy.

The paper presents the method and algorithm of pitch frequency detection which doesn't provide gross errors and has the high accuracy.

1 PROBLEM STATEMENT

The number of papers about pitch frequency detection is already published to the time. For our analysis let's make use of last publication describing modified cepstrum method [8]. In a noise environment the accuracy increasing is obtained due to additional operations of spectrum smoothing, clipping and removing of high frequency components.

Let's note some defects of the method. The spectrum smoothing does not consider the pitch frequency therefore at its low values in the smoothing domain inevitably more pitch harmonics enter than at low pitch frequency. It leads to "tremor" of the smoothed spectrum at high pitch frequency or to incomplete extraction of formant speech structure at low pitch frequency.

To accomplish the clipping operation a threshold value (level of clipping) depending only on minimum and maximum values of the spectrum logarithm and clipping coefficient is used. As a clipping coefficient is fixed then depending on spectrum characteristic either part of productive pitch harmonics can be left out of account or some noise components can be ignored for the farther pitch frequency detection.

The procedure of high frequency components removing is based on assumption that in this area of spectrum the major part of noise component power is concentrated. In the nature conditions such assumption does not correspond to reality. So, fig.1 shows spectrum logarithm of speech fragment with noisy average band. Obviously, in the such domains the modified cepstral method performs with larger errors. Relatively high length of input time window (about 50 ms) in the domains of fast pitch changes must assist to excessive pitch contour smoothing that reduces the method accuracy. At last, it should be noticed the high computing complicity of the algorithm: only for the cepstrum computation it is requested to implement 1024-point FFT.

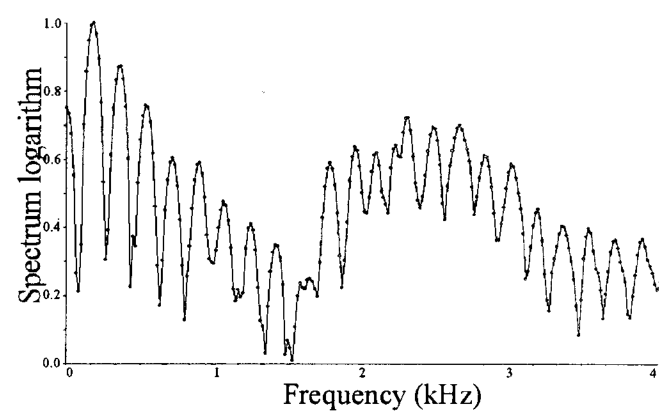

Fig.1. Normalized spectrum logarithm

The listed above defects make to define the directions of farther method perfection. The problem is to develop a method of pitch frequency detection that would use shorter interval of the analysis, pitch correlated spectrum smoothing, adaptive clipping threshold computation, more effective procedure of high noise bands separation, less computing complicity.

*) This work was supported in part by the Computer Science Department, University of Technology Bialystok (Poland) and the Computer Engineering Department, BSUIR (Belarus) under the grants W/II/6/96 and #98-3090 respectively.

2 THE METHOD DESCRIPTION

The method of pitch detection which is presented in the given paper is based on the approach of harmonic peak detection [9]. This approach may be considered as a maximum likelihood technique in the presence of additive noise [10], making it very robust.

Fig.2 shows the diagram of proposed method of pitch frequency detection.

The method and algorithm of pitch frequency description as follows:

#1. Short time Fourier transform is computed:

1.1) the input speech signal is weighted 256-point time window (sampling rate is 8kHz, Kaiser window with parameter β=7 and overlapping coefficient 0.75 are used);

1.2) the signal segments are added by zero up to 512 points for the purpose of the spectral analysis frequency resolution increase;

1.3) FFT is performed. The frequency resolution for each frequency components is 15.655 Hz that represents the compromise between the requirements of the good resolution on time and frequency domain.

#2. Its spectrum power log is calculated.

(For an elimination of influence of various loudness the log of a spectrum is normalized so that the maximum value in each segment of the analysis was equal 1, as it is shown in the fig. 1.)

#3. The preliminary estimation of pith frequency is implemented.

(Research have shown that with confidence to tell about vocal segment is possible availability even four first fundamental frequency harmonics. This fact was used for preliminary estimation of the pitch frequency:

$$\hat{f}_0 = \frac{1}{4}\sum_{i=1}^{4} f_i / i,$$

where f_i is mean frequency of the pitch frequency i-th harmonic peak.)

#4. Operation of spectrum smoothing is carried out.

(With the purpose of an exclusion of influence the formant structure of speech, as well as in [8] the equalization (smoothing) of a spectrum is used. To supply an identical degree of smoothing at various frequencies of pitch frequency a length of the spectrum average window should depend on fundamental frequency. In the given method it is selected equal 1.5 $\hat{f}_0$ providing in practice good results.) In the fig. 3 the spectrum smoothing result is shown.

#5. The clipping operation is carried out:

5.1) the threshold of clipping is calculated for each segment of the analysis so that the quantity of maximum above it was equal to computational number of pitch frequency harmonics;

5.2) at clipping the spectrum value smaller some threshold are installed equal to zero. This operation allows to increase quality of a method in conditions of an additive noise.

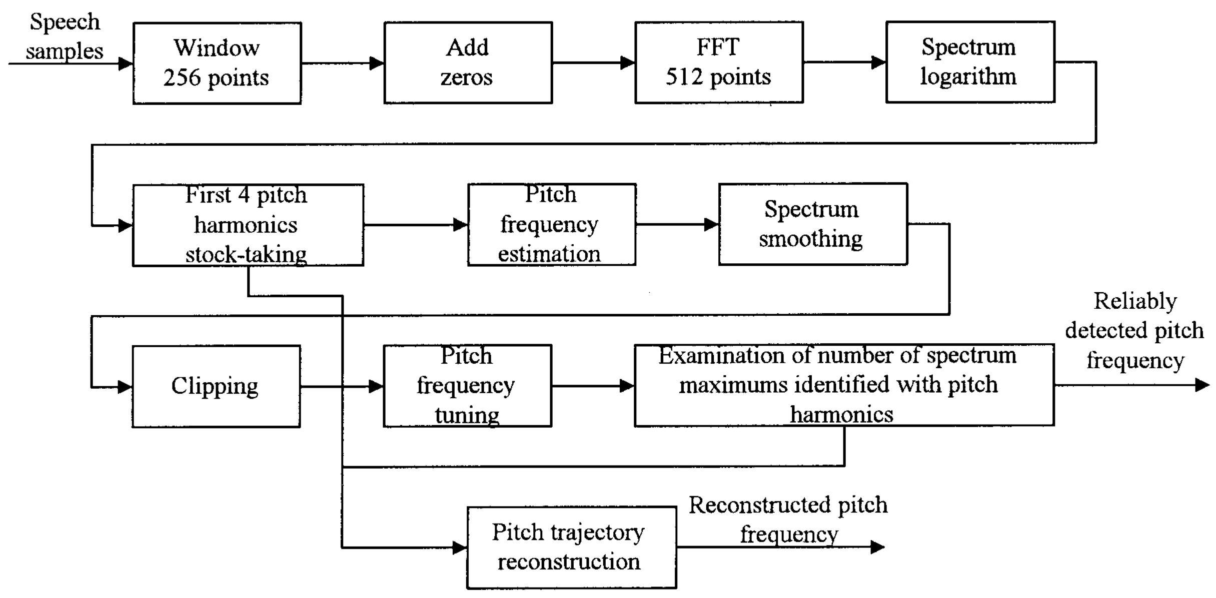

Fig. 2. The diagram of the method of pitch frequency detection on the base of tuning to its harmonics

The example of the spectrum after clipping is adduced in the fig. 4.

#6. The pitch frequency tuning.

#7. The pitch frequency contour reconstruction.

In contrast to majority of methods, for an example [11], the value of a threshold is not fixed. Such way allows effectively to adapt for segments with a various degree vocalization.

3 THE PITCH FREQUENCY TUNING

The pitch frequency tuning is concluded in the optimizing task resolving: it is founded the frequency when the number of clipping spectrum maximums situated in some vicinity of the calculation harmonics of the pitch frequencies is the biggest and weighted sum of departures of the maximum centers from the given calculation frequencies is minimum.

As the weighting coefficients the maximum amplitudes are used. By pitch frequency detection the such approach allows not to take account of those spectrum maximums that are not on «their places» and, hence, are not true pitch harmonics.

For pitch frequency tuning is used the following algorithm:

```
s := 1000;    N_peak := 0;
ff := f̂_0 -Δ;
   repeat
   ss := 0;  n := 0;
   for i := 1 to N_max do
      begin
      δ := | f_i - ff · {f_i / ff} |;
      if δ < Δ then
         begin
         n := n + 1;  ss := ss + δ · h_i;
         end; {if}
      end; {for}
   if (n > N_peak) or ((n = N_peak) and (ss < s)) then
      begin
      f_0 := ff; N_peak := n;  s := ss;
      end; {if}
   ff := ff+0.01
   until ff > f̂_0 + Δ;
```

In adduced algorithm the following labels are adopted: {} is a rounding off operation; N_{peak} is number of the spectrum peaks which identified with pitch harmonics; N_{max} is number of the maximums of the clipped spectrum; f_i is the central frequencies of these maximums; h_i is interpolated values of the spectrum amplitude on the central frequencies. The constant Δ characterizes a maximum difference between the central frequency of the spectrum peak and the computational frequency of the pitch harmonic, at which this peak still is considered appropriate to the harmonic. Experimentally was obtained that the best value Δ is 19 Hz. The step of the pitch frequency tuning is selected that a computational error of the maximum harmonic of the given frequency did not exceed the spectral resolution.

The computed centers of pitch harmonics after the frequency tuning are shown on the fig.4 with vertical dashes.

It is possible to judge about a degree of pitch frequency calculation accuracy, examining the number of spectrum maximums identified with pitch harmonics. For the stable work this number must be in excess of some percentage of calculated maximum number of harmonics for the given pitch frequency. Let's call the value of pitch frequency obtained in such a way the reliable one.

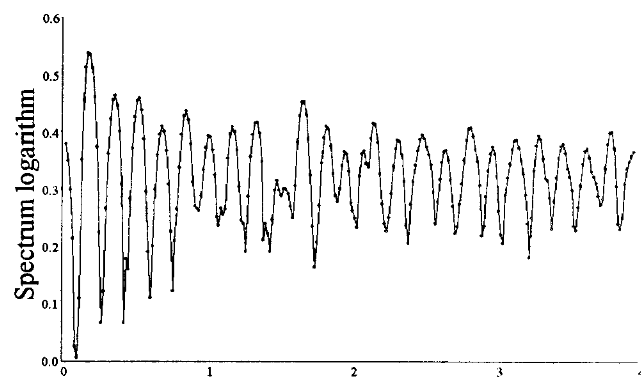

Fig.3. The spectrum logarithm after smoothing operation

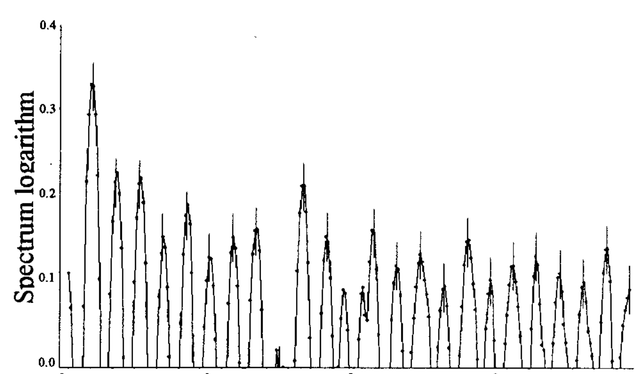

Fig.4. The spectrum after clipping operation

4 THE PITCH FREQUENCY CONTOUR RECONSTRUCTION

If by the analysis it doesn't manage to detect the first 4 pitch harmonics or the number of identified spectrum maximums is less than threshold an attempt to reconstruct the pitch frequency contour will be implemented. Let assume that a pitch frequency was detected for *i*-th time window. Then it can be chosen as a preliminary estimation for $i+1$-th time window. After spectrum smoothing, clipping operation, pitch frequency tuning and examining of identified number of the spectrum maximums are accomplished it is possible to detect the pitch frequency.

The similar procedure can be used for $i+2$-th time window and so on. The pitch contour reconstruction is carried out until not less than 3 of the first 4 pitch harmonics are identified with the spectrum maximums. By examining the number of identified spectrum maximums the permitting threshold must be chosen lesser than this one for the reliable pitch frequency detection. The pitch frequency accuracy in a case of the contour reconstruction will be lower therefore it could be usefully to make smoothing of reconstructed points of the contour. If a delay in the pitch frequency detection is allowed then the procedure of contour reconstruction should be carried out with time reducing, taking into account the pitch frequency of i-th time window for $i-1$-th time window and so on within the limits of permitted delay.

The fig.5 shows the pitch contour for two second speech fragment: "one , two, three, four", but the given fragment is told in Russian by Chinese.

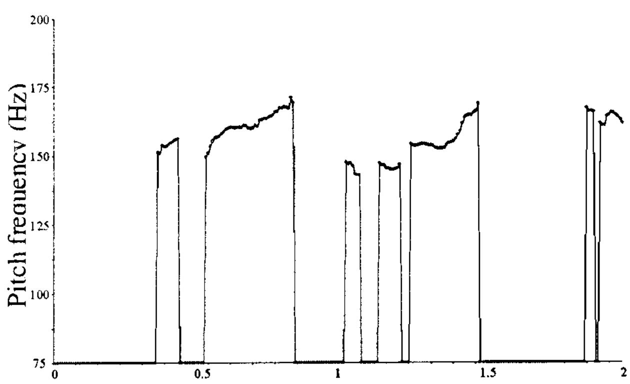

Fig.5. Pitch frequency contour

CONCLUDING REMARKS

The problem of pitch detection has been investigated. The robust method of pitch frequency estimation on the base of tuning to its harmonics was proposed which use shorter interval of the analysis, pitch correlated spectrum smoothing, adaptive clipping threshold computation, more effective procedure of high noise bands separation, less computing complicity.

The contour reconstruction technique allows to judge about the pitch frequency at the moments of attenuation and establishment of tone where the majority of tone-noise detectors do not locate yet the tone component.

The described above method was realised in Delphi 2.0 language. By testing on PC with Pentium-120 processor the algorithm worked in the real time.

The offered method of pitch frequency detection is well intend to real-time implementation in MBE-vocoders [12] which based on the single DSP TMS320C50 (80 MHz).

REFERENCES

[1]. Jayant N. Signal compression: technology targets and research directions. / IEEE Journal on Selected areas in communications, vol. 10, No. 5, June 1992, - pp.796-818.

[2]. Gersho A. Advances in speech and audio compression. / Proc. of the IEEE, vol. 82, No. 6, June 1994, - pp.900-918.

[3]. Spanias A.S. Speech coding: a tutorial review. / Proc. of the IEEE, vol. 82, No. 10, Oct. 1994, - pp.1541-1582.

[4]. Rabiner L.R., Schafer R.W. Digital processing of speech signals.- Prentice-Hall, N.J., 1978, - 512p.

[5]. Steele R. Mobile radio communications. - Pentech Press, London, 1992, - pp.186-346.

[6]. Advances in speech signal processing / Edited by S.Furui, M.M.Soudhi. - Marcel Dekker Inc., N.Y., 1992, - 871p.

[7]. Oh K.A and Un C.K. A performance comparison of pitch extraction algorithms for noisy speech. / Proc. Inter. Conf. Acoust. Speech and Signal Process., 1984, pp.18B4.1.–18B4.4.

[8]. Shimamura T., Kunieda N. and Suzuki J. A robust cepstrum pitch detector. / Proc. the 1st European conf. "Signal analysis and prediction", Prague, Czech, 1997, - pp.285-288.

[9]. A.M.Kondoz. Digital speech: coding for low bit rate communication systems. - John Wiley & Sons, New York, 1996, pp.65-77.

[10]. E.Chilton. Factors affecting the quality of linear predictive coding of speech at low bitrates. PhD thesis, University of Surrey, Guildford, Surrey, UK, October 1990.

[11]. J.J.Dubnowski, R.W.Schafer, and L.R.Rabiner. Real-time digital hardware pitch detector. /IEEE Trans. on Acoustics, Speech, and Signal Processing, vol. 24, No. 2, February 1976, - pp.2-8.

[12]. Griffin D.W., Lim. J.S. Multiband excitation vocoder. / IEEE Trans. on Acoustics, Speech, and Signal Processing, vol. 36, No. 8, August 1988, - pp.1223-1235.