Analysis of speech signal restored from the pitch harmonics

Snejana G. Pleshkova-Bekyarska[1] and Michail B. Momcshedjikov[2]

Abstract— **The most known speech coding method using decoding as synthese of the speech signal with the coded speech characteristics. One of these characteristics is the pith frequency of the voice sounds. It is very important to transmit the values of the pitch frequency and its harmonics with the minimum numbers of bits. The goal of this article is to make an analysis of the restored speech signal with the pitch harmonics and to make an estimation of the distortion in the restored speech signal.**

Keywords— **Speech analysis, speech coding, speech synthesis, speech coding distortions**

I. Introduction

The pitch frequency determination can be made in the time or frequency domain [1]. Each of these methods give some precision for the current values of the pitch frequency. In this article the goal is to analyze the restoration of speech signal from the pitch harmonics and not the goal to study the pitch frequency determination methods.

II. Speech representation with pitch harmonics

It is chosen in this article to work with the pitch frequency determination method in frequency domain [2]:

$$E\left(\omega_0\right)=\frac{1}{2\pi}\int_{-\pi}^{+\pi}\left|S_\omega\left(\omega\right)-S_\omega\left(\overset{\wedge}{\omega},\omega_0\right)\right|^2 d\omega, \qquad (1)$$

where:

ω_0 is the pitch frequency;

$S_\omega(\omega)$ and $S_\omega(\omega,\omega_0)$ - the spectrum of the original and restored speech signals, respectively, with the use of window function ω.

The determination of the pitch frequency ω_0 is used in the speech coding methods with restoration of the speech signal from the harmonics of the pitch frequency [3]:

$$A_l\left(\omega\right)=\frac{\sum\limits_{n=a_l}^{b_l} S_\omega\left(n\right).W_R\left(n\right)}{\sum\limits_{n=a_l}^{b_l}\left|W_R\left(n\right)\right|^2}, \qquad (2)$$

where:

$\grave{A}_l(\omega)$ is l^{th} harmonics of the pitch frequency ω_0;

$S_\omega(n)$ – the spectrum of a frame of speech signal;

$W_R(n)$ – the window function for separating the l^{th} harmonic of the pitch frequency from of the whole spectrum $S_\omega(n)$;

$n=a_l\div b_l$ – frequency interval around the l^{th} harmonic of the pitch frequency, chosen very narrow to determine with a good precision the average amplitude of the l^{th} harmonic of the pitch frequency.

Using the values $\grave{A}_l(\omega_0)$ of harmonics of pitch frequency from (2) in the synthesis of restored speech signals in the decoders must be considered with the characteristics of the speech signals and the chosen method for coding characteristics determination in the domain of a sequence of frames with 20÷30mS duration [1]. It is not real to decide that the pitch frequency is constant even if there is a case of a speech signal for a separation voiced letter, with the same values for the frequency and phase in the two adjacent frames.

III. Speech frame transition cases

It is necessary to consider in this analysis of the distortions the following cases of transition between two frames j-1 and j in accordance with values of logical signal v/uv for voiced (v/uv=1) and unvoiced (v/uv=0) speech signals:
the transition between two frames, belonging to unvoiced speech:

$$v/uv\left(j-1\right)=0,\quad v/uv\left(j\right)=0; \qquad (3)$$

- the transition between two frames, belonging to voiced speech:

$$v/uv\left(j-1\right)=1,\quad v/uv\left(j\right)=1; \qquad (4)$$

the transition between two frames, first frame belonging to unvoiced speech and the second – to voiced speech:

$$v/uv\left(j-1\right)=0,\quad v/uv\left(j\right)=1; \qquad (5)$$

or vice versa:

$$v/uv\left(j-1\right)=1,\quad v/uv\left(j\right)=0. \qquad (6)$$

For the case (4) the transition between two frames, belonging to voiced speech, can be define two conditions:
the speech signal for one voiced letters:

$$\omega_0\left(j-1\right)-\omega_0\left(j\right)\le\Delta\omega_0; \qquad (7)$$

the speech signal for two difference voiced letters:

$$\omega_0\left(j-1\right)-\omega_0\left(j\right)\ge\Delta\omega_0, \qquad (8)$$

[1] Snejana G. Pleshkova-Bekyarska is with the Faculty of Communications and Communications Technologies, Technical University, Kliment Ohridski 8, 1000 Sofia, Bulgaria, E-mail: snegpl@vmei.acad.bg

[2] Michael B. Momschedjikov is with the Faculty of Communications and Communications Technologies, Technical University, Kliment Ohridski 8, 1000 Sofia, Bulgaria, E-mail: snegpl@vmei.acad.bg

where:

$\Delta\omega_0$ is a predefined value of the difference between ω_0(j-1) and ω_0(j) in frame (j-1) and j, respectively.

IV. Restored speech distortion analysis

In the condition (7) is satisfied, then the synthesis of the speech signal with pitch harmonics must be done with the concordance of the amplitudes a$_l$(n) and phases θ_l(n) of the pitch harmonics in the two adjacent frames. It is used very often the methods of interpolation between amplitudes M$_l$(j-1) and M$_l$(j) and phases φ_l(j-1) and φ_l(j) for two adjacent frames:

$$a_l\left(n\right) = f_a\left(M_l\left(j-1\right), M_l\left(j\right)\right) \tag{9}$$

$$\theta_l\left(n\right) = f_\theta\left(\varphi_l\left(j-1\right), \varphi_l\left(j\right)\right), \tag{10}$$

where:

- f$_a$ and f$_\theta$ are respectively the functions chosen for interpolation of amplitudes and phases of each lth harmonic in two adjacent frames ;

- M$_l$(j-1), M$_l$(j), φ_l(j-1) and φ_l(j) – the amplitudes and phases of lth pitch harmonic in the transition between two adjacent frames (j-1) and j.

The amplitudes M$_l$(j-1) and M$_l$(j) and phases φ_l(j-1) and φ_l(j) can be determined from $\grave{A}_l(\omega)$ (2) – the complex value of lth pitch harmonic in frame (j-1) and j, respectively.

The most known speech coding methods, used the linear interpolation for the functions f$_a$ and f$_\theta$ in (9) and (10):

$$a_l\left(n\right) = M_l\left(j-1\right) + \left[M_l\left(j\right) - M_l\left(j-1\right)\right]\frac{n}{N} \tag{11}$$

$$\begin{gathered}\theta_l\left(n\right) = \varphi_l\left(j-1\right) + \left[l\omega_0\left(j-1\right) + \Delta\omega_l\left(j\right)\right]n + \\ + \left[\omega_0\left(j\right) + \omega_0\left(j-1\right)\right]\tfrac{l.n^2}{N},\end{gathered} \tag{12}$$

where:

$\Delta\omega_l$(j) is the variation of the frequency:

$$\Delta\omega_l\left(j\right) = \frac{1}{N}\left(\Delta\Phi_l\left(j\right) - 2\pi\left[\frac{\Delta\Phi_l\left(j\right)+\pi}{2\pi}\right]\right) \tag{13}$$

$\Delta\Phi_l$(j) – the variation of the phase angle:

$$\Delta\Phi_l\left(j\right) = \Phi_l\left(j\right) - \Phi_l\left(j-1\right) - \left[\omega_0\left(j-1\right) + \omega_0\left(j\right)\right]\frac{l.N}{2}. \tag{14}$$

The expressions (11), (12), (13) and (14) guarantee the continuos variations of the pitch harmonics in the order of two adjacent frame (j-1) and j.

For the case (7) the restored signal for the voiced speech S$_v$(n) is:

$$S_v\left(n\right) = \sum_{l=1}^{L} a_l\left(n\right)\cos\left[\theta_l\left(n\right)\right], \tag{15}$$

where:

L is the chosen maximum number of pitch harmonics for the synthesis of speech signal.

For the case (8) it is necessary to use the overlap-add method for concordance of the amplitudes and phases in two adjacent frames (j-1) and j. This mean that in the place of transition the restored speech signal is the sum of two signals S$_{v1}$(n) and S$_{v2}$(n), respectively for the pitch harmonics of ω_0(j-1) the voiced speech in the (j-1) frame and for the pitch harmonics of ω_0(j) for the voiced speech in the j frame:

$$S_v\left(n\right) = S_{v1}\left(n\right) + S_{v2}\left(n\right), \tag{16}$$

where:

$$S_{v1}\left(n\right) = \sum_{l=1}^{L}\omega_s\left(n\right)M_l\left(j-1\right)\cos\left[\omega_0\left(j-1\right)nl + \Phi_l\left(j-1\right)\right] \tag{17}$$

$$S_{v2}\left(n\right) = \sum_{l=1}^{L}\omega_s\left(n-N\right)M_l\left(j\right)\cos\left[\omega_0\left(j\right)\left(n-N\right) + \Phi_l\left(j\right)\right], \tag{18}$$

where:

$\omega_s(\ldots)$ – window function for the synthesized speech.

It is seen from the expressions (15) and (16), that in the place of transition between two adjacent frames of speech signal, it is possible to have the distortions from the variations of the phase of synthesized speech for the case of the single variations of pitch frequency ω_0(j-1)≈ ω_0(j) in two adjacent frames or from the variations of the amplitude and phase of the synthesized speech signal for the strong variations of the pitch ω_0(j-1)≠ ω_0(j) in two adjacent frames. The quantity estimation of these of these distortions must be done from the expression (11) for the amplitude distortions (13) for the frequency distortions and (12,14) for the phase distortions.

V. Simulation results

Some of the simulation results for different methods for speech signals synthesis in the place of transition between two adjacent frames are shown for visual impression from Fig.1 to Fig.6.

On the all figures are shown respectively the time and spectrum graphics in the following sequence: original and restored signal. The time interval in the time graphics is chosen such, that to display more then one frame for analysis or synthesis of speech signal. With an arrow mark it is shown the place where there, are the distortion in the restored speech signal and the respectively place in the spectrum of the restored speech signal, where there are the differences with the spectrum of the original speech signal.

VI. Conclusion

The time and spectrum graphics in Fig.1, 2, 3, 4, 5 and 6 are different each to an other with the choice of speech synthesis method. The comparison of the graphics for these three methods shown, that the choice of method of speech synthesis is important to the value of amplitude distortion of the speech signals and their corresponding difference of spectrum in respect of original speech signals.

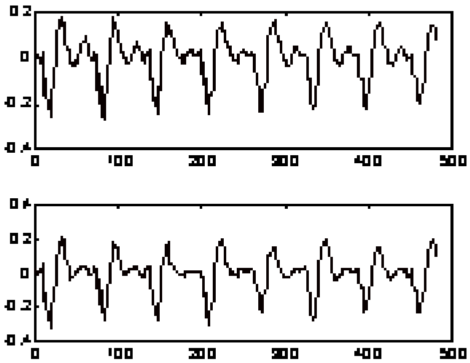

Fig. 1. The time graphics in the following sequence: original and restored signal of method 1.

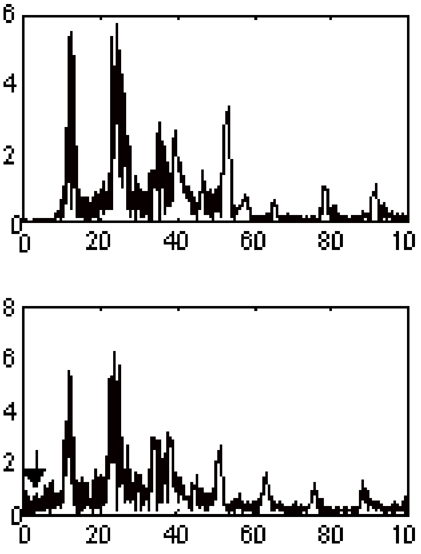

Fig. 2. The spectrum graphics in the following sequence: original and restored signal of method 1.

References

[1] Kondoz A.M, *Digital Speech*, John Wiley and Sons LTD, New Jork, 1994.

[2] Griffin D.W. and J.S.Lim, "Multi-Band Exitation Vocoder", *IEEE Transactions an ASSP*, 36(8) August 1988, pp.664-678.

[3] Spanias A.S, "Speech coding:A Tutorial Review", *Proceedings of the IEEE*, Volume 82, No10, October 1994..

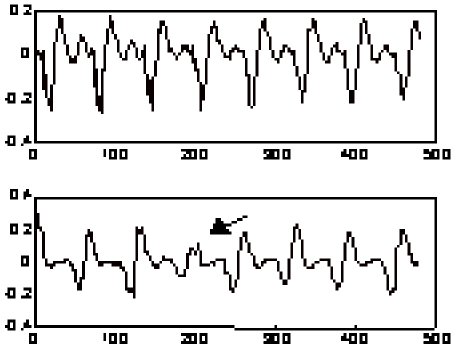

Fig. 3. The time graphics in the following sequence: original and restored signal of method 2.

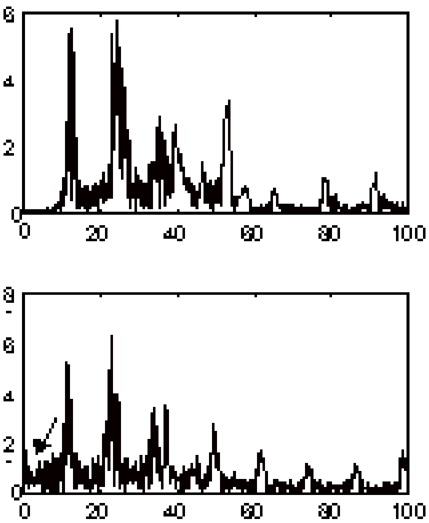

Fig. 4. The spectrum graphics in the following sequence: original and restored signal of method 2.

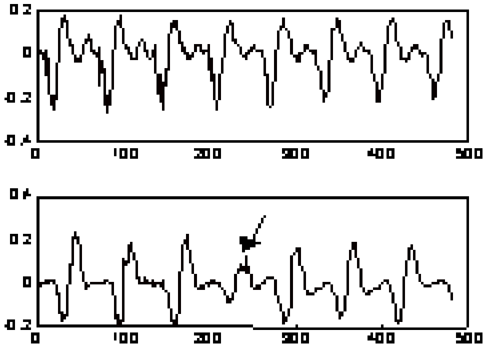

Fig. 5. The time graphics in the following sequence: original and restored signal of method 3.

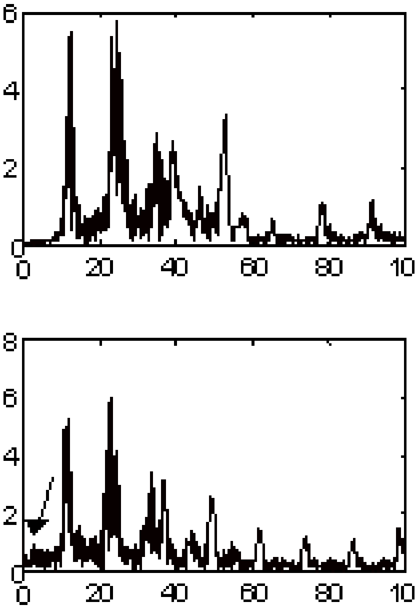

Fig. 6. The spectrum graphics in the following sequence: original and restored signal of method 3.