Turkish Journal of Electrical Engineering and Computer Sciences

Volume 14 | Number 2 Article 2

1-1-2006

Speech Pre-Processing for Pitch and Pitch-Cylce Evolutions Smoothing

HASSAN FARSI

Turk J Elec Engin, VOL.14, NO.2 2006, © TÜBİTAK

Speech Pre-Processing for Pitch and Pitch-Cylce Evolutions Smoothing

Hassan FARSI
Department of Electronic and Electrical Eng., Faculty of Eng.,
University of Birjand, Birjand-IRAN
e-mail: hs_farsi@birjand.ac.ir

Abstract

In low bit rate speech coders, pitch is usually transmitted once per frame and, when needed, the intermediate pitch values are obtained by interpolation between 2 adjacent pitch values. Although pitch usually evolves slowly, sometimes it has irregular variations and the estimated pitch differs from the real one. In addition, some speech coders, e.g., waveform interpolation coders, rely on smooth pitch-cycle evolutions to extract speech model parameters in the analysis stage. However, non-stationary characteristics of speech may lead to inaccurate estimation of the parameters. This affects the synthesised speech quality.

We propose a pre-processor, which modifies the residual speech signal to provide smooth pitch variations and pitch-cycle evolutions, without distorting perceptual speech quality. Thus, the pitch and the voicing level can be more accurately determined.

Key Words: *Pitch, speech coder, bit rate, pitch cycle.*

1. Introduction

In low bit rate speech coders, pitch estimation and voicing-level computation play major roles in speech quality. For instance, waveform interpolation (WI) encoders require the pitch period at every extraction point in order to perform characteristic waveform (CW) extraction. In WI decoding, a pitch value at every sample point is required to construct a phase track, and then to convert a two-dimensional surface to a one-dimensional signal [1]. In mixed excitation linear prediction (MELP), the voicing level of speech is calculated by using the estimated pitch and the normalized autocorrelation value for 5 frequency bands [2].

The basic assumptions for pitch estimation algorithms are:

- Voiced speech samples are correlated at specific time intervals, called pitch periods. Although these samples are usually highly correlated, they sometimes have low correlation and, therefore, the resulting normalised autocorrelation is unreliable for pitch estimation.
- Pitch evolves smoothly during a voiced frame. Although this usually happens, pitch occasionally has irregular variations, which can lead to inaccurate pitch estimation. This affects the CW extraction in WI encoders (see section 2) and phase track construction in WI decoders. Moreover, in MELP, the

voicing level of speech is affected by inaccurate pitch estimation, which degrades synthesised speech quality. The irregular pitch variations usually happen in transition frames, where speech characteristics change from quasi-periodic to random-like signals, or vice versa, and therefore, the long-term correlation of the speech signal is affected.

In order to overcome these problems, a methodology that provides more accurate estimation of the parameters is required. This methodology can be performed either inside a speech coder using alternative algorithms, or outside, as a pre-processor. In the first case, no modification is performed on the speech signal and alternative factors, such as history and future of a parameter, and spectral matching, etc. are used to find a better estimation, whereas creating more regular speech is the objective of the second method. This enables the algorithms used in the speech coder to calculate a more accurate pitch. This technique is called speech pre-processing since the speech signal is modified before passing to the speech coder.

Existing pre-processors have been designed for special coders. For instance, in [3], Kleijn introduces a pre-processor in combination with a block-DFT-based WI coder. This coding structure maintains the advantages of earlier WI coders and adds the asymptotically perfect reconstruction property. Since the alignment procedure employed by a WI encoder causes the relative phase loss of the CW, the WI coder does not produce a perfect reconstruction [3]. The alignment procedure is included as a part of the pre-processing performed outside the WI coder. This is performed by moving a pitch pulse to the centre of a CW so that the modified segment is maximally correlated to the previous cycle, and thus, the employed alignment procedure in earlier WI coders is not required.

The pre-processor employed in [4] performs high-pass and adaptive noise suppression filtering before the estimation of speech parameters. After speech parameter estimation, the residual signal is modified to generate a target residual for the fixed codebook search in a CELP coder. A shifted target residual is generated using the past-modified residual and the delay contour of the current frame. This shifted residual is used as a target for shifting the residual of the current sub-frame. All pitch pulses in the original residual are shifted individually to match the delay contour of the modified target residual.

In this paper, pitch and pitch-cycle evolutions smoothing is applied as a pre-processing method independent of a low bit rate speech coder, which modifies the residual signal so that a more smoothly evolving pitch contour and pitch-cycle waveform are achieved. This paper is organised as follows. In section 2, we present the pitch estimation problems that lead to incorrect intermediate pitch values in WI and also inaccurate voicing level in MELP, when pitch evolves non-linearly. In section 3, the new pre-processing method is introduced. The results are analysed in section 4 and the conclusions are drawn in section 5.

2. Position of the Problem

In order to estimate pitch and intermediate pitch values correctly, speech coders usually assume that the pitch evolves slowly during a frame. However, when the pitch has irregular variations, the estimated pitch values may be incorrect. As an example, in Figure 1a, a voiced residual speech, whose pitch period evolves non-linearly, is shown. We apply the procedure in [1] to obtain the pitch values. In Figure 1b, the estimated pitch values are shown in comparison to the real pitch values. It is observed that the estimated pitch values differ from the real ones. In the first experiment, the effect of the incorrectly estimated pitch values in a WI coder is studied. The CWs are extracted every 2 ms with lengths given by the intermediate pitch values [1]. In Figure 2, two of the extracted CWs are shown. It is observed that when the interpolated pitch is longer than the real pitch value, there is one more pitch pulse during the extracted CW (Figure 2a). On

the other hand, when the interpolated pitch is shorter than the real one, the CW does not contain any pitch pulses (Figure 2b). After the alignment and the normalisation procedures on the extracted CWs, the CW surface is constructed as shown in Figure 2c. Since the lengths of the intermediate CWs are incorrect, some of the pitch pulses are not phase-aligned. Next, the resulting CW surface is decomposed into a slowly evolving waveform (SEW) and a rapidly evolving waveform (REW) by low-pass and high-pass filtering, with a cut-off frequency of 20 Hz. It is expected that the voiced speech is transferred to the SEW surface, but the misalignment of the pitch pulses causes the bandwidth of the evolution spectrum to exceed 20 Hz. As a result, a part of the voiced speech is transferred into the REW surface and the decomposition is performed incorrectly.

In the second experiment, the effect of non-linear pitch variations on voicing level estimation in MELP is studied, where voicing decision is performed for 5 frequency bands. In Figure 1a, it is observed that in spite of having strongly voiced speech for the second and third frames, only 2 bands (the first and third) for the first frame and the first band for the second frame are considered as voiced. However, all frequency bands of the first frame are estimated as voiced.

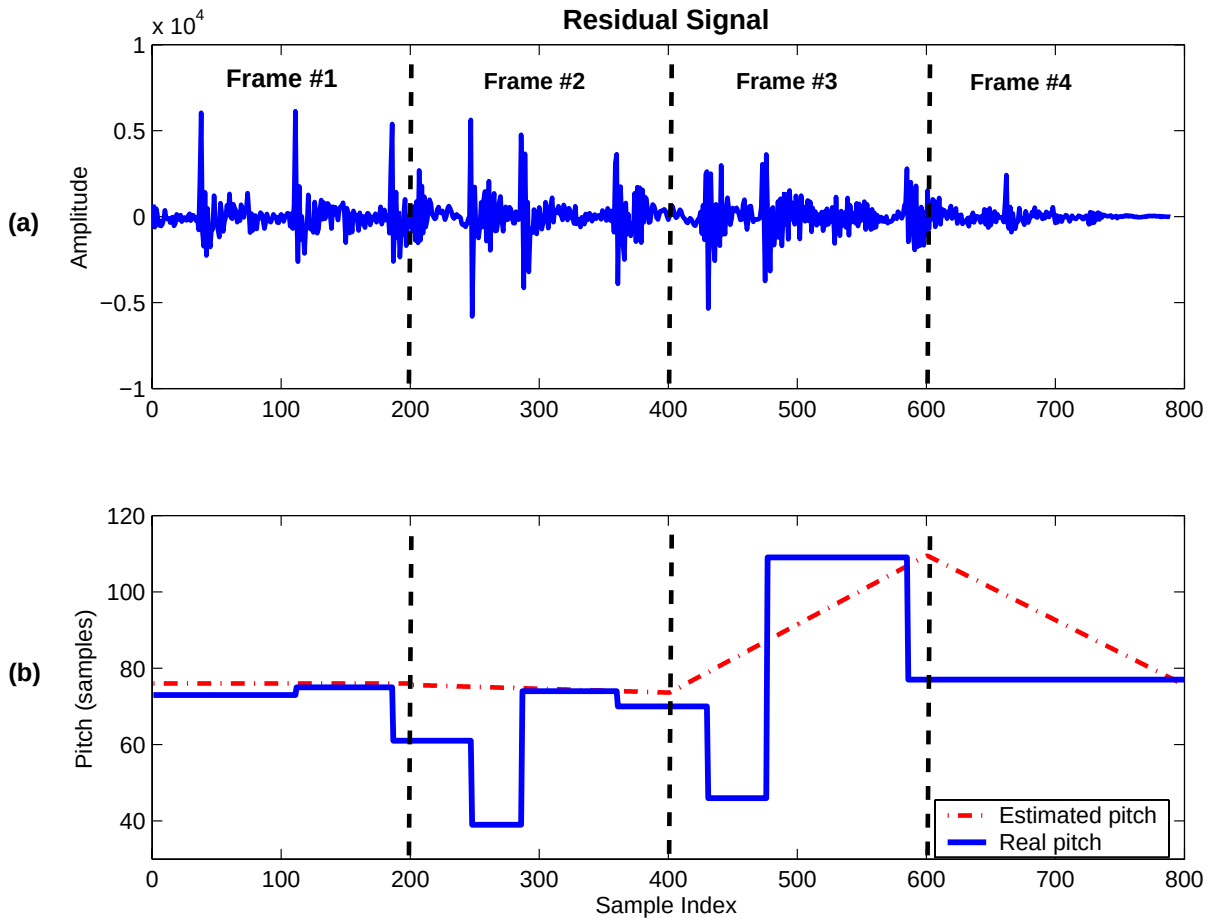

Figure 1. The voiced residual speech signal with non-linear pitch variations (a) and the estimated pitch values (dash-dot) in comparison to the real pitch values (solid) (b).

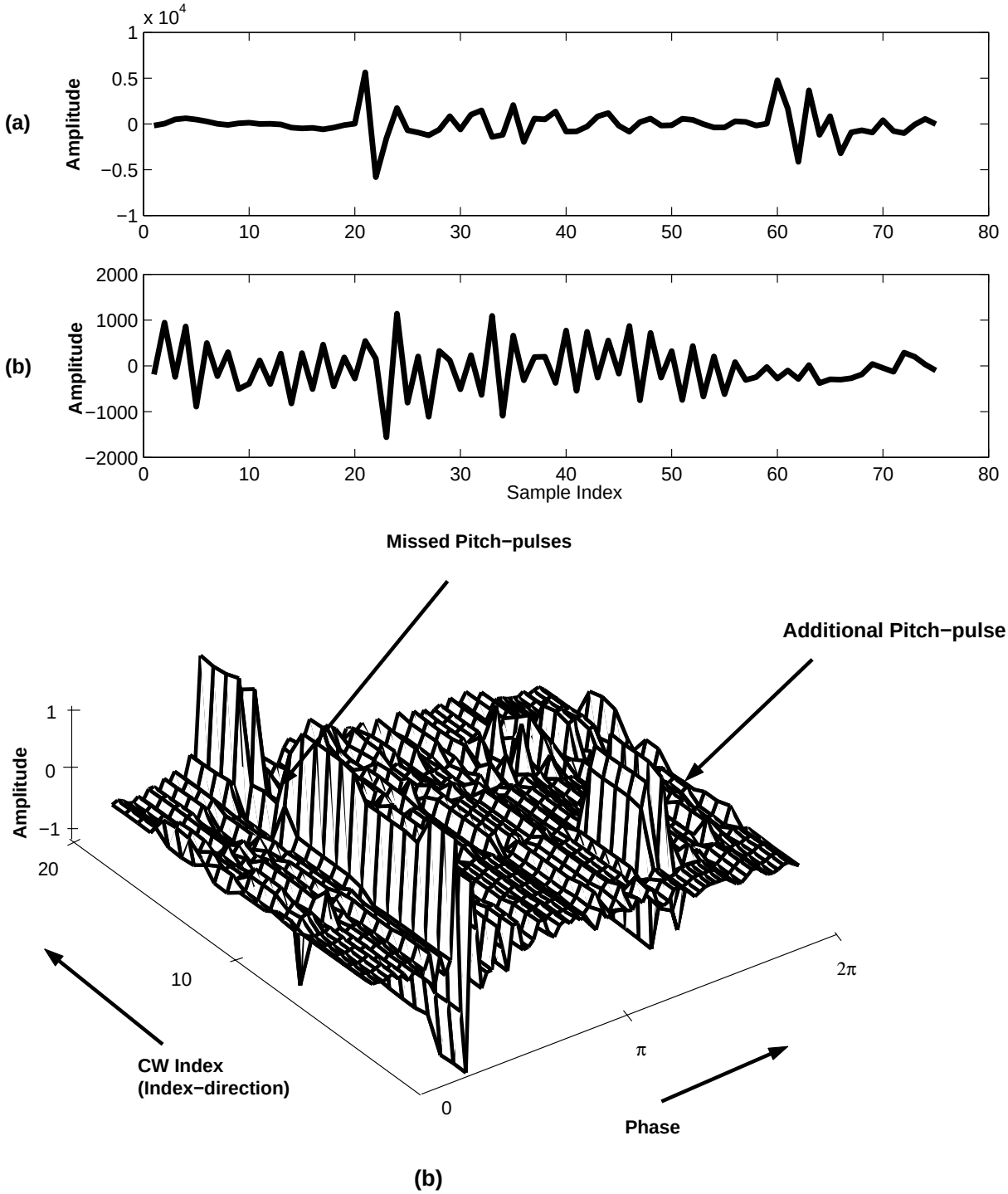

Figure 2. The CW corresponding to an estimated pitch longer than the real pitch (a) and the CW corresponding to an estimated pitch shorter than the real pitch (b). The resulting CW surface in the case of irregular pitch variations (c). The resulting CW surface after proposed pitch modification (d).

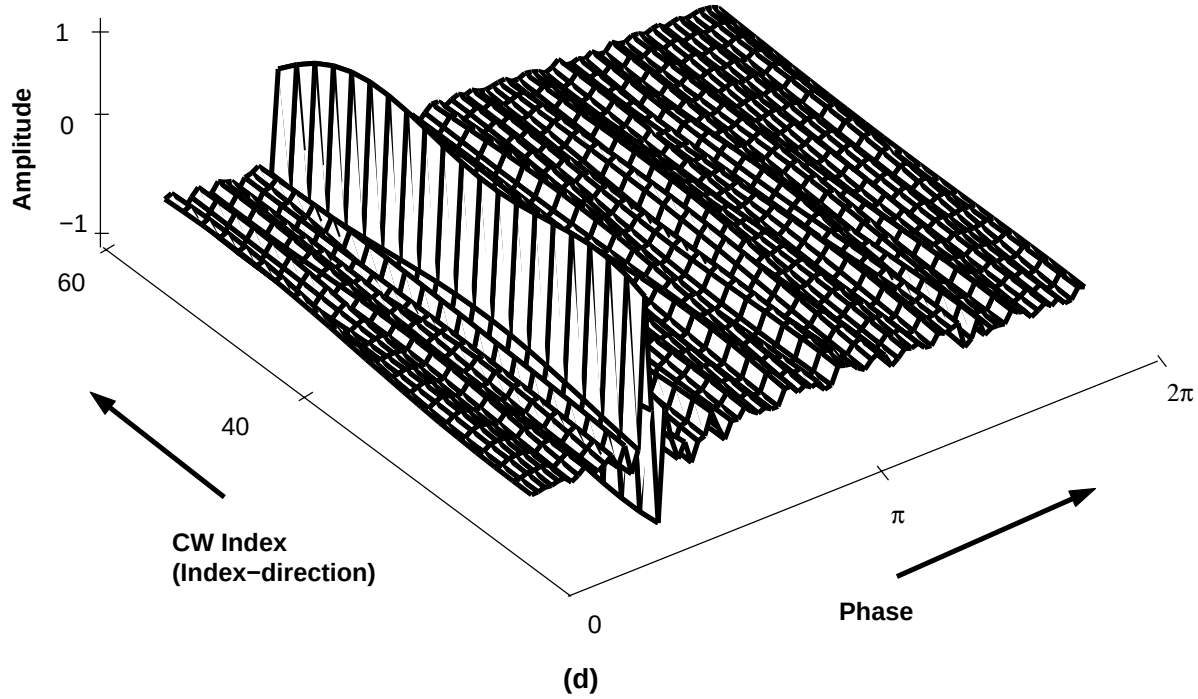

Figure 2. Contunied.

In the next section, a method, which slightly modifies the residual signal, is introduced. This modification ensures that the pitch values evolve more smoothly during a frame and the voicing level of speech is more accurately estimated.

3. Proposed Pre-processing

The proposed pre-processing algorithm modifies the residual signal so that it is more suitable for coding. The inputs to the pre-processor are the linear prediction residual of the speech signal, an associated pitch track, and the voicing decision. The pitch period is estimated once in a frame using conventional autocorrelation-based methods, and the resulting estimate is then linearly interpolated for each pitch cycle. The output is a modified linear prediction residual, which is constructed by concatenation of the modified/unmodified pitch cycles of the residual signal.

If the frame is unvoiced, no modification is made to the pitch value. During voiced sections, the main task of the pre-processor is to smooth the pitch variations and pitch-cycle evolutions, while increasing long-term correlation of the speech signal, and also maintaining the speech perceptually identical to the original. Figure 3 shows a block diagram of the proposed pre-processor, and we now discuss the operation of the pre-processor on a step-by-step basis.

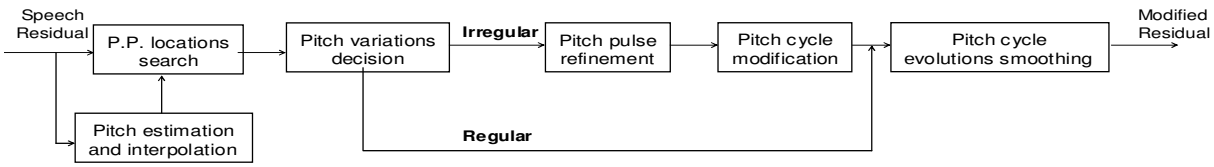

Figure 3. A block diagram of the proposed pre-processor.

3.1. Local pitch estimation

During voiced sections, the locations of the pitch pulses in a frame are searched. The search is based on the concentrated energy of residual samples and on the interpolated pitch period estimate. Since both the valid and invalid pitch pulses are detected using concentrated energy measure, the envelope of windowed LP residual (HEWLPR) with adaptive threshold is used to separate the invalid pitch pulses from the valid ones [5]. If 2 potential pitch pulse locations at n_k and n_{k+1} are found, the distance $d_k = n_{k+1} - n_k$ indicates one local pitch period between those instances. Next, a vector including local pitch values V_{lp} is created. The size of the vector V_{lp} depends on the number of pitch pulse locations. Since V_{lp} contains local pitch values and the difference between 2 successive pitch pulse locations indicates the local pitch value, the size of the vector V_{lp} is N-1, where N is the number of pitch pulse locations found in a frame. If the standard deviation of this vector is less than a predetermined threshold, the pitch pulse locations of the current frame are not modified and step 3.4 is applied. Otherwise, the first stage of the modification is performed.

3.2. First stage: Pitch-pulse refinement

Our experiments show that the proper pitch pulse locations are sometimes estimated incorrectly. Thus, in the first stage of the proposed pre-processor, pitch-pulse validity is tested. A pitch-cycle centred on the pitch-pulse and with a length of twice the minimum pitch period is used for every pitch-pulse. Then, the correlation between the pitch-cycles is computed and a vector V_{Corr} is created. The standard deviation of the vector V_{Corr} is calculated and if it is less than a threshold, the current pitch-pulses are validated and the second stage of the modification is performed; otherwise, we find which pitch-pulse requires refinement based on the minimum correlation. The assumption that validates the refinement process is that the energy of the samples around the pitch-pulse (excluding the pitch-pulse) changes as a fraction of the samples' energy around the previous pitch-pulse. We thus consider a moving window of length 5 to compute the samples energies and define the fraction factor (β) as follows [6]:

$$\beta = \sqrt{\frac{E_{ii}}{\sqrt{E_{mi}.E_{ni}}}} \tag{1}$$

Where E_{ii} , E_{mi}, and E_{ni} are the respective means of the energy of the samples around the current, the previous, and the next pitch-pulses. Then, the samples around the current pitch-pulse (including the current pitch-pulse) are normalised by β. As an example, 2 invalid pitch-pulses in the speech signal shown in Figure 4a are refined using the described algorithm. The synthesised speech is shown in Figure 4b.

In order to study the effect of the refinement procedure on pitch estimation, the original and the modified speech signals are filtered using a low-pass filter and then the normalised autocorrelation is computed for pitch lags between 40 and 160 samples. The results depicted in Figure 5 show that the maximum of the normalised autocorrelation (which occurs at a pitch lag of 78 samples) of the modified speech is around 0.1 higher than the original one. In addition, the maximum that occurred at the sub-multiple of the pitch (at a pitch lag of 117 samples) is more attenuated. This leads to a more reliable pitch estimate.

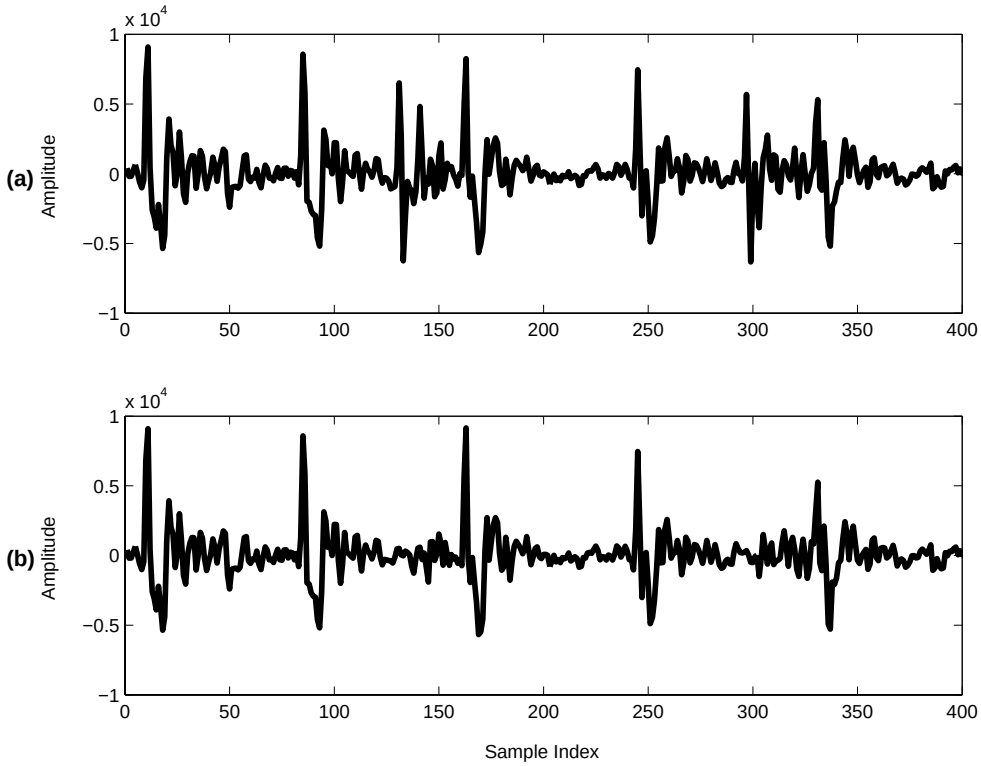

Figure 4. The speech signal including 2 high-energy invalid pitch pulses (a). The modified speech signal (b).

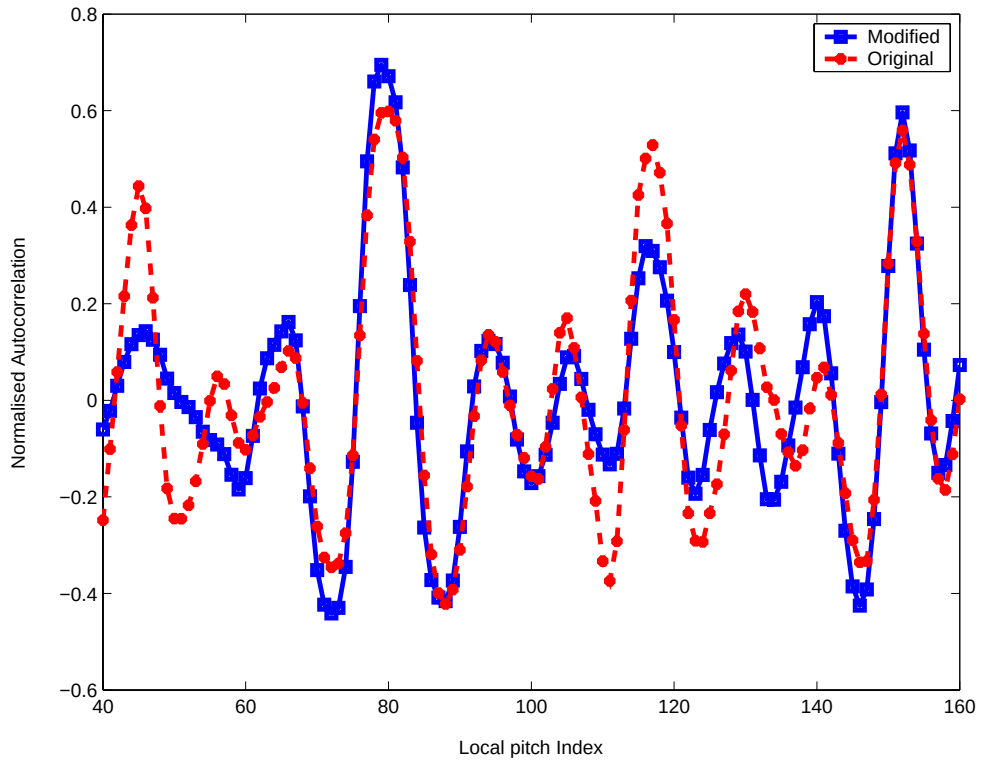

Figure 5. The normalised autocorrelation function computed for both original (Stars) and modified (Squares) speech.

After the first stage of the modification, the vector V_{lp} is updated and the standard deviation of the new vector is calculated. If it is less than the predetermined threshold, the current frame is copied to the output. Otherwise, we find which local pitch has to be modified based on the minimum correlation, and the second stage is performed.

3.3. Second stage: Pitch-cycle modification

A local pitch cycle is defined as the samples located between 2 pulses. Our experiments show that when the pitch has around 5% deviation, the pitch and the voicing level are estimated correctly; thus, we allow 5% deviation of the pitch.

If $C_k(n)$ with length P_k is the local pitch cycle that has to be modified, and $C_{k-1}(n)$ with length P_{k-1} is the previous local pitch cycle, we consider the modified length $\widetilde{P}_k$ as follows:

$$\tilde{P}_k = \begin{cases} 1.05P_{k-1} & P_k > P_{k-1} \\ 0.95P_{k-1} & P_k < P_{k-1} \end{cases} \tag{2}$$

We propose the pitch cycle modification as follows:

If $P_k < P_{k-1}$, a segment r_m from the regular pitch cycle, is inserted into the irregular pitch cycle at the same position, the modified pitch cycle is calculated by Eq. (3).

$$\tilde{C} \begin{cases} C_k(n) & 0 \le n < N_p \\ \delta r_m(n - N_p) & N_p \le n < L + N_p \\ C_k(n - L) & L + N_p \le n < \tilde{P}_k \end{cases} \tag{3}$$

N_p is the start point of the segment r_m, which is the connection point, and since r_m is selected from the previous pitch cycle, N_p is in the interval of $[0, P_{k-1} - L]$. The length L is given by Eq. (4).

$$L = \left| \widetilde{P}_k - P_k \right| \tag{4}$$

δ is a factor that changes the energy of the segment r_m based on the irregular pitch cycle energy and is given by Eq. (5).

$$\delta = \sqrt{\frac{E_k}{E_{k-1}}} \tag{5}$$

Where E_k and E_{k-1} are the energy of the pitch cycles C_k and C_{k-1}, respectively.

If $P_k > P_{k-1}$, the segment r_m from the irregular pitch cycle is discarded, the modified pitch cycle is calculated by Eq. (6).

$$\tilde{C}_k(n) = \begin{cases} C_k(n) & 0 \le n < N_p \\ C_k(n + L) & N_p \le n < \tilde{P}_k \end{cases} \tag{6}$$

In this case, due to selecting the segment r_m from the current pitch cycle C_k, the start point of the segment r_m, N_p, is in the interval of $[0, P_k - L]$.

In order to find the optimum segment r_m, we define a two-variable function $\Psi(\alpha, , \Delta)$ as follows:

$$\Psi(\alpha, \Delta) = \alpha(\tilde{c}_k, c_{k-1}) + \Delta(r_m, \tilde{c}_k) \tag{7}$$

α indicates the normalised correlation between the previous pitch cycle C_{k-1} and the modified pitch cycle $\widetilde{C}_k$. Δ is proportional to the reverse of discontinuity energies at the connection points. Since $|\alpha| < 1$, we normalise the discontinuity energies to maximum energy so that their reverse is fitted at the same range with α as follows:

$$\Delta(r_m, \tilde{C}_k) = -\frac{1}{M} E(N_p, \tilde{C}_k) + 1 \tag{8}$$

Where M depends on the number of connection points:

$$M = \begin{cases} 1 & \text{if} \quad r_m \text{ is discarded} \\ 2 & \text{if} \quad r_m \text{ is inserted} \end{cases} \tag{9}$$

and E (.) is the normalised discontinuity energy and given by:

$$E(N_p, \tilde{C}_k) = \frac{\left|\tilde{C}_k(N_p) - \tilde{C}_k(N_p + 1)\right|^2}{\left|Max(\tilde{C}_k) - Min(\tilde{C}_k)\right|^2} \tag{10}$$

With maximisation of the function $\Psi(.)$ in terms of the segment r_m , the optimum segment r_{m-opt} is obtained:

$$r_{m-opt} = \arg\max \ [\alpha(\tilde{c}_k, c_{k-1}) + \Delta(r_m, \tilde{c}_k)] \tag{11}$$

Using the optimum segment, r_{m-opt}, the modified local pitch cycle $\widetilde{C}_k$ is constructed by Eq. (3) or (6). This procedure is performed for all of the irregular pitch cycles.

Next, the vector V_{lp} is updated by using the new pitch pulse locations. If the standard deviation of vector V_{lp} is larger than the predetermined threshold, the procedure described in 3.3 is then repeated; otherwise, the pitch cycle modification is stopped.

By concatenation of the modified/unmodified local pitch cycles, the modified residual signal is constructed and is then passed to the synthesis filter, whose coefficients are pitch-cycle variables for the reconstruction of the speech signal. Figure 6 shows the original and the modified residual signals with the local pitch variations. In order to show the effect of modification, the pitch values are estimated by using the algorithm used in the standard 2.4 Kb/s MELP. The estimated pitch values are 72 and 73 samples for the second original and the modified frames, respectively. Figure 7 shows the normalised autocorrelation function used for estimating the pitch values. Obviously, due to having a high correlation at a pitch lag of 73 samples and low correlation for other pitch lag values, especially at a pitch lag of around 100 samples, the modified speech signal will produce a more reliable pitch estimate.

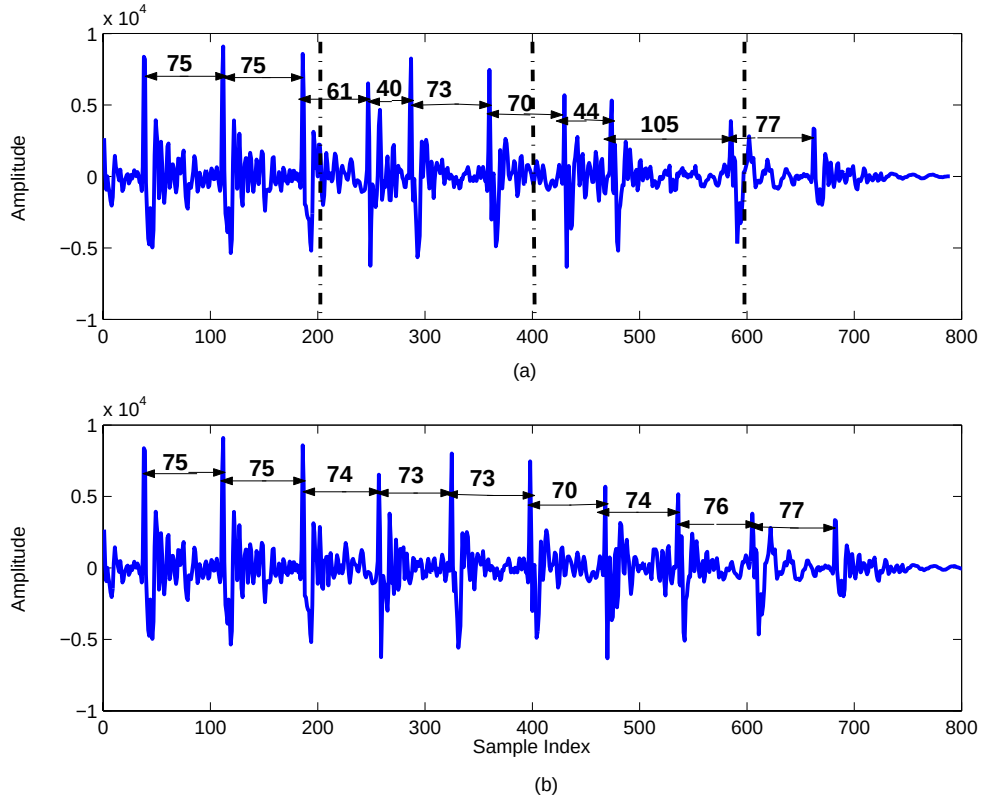

Figure 6. The original speech including irregular pitch variations (a) and the modified speech including smooth pitch evolution (b).

3.4. Pitch-cycle evolutions smoothing

In order to make sure that the pitch cycles evolve smoothly during a voiced frame and also frame-to-frame, we propose a method that increases the correlation of the pitch cycles based on a highly correlated target signal. Thus, the aim of this stage is to make more regular speech. This is performed for the voiced speech frames, regardless of whether of smooth or irregular evolutions. The inputs are the modified/unmodified residual signal and the predetermined local pitch locations. The output is the modified residual signal for which the pitch cycles evolve smoothly.

3.4.1. Target correlation concept

In the following sections, a pitch cycle is centred by a pitch pulse with the length of the interpolated pitch value. The basic idea of the target correlation approach is to modify low correlated pitch cycles of the residual signal. Thus, it is required that the target contains highly correlated pitch cycles. The low correlated cycles are searched by computing the normalised cross-correlation between the pitch cycles of residual and the relative pitch cycles of the target signal and comparing it against a threshold (Figure 8). These cycles are modified using the high correlated cycles as described in section 3.4.3. Thus, in the first step, the requirement is to construct the target signal.

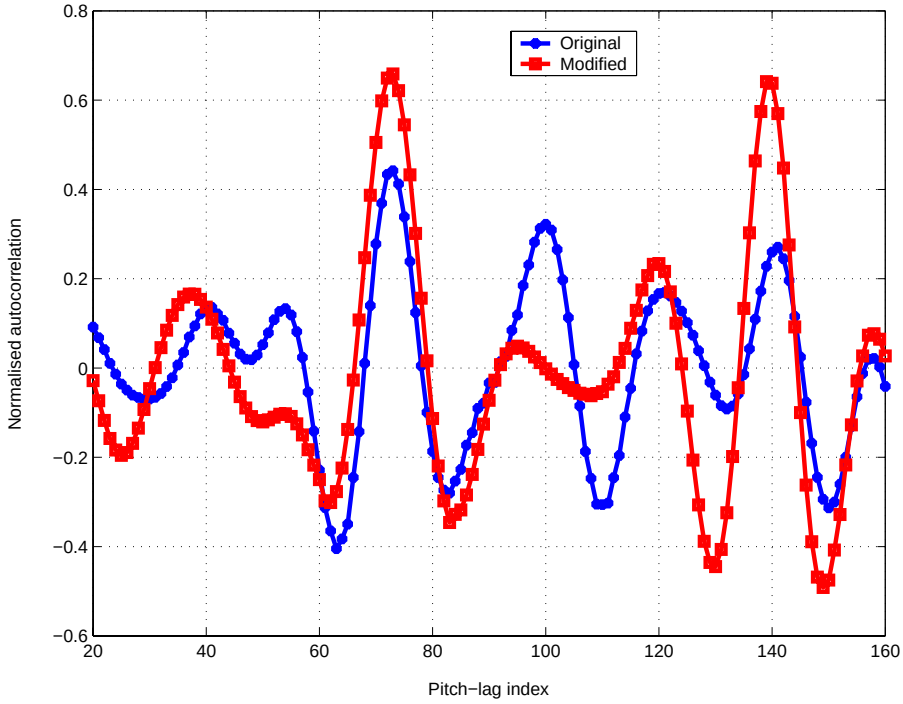

Figure 7. The normalised autocorrelation function computed for the second frame of the original (stars and solid line) and the modified (square and dash-dot line) speech.

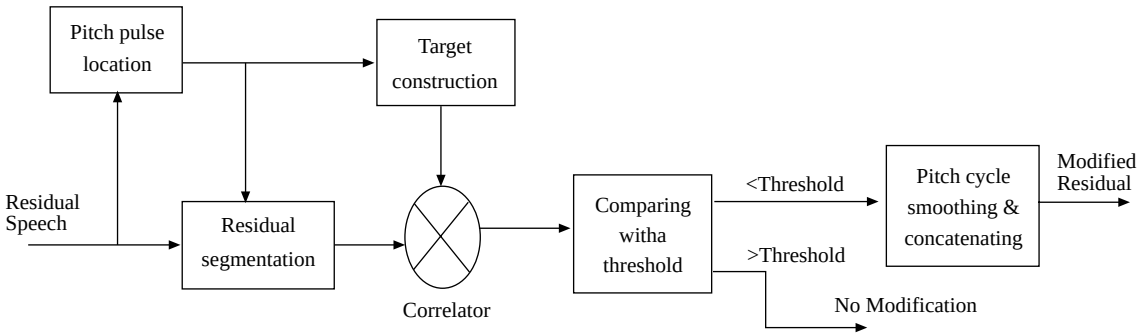

Figure 8. A block diagram of target correlating based on pitch pulse evolutions smoothing.

3.4.2. Target signal construction

In the first step, the local pitch cycles are extracted using the estimated pitch value P and pitch pulse locations as described in section 3.1. Each pitch cycle contains one centred pitch pulse to minimize boundaries energies. Therefore, the number of pitch cycles L is equal to the number of pitch pulse locations. Consider L consecutive pitch cycles in the LP residual signal. After normalisation to unit energy and appropriate alignment, we obtain the set of cycles $y_0, y_1, \ldots, y_{L-1}$. We consider the target cycle x to have maximum correlation with vectors $y_0, y_1, \ldots, y_{L-1}$.

$$x = \arg\max \sum_i \left\| y_i^T x \right\|^2 \qquad \|x\| = 1 \tag{12}$$

It is shown [6] that the vector x is given by:

$$YY^T.x = 0 \tag{13}$$

Where the $P \times L$ matrix Y $(L < P)$ is given by:

$$Y = [y_0\ y_1\ \ldots\ y_{L-1}] \tag{14}$$

Since the rank of the $P \times P$ matrix YY^T cannot exceed the rank of Y or Y^T (which is L) [6] and $L < P$, YY^T is not a full rank matrix and, therefore, there is a non-zero solution. Eq. (13) is solved by the singular value decomposition (SVD) method [7]. In this method, a matrix A can be rewritten as the product of a column-orthogonal matrix **U,** a diagonal matrix **W,** with positive or zero elements (the singular values), and the transpose of an orthogonal matrix **V.** Any column of **V,** whose corresponding $\mathbf{w}_j$ of **W** is zero, yields a solution. The vector, which maximizes Eq. 12, is the target cycle x.

Since the dimension of the matrix YY^T is $P \times P$, searching the target cycle x may require much computation, especially for large pitch values. In order to reduce the cost of computation, we restrict the length of the cycles $y_0, y_1, \ldots, y_{L-1}$ to the minimum pitch value (2.5 ms equal to 20 samples for sampling frequency of 8 kHz). Since the high-energy part of the cycles (which is pitch pulse region) has the main contribution in the cross-correlation value, the cycles $y_0, y_1, \ldots, y_{L-1}$ are centred at pitch pulse location with a length of $P = 20$.

3.4.3. Pitch-cycle smoothing using target correlation signal

After constructing the target signal, the normalised cross-correlation is computed between the cycles of the residual signal and the relative target cycles. Next, the ratio of the minimum to the maximum of the resulting values is computed. If the ratio is higher than a threshold (threshold = 0.85), no modification is performed; otherwise, the low-correlated cycles are replaced based on linear interpolation between the high-correlated cycles as given in Eq. (15).

$$\begin{aligned} &y(i) = \alpha C_{b0}(i) + (1-\alpha)C_{b1}(i) \\ &\tilde{C}_k(i) = \mu\, .y(i) \end{aligned} \qquad 0 \le i \le L \tag{15}$$

In this formula, C_{b0} and C_{b1} are the high correlated cycles before and after low correlated cycle C_k, $\alpha \le 1$ is searched so that the cross-correlation between the resulting cycle $\tilde{C}_k$ and its relative target cycle is maximised. μ controls the resulting cycle energy to be identical to the original one and is given by Eq. (16)

$$\mu = \sqrt{\frac{E_k}{E_y}} \tag{16}$$

where E_k and E_y are the energy of the original cycle and the reshaped cycle y given in Eq. (15). In order to reduce the discontinuity energy, an overlap-and-add method is applied at the connection points. Figure 9 shows the effect of smoothing pitch cycles on the cross-correlation between successive residual cycles and the normalised autocorrelation function. Obviously, due to higher correlation at a pitch lag of 45 samples and lower correlation at other pitch lags, especially at a pitch lag of 68 samples, a reliable pitch can be estimated.

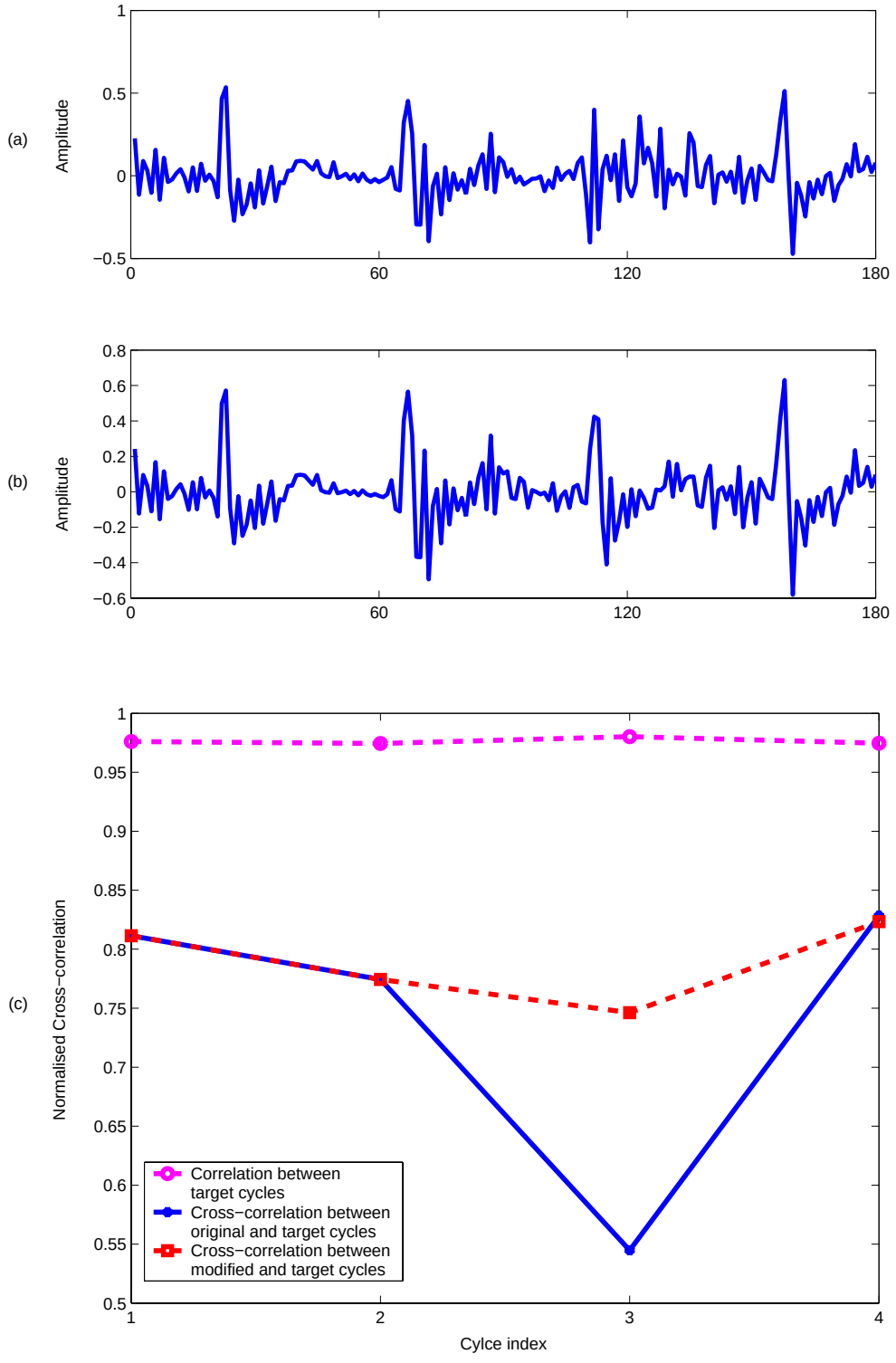

Figure 9. The original and the modified residual signals (a, b), the normalised cross-correlation between the target cycles and the original/modified residual cycles (c), and the normalised autocorrelation of the modified and original speech signals (d).

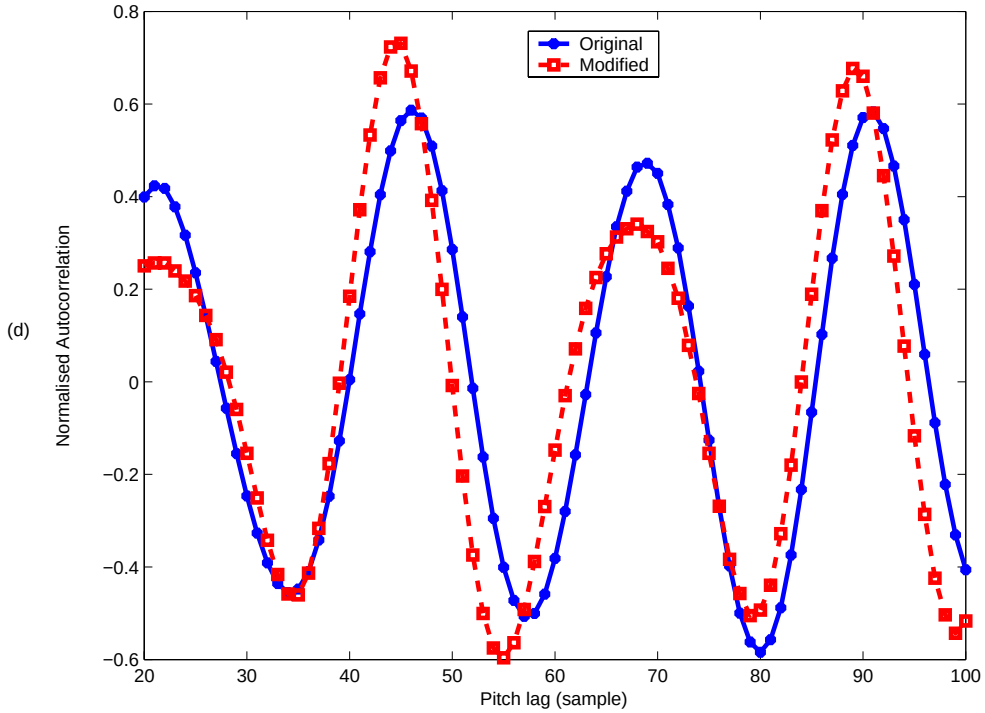

Figure 9. Contunied.

4. Pre-processor Evaluation

4.1. Experiments and Quantitative Results

The proposed pre-processor was evaluated in pitch and voicing-level estimations. Pitch prediction gain (PPG) in the time domain and synthetic spectral matching (SSM) in the frequency domain were used to measure the acuracy of the estimated pitch and voicing-level. In [8] it is shown that more accurate estimated pitch leads to higher PPG and lower SSM distortion. The estimated pitch periods of the modified and the original speech signals were used to calculate the PPG and SSM. The pre-processor detailed in section 3 was applied on sixty 8-second sentences from the NTT speech database [9]. The experiments show that the pre-processor gets activated in only 12 sentences. Thus, in order to evaluate the pre-processor, only short sentences and words were selected (for both male and female) for which the pre-processor was activated. The PPG and SSM were calculated for the modified and the relative original frames. Results are shown in Figure 10.

In order to show the effect of the proposed pre-processor on voicing-level estimation, we apply both the original and the modified speech separately as inputs to the standard 2.4 Kb/s MELP and compute the spectral distortion between the input and the synthesised speech for frequency bands of 0-500, 500-1000, 1000-2000, 2000-3000, and 3000-4000 Hz, and for the full frequency band. Results are shown in Figure 11.

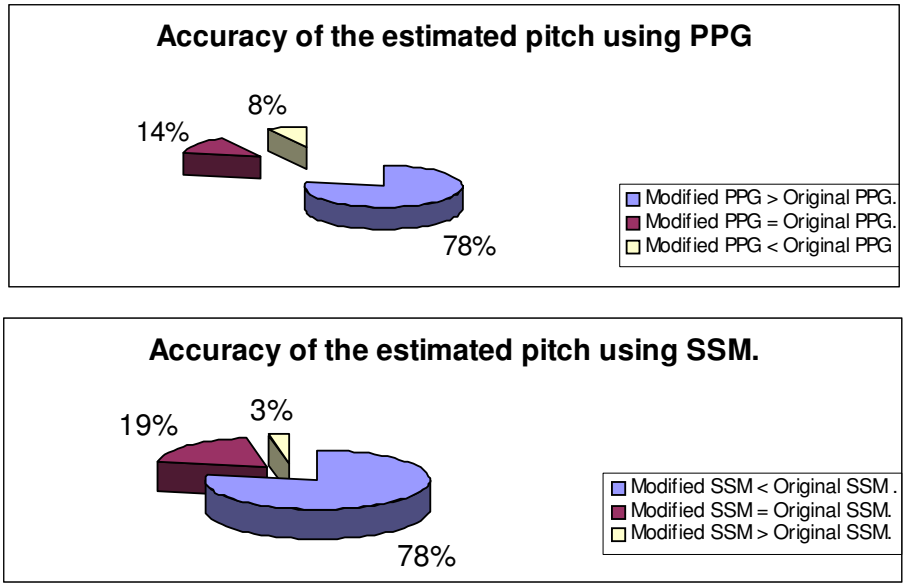

Figure 10. Accuracy of the estimated pitch values of the modified speech compared to the original using PPG and SSM.

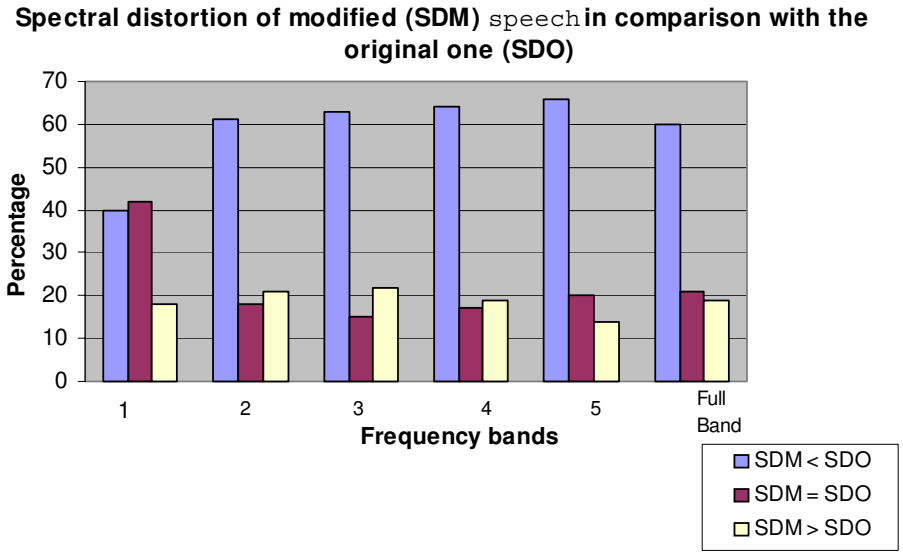

Figure 11. Accuracy of the estimated voicing level using the spectral distortion.

4.2. Subjective Listening Test Results

Due to the non-existence of a quantitative measure to evaluate the modified speech perceptually [10], subjective listening tests are proposed to be used as a perceptual measure. Therefore, for testing purposes, an AB-test [10] with 15 listeners was carried out on the original and the processed speech files of the database used in section 4.1. In order to evaluate the effect of the pre-processor on the speech coder, the original and the modified speech were applied to standard 2.4 kbps MELP and an A vs. B comparison test was carried out on the speech files. The results are shown in Table 1. The results obtained indicate that there is no statistical difference in perceptual quality between the original and the modified speech. However, the

pre-processor in combination with the MELP provides significantly better perceptual speech quality than the standard MELP.

Table 1. A vs. B test for modified and original speech, 2.4 kbps MELP in combination with pre-processor and 2.4 kbps MELP.

Speech Type	A (%)	Equal (%)	B (%)
Modified Speech vs. Original Speech	11.4	76.3	12.3
Pre-processor + 2.4kbps MELP vs. 2.4kbps MELP	55.6	30.1	14.3

5. Conclusions

In this paper, we discussed the effect of inaccurate pitch and voicing level estimation in a WI coder and a 2.4 Kbps MELP. The pitch and voicing level errors may significantly degrade subjective quality provided by the speech coders.

We proposed a new pre-processor, which modifies the residual signal to make more regular speech. This reduces the inaccuracy of the pitch and voicing level estimation in speech coders. The subjective listening tests show that the pre-processor maintains the original speech quality and can therefore be used in combination with any speech coder. The proposed pre-processor in combination with the standard 2.4Kbps MELP provides significantly better quality than MELP alone.

References

[1] W. B. Kleijn, "Waveform Interpolation for Speech Coding and Synthesis," in *Speech Coding and Synthesis,* pp. 175-208, Elsevier Science B.V., 1995.

[2] S. Lynn, C. Ronald and C. John, "MELP: The New Federal Standard at 2400 bps", in *IEEE ICASSP'97 Conference, Munich*, Germany, pp. 1591-1594.

[3] T. Eriksson and W. B. Kleijn, "On waveform interpolation coding with asymptotically perfect reconstruction," *Proc. Int. Conf. Acoust. Speech Sign. Process*, pp. 147-150, 1999.

[4] TIA/EIA, Enhanced variable rate codec, speech service option 3 for wideband spread spectrum digital systems, IS-127, 1997.

[5] Y. M. Cheng and D. O'Shaughnessy," Automatic and reliable estimation of glottal closure instant and period," *IEEE Transaction on Speech and Audio Processing,* Vol. 37, No. 12, December 1989.

[6] H. Farsi, Advanced Pre-and-post processing techniques for speech coding. Ph.D. thesis, University of Surrey, June 2003.

[7] William H. Press, Saul A. Teukolsky and William T. Vetterling, Numerical Recipes in C. Cambridge University Press, Cambridge, 2002.

[8] A. M. Kondoz, Digital speech: coding for low bit rate communication systems. John Wiley, UK, 1994.

[9] http://www.ntt-at.com/products_e/speech2002/

[10] W. B. Kleijn and K. K. Paliwal, Speech Coding and Synthesis. Elsevier Science, Amsterdam, The Netherland, 1998.