

An Improved Packet Loss Concealment Method for Mobile Audio Coding

Kai Xiang^{1,2} and Ruimin Hu^{1,*}

¹National Engineering Research Center for Multimedia Software, School of Computing, Wuhan University, Wuhan, China

²School of Information Management, Hubei University of Economics, Wuhan, China

Abstract: In this paper, we propose an improved packet loss concealment (PLC) algorithm for ACELP-based speech coders in order to improve the quality of decoded speech at the receiver. Conventional receiver-based PLC algorithms in the AMR-WB+ speech codec are usually based on speech correlations to reconstruct the decoded speech of the lost frames by using parameter information obtained from the previous or adjacent correctly received frames. However, this approach has difficulties in reconstructing accurate voice signals when consecutive packet loss and bursts of packet loss occur. Thus, we propose a new PLC approach based on data-driven principle to recover the Immittance Spectral Frequency (ISF) coefficients of lost packet, referring to pitch period parameters of adjacent correctly received voice packets. The proposed PLC algorithm is designed for ACELP-based audio codec. We make the performance comparison between the data-driven method and the PLC algorithm currently employed in AMR-WB+ speech codec through the PESQ objective evaluation test and MOS subjective evaluation test. The results prove that our algorithm can achieve better speech recovery quality than original PLC algorithm employed in AMR-WB+ speech codec.

Keywords: AMR-WB+, packet loss concealment, speech coding, VoIP.

1. INTRODUCTION

VoIP (Voice over Internet Protocol) is a recently emerging technology in the area of voice communications. As more and more cell phone users are taking advantage of mobile VoIP services, the markets of VoIP-based communication systems and consumer electronics grows dramatically. Unfortunately, it is a major challenge that VoIP has not yet provided toll-quality voice over the networks [1, 2]. A number of factors may affect the service quality of VoIP, such as packet loss, packet delay, and jitter [3, 4]. Considerable efforts have been made within different layers of current communication systems to reduce delay, smooth jitters and recover loss. We focus on the work that enhances VoIP speech quality at the receiver of a mobile Internet phone.

For concealing packet loss or recovering lost packets at the receiver, several methods have been developed. These methods can be roughly divided into several categories: In “waveform substitution” method, the missing frames are replaced by the other already-received frame using pitch replication [5, 6], pattern matching [7, 8] and “model-based repetition”. Among these, the model parameters of the previously received signals are extrapolated or interpolated, which leads to the recovery of the lost signal part [9-12]. Gündüzhan *et al.* proposed a linear prediction (LP)-based packet loss concealment (PLC) to construct the excitation signal for the autoregressive model of the missing segment by repeating the excitation signal of the previously received

speech with a periodicity equal to the pitch period [9]. Wong *et al.* proposed to classify the missing segment into voiced, unvoiced, or partially voiced types, and then construct the excitation signal correspondingly [13]. More recently, PLC method based on a hidden Markov model was proposed [14]. Miralavi *et al.* proposed a state-space-based method, which uses a Kalman filter and an estimated excitation signal to predict and estimate the speech signal [15].

Model-based PLC algorithms effectively improve the quality of the reconstructed speech in voice over packet networks. However, consecutive reconstructed speech is sensitive to pitch estimation errors, such as doubling or halving periods and do not offer natural sound. Therefore, accurate pitch estimates are needed for generating model-based synthesized speech and gradual muting for the duration of the loss period.

In the paper, we propose an efficient receiver-based enhancing method of VoIP speech quality. The important functionality to be implemented at the receiver is PLC scheme using the effective prediction model based on the variety trend of pitch period parameters of good speech packets to predicate the ISF parameters of the lost signal packets.

2. PLC SCHEME IN AMR-WB+

AMR-WB+ speech codec is based on the ACELP (Algebraic Code-Excited Liner Prediction) mode. The speech coding characteristic parameters are pitch period parameter, ISF coefficients, and etc. The pitch period parameter and ISF spectrum coefficients play an important role in speech signal

recovery. We then introduce the recovery method of the two parameters.

2.1. Recovery of Pitch Period Parameter

The method of recovery of the pitch period parameter in AMR-WB+ encoder is divided into two parts: the error detection unit and the error concealment unit. The error concealment unit utilizes the state transition machine to transit among seven states, and there are twenty-six different kinds of pattern sets. The decoder needs to restore and recover the pattern sets of the lost frames from the pattern sets of correct received ones. When the lost frame is with the type of ACELP, the pitch period parameter is reconstructed by the following algorithm 1.

Algorithm 1 Recovery of The Pitch Period

```

1: Input:  $bfi$ 
2: Output:  $T_0, T_0\_frac$ 
3: If ( $bfi$ ) Then
4:  $old\_T_0\_frac \leftarrow old\_T_0\_frac + 1$ ;
5: If ( $old\_T_0\_frac > 3$ ) Then
6:  $old\_T_0\_frac \leftarrow old\_T_0\_frac - 4$ ;
7:  $old\_T_0 \leftarrow old\_T_0 + 1$ ;
8: End If
9: If ( $old\_T_0 \geq PIT\_MAX$ ) Then
10:  $old\_T_0 \leftarrow PIT\_MAX - 5$ ;
11: End If
12:  $T_0 \leftarrow old\_T_0$ ;
13:  $T_0\_frac \leftarrow old\_T_0\_frac$ ;

```

The definition of parameters in algorithm 1 is described as follows:

bfi —the flag of error frame, T_0 —the integer part of the reconstructed pitch period, T_0_frac —the fractional part of the reconstructed pitch period, old_T_0 —the integer part of the pitch period of the previous frame, $old_T_0_frac$ —the fractional part of the pitch period of the previous frame, PIT_MAX —the upper limit of the pitch period.

The defects of recovery method with the pitch period parameter in AMR-WB+ codec are that the pitch delay increases 1/4 sample at each time. The step size is too small to curb excessive cyclical effect, so that the accumulation error increases even with the successive lost frames.

2.2. Recovery of ISF Coefficients

ISF coefficients used in AMR-WB+ speech coder are of the 16-order. When the frame errors occur, AMR WB+ voice decoder obtains ISF spectrum coefficients of the error frame by shifting the coefficients of previous frames to adaptive mean value of ISF spectrum coefficients. The value can be calculated through the Formula (1).

$$ISF_q(i) = \alpha * past_ISF_q(i) + (1 - \alpha) * ISF_{mean}(i) \quad (1)$$

Where, i is from 0 to 15, $\alpha = 0.9$, represents the value of ISF vector of current frame, $past_ISF_q(i)$ represents the value of ISF vector of previous frame, $ISF_{mean}(i)$ represents the adaptive mean value of ISF vector, which can be calculated by Formula (2).

$$ISF_{mean}(i) = \beta * ISF_{const_mean}(i) + (1 - \beta) * ISF_{adaptive_mean}(i) \quad (2)$$

Where, the range of i values from 0 to 15, $\beta = 0.25$, represents the long term average constant

vector of ISF vector, represents the averaged value of ISF vector of three consecutive good frames, which can be calculated by Formula (3).

$$ISF_{adaptive_mean}(i) = \frac{1}{3} \sum_{i=0}^2 past_ISF_q(i) \quad (3)$$

The drawbacks of the recovery method with ISF spectrum coefficients are analyzed as follows:

(1) The severity of the bad frames is not considered. When consecutive bad frames occur, the correlations between the adjacent good frames and the bad frames will be vanished.

(2) In fact, ISF coefficients of the adjacent good frames and the bad frames have more relevances than other good frames. We can ignore the other good frames to reduce the computational complexity.

3. PROPOSED OPTIMAL METHOD

3.1. Recovery of Pitch Period Parameter

The algorithm in AMR-WB+ standard do not consider the type of the adjacent frames, we proposed a way to adaptively select the error hidden schemes by predicting the variation trend of the pitch period of adjacent frame. The idea of the algorithm is to maintain the pitch information of the good frame that is the one before the error frame. And then predict the trend of change with the pitch information. If pitch period increases, the pitch period of lost frame will also increase, and vice versa. The prediction process is shown in Fig. (1).

We adopt a linear prediction model to predict pitch period of current loss of speech frames based on the variation trend. The model can be described as Formula (4).

$$P'_{lost}(k) = a + b * P(k) \quad (4)$$

Previous five correctly received pitch frames are used to predict pitch information of the current frame loss. The previous pitch value received correctly is denoted as $P(k)$, the minimum mean square error criterion is used to estimate the parameters b and a . The Formula (5) is shown as follows.

$$E = \sum_{i=0}^4 [P'_{lost}(i) - P(i)]^2 = \sum_{i=0}^4 [(a + b * P(i)) - P(i)]^2 \quad (5)$$

The calculating results are shown as follows.

$$a = \frac{3 \sum_{i=0}^4 P(i) - \sum_{i=0}^4 iP(i)}{5}, b = \frac{\sum_{i=0}^4 iP(i) - 2 \sum_{i=0}^4 P(i)}{10} \quad (6)$$

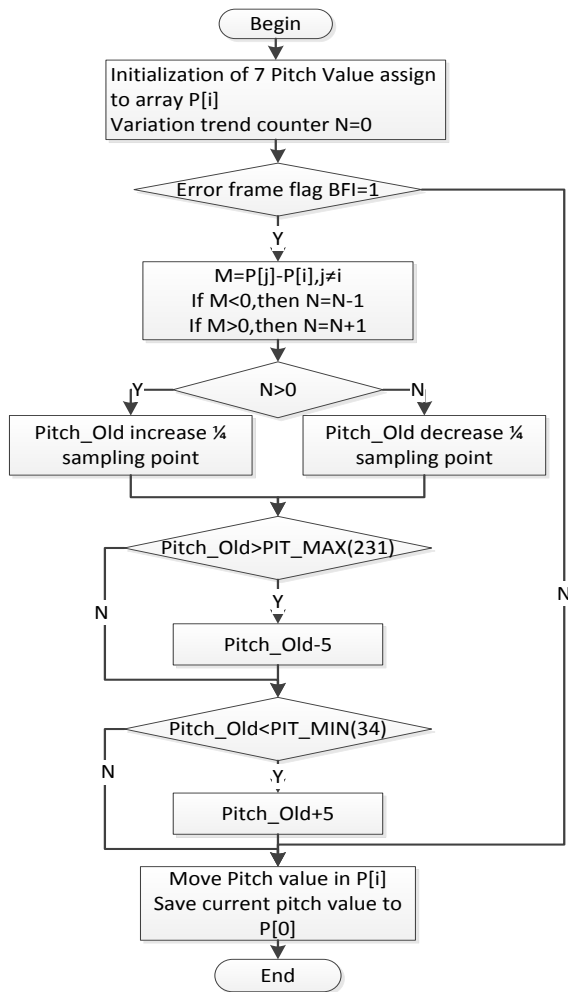


Fig. (1). Flowchart of the prediction method.

3.2. Recovery of ISF Coefficients

SPSA (Simultaneous Perturbation Stochastic Approximation) is an online data-driven approach without a model proposed by professor Spall in 1993 [16]. The data-driven method does not need to make modeling analysis of the con-

trol system but only needs to determine the control parameters effectively by processing input and output data of closed loop.

The data-driven approach can be used to recover the ISF coefficients. The principle of the method is to monitor a reliable speech frame parameter and then to predict whether the variation trend of the adjacent frames received are correctly. With the information achieved, then guide the error hidden process of the other speech frame parameters. The process can be described as in Fig. (2).

The process of recovery of the ISF coefficients can be divided into three steps; State Classification, Model Training and Parameter Prediction. The prediction process can be described as shown in Fig. (3).

The description of the three steps is shown as below:

(1) **State Classification:** Before the processing of the speech signal, we need to sub-frame processing for the input code stream. Considering the differences of state transitions of different frames, the state needs to determine its transfer characteristics of each frame based on the properties of the speech. Before training the model, we can decide which frames are chosen for model training before the first M frames according to the prior knowledge. Prior knowledge is stored in a ROM table of decoder.

(2) **Model Training:** Training the model utilizes the mono test sequences, including speech sequences, music sequences and mixed sequences. The linear regression model utilizes the data of the model training set with a prior knowledge, while training the selected parameters to make the minimum mean square error based on the minimum mean square error criterion. The parameter values obtained by training are stored in the ROM table of decoder.

(3) **Parameter Prediction:** In the regression process, the decoder for ISF coefficients recovery of lost frames, received with reference to neighboring frames, to select the correct model parameter values from a ROM table. The model parameters are shown in Table I.

The current ISF coefficients can be predicted by Formula (7).

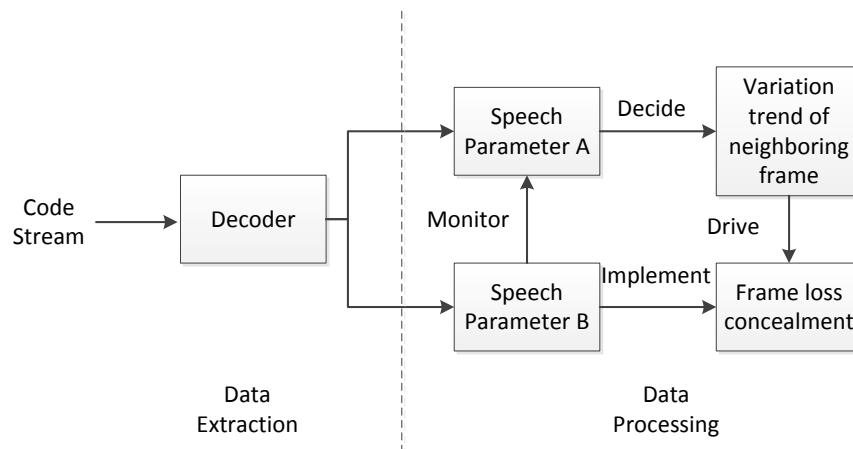


Fig. (2). Flowchart of the data-driven method.

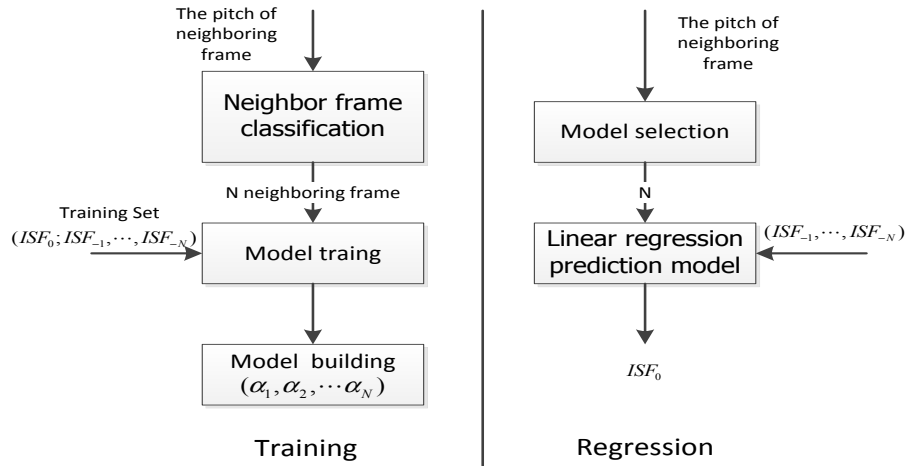


Fig. (3). Flowchart of the prediction process.

$$ISF_0 = \alpha_{1,i} ISF_{-1} + \alpha_{2,i} ISF_{-2} + \dots + \alpha_{N,i} ISF_{-N} + \beta_i ISF_{CONST_MEAN}, i = 1 \dots 2^M \quad (7)$$

4. EXPERIMENTS AND RESULTS

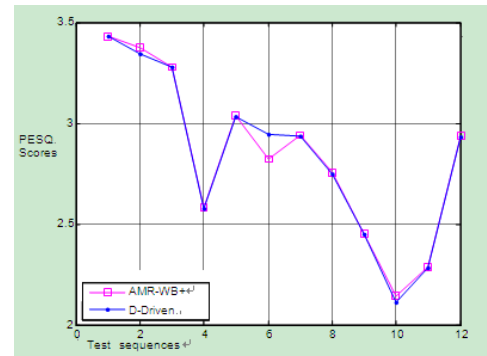
We made the experiments when compared with the original frame loss recovery method employed in AMR-WB+ codec. In the experiments, we use the same group of test sequence, with different emergent loss circumstances. The data-driven method and original algorithm in AMR-WB+ codec were separated and used to decode and recover the lost data, and analyze the output sequence. Then the MOS subjective evaluation method and PESQ objective evaluation method are used to evaluate the results of the redemptive audio sequences.

The twelve test sequences were from MPEG standard test sequences, sampled at 16 kHz mono audio sequences. Coding mode selects the sample rate at 25.6 kHz, 12.0 kb/s bit rate, and uses only the ACELP coding mode. Fig. (4) is the performance comparison for data-driven algorithm and the original algorithm with the burst packet loss rate at 1%, 3%, 5% and 10% with the PESQ evaluation method.

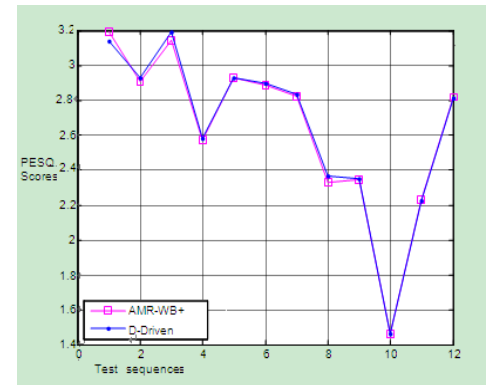
From Fig. (4), we can see that PESQ scores of the proposed algorithm are significantly higher than that of the original algorithm in the majority of the test sequences for four kinds of burst packet loss rate.

Table 1. Prediction model parameters.

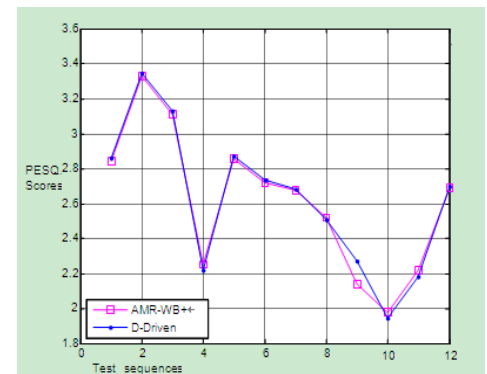
State Transition	Model Parameters
$(s_{1,-M}, s_{1,-M+1}, \dots, s_{1,-1})$	$(N_1, \alpha_{1,1}, \alpha_{1,2}, \dots, \alpha_{1,N_1}, \beta_1)$
$(s_{2,-M}, s_{2,-M+1}, L, s_{2,-1})$	$(N_2, \alpha_{2,1}, \alpha_{2,2}, \dots, \alpha_{2,N_2}, \beta_2)$
\vdots	\vdots
$(s_{2^M,-M}, s_{2^M,-M+1}, L, s_{2^M,-1})$	$(N_{2^M}, \alpha_{2^M,1}, \alpha_{2^M,2}, \dots, \alpha_{2^M,N_{2^M}}, \beta_{2^M})$



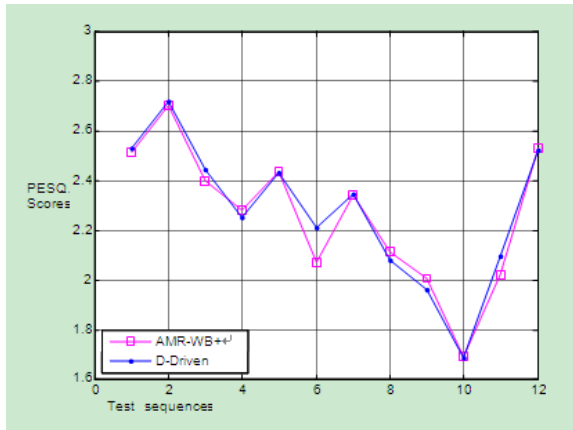
(a) Packet loss rate at 1%



(b) Packet loss rate at 3%

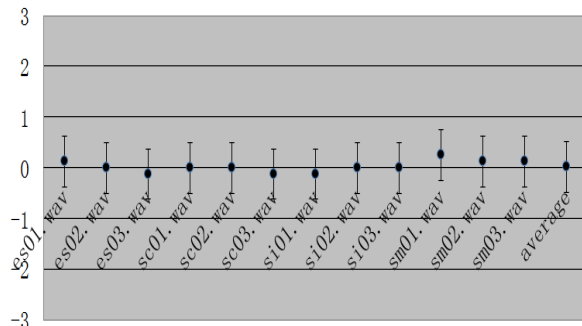


(c) Packet loss rate at 5%

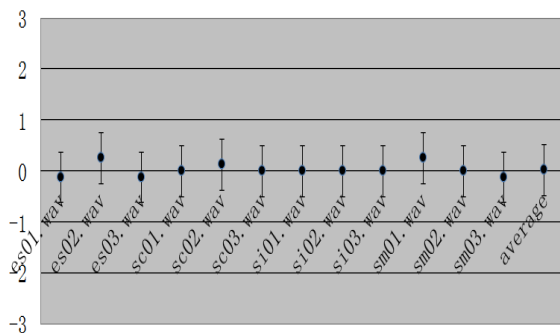


(d) Packet loss rate at 10%

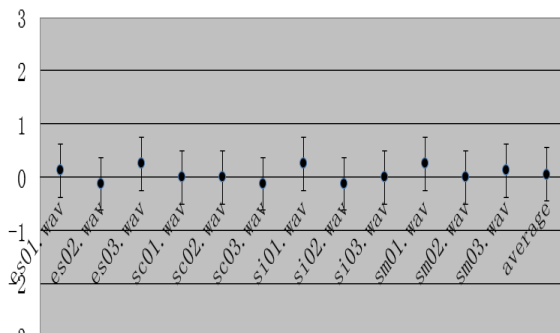
Fig. (4). Performance comparison at different packet loss rate.



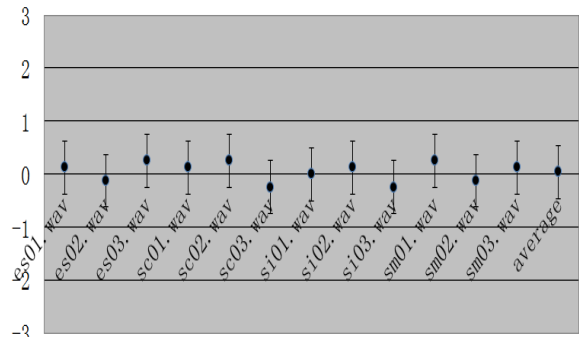
(a) Packet loss rate at 1%



(b) Packet loss rate at 3%



(c) Packet loss rate at 5%



(d) Packet loss rate at 10%

Fig. (5). MOS scores at different packet loss rate.

Ten testers that work actively in the domain of audio compression participated in the subjective test based on the proposed algorithm and the original algorithm. Test result is shown in Fig. (5).

From the results, our algorithm achieves improvement at the low packet loss rate of 1% and 3% with the comparisons of conventional algorithm. In the high dropout rate at 5% and 10%, MOS scores changes in amplitude, some of subjective scores of the test sequence decline, but overall most improved MOS scores.

CONCLUSION

In this paper, we proposed a new ACELP-based packet loss concealment method to recover the essential decoder parameters pitch period and ISF coefficients for lost packets. In particular, the introduction of data-driven control theory approach is used to restore ISF coefficients. Under multiple sound sources with different burst packet loss rate, our data-driven method produce better decoded audio quality than conventional method employed in AMR-WB+ according to the test results of PESQ objective evaluation and MOS subjective evaluation. The optimized packet loss concealment method can also be used in other ACELP-based audio codec to improve the decoded audio quality. Our future work includes the recovery of the fixed code book gain and the adaptive code book gain with reference to the data-driven method. Because of the importance of 3D audio coding in applications, the recovery of 3D audio signal processing parameters such as ILD, ITD, and IACC will also be included in follow-up study on PLC method for 3D stereo audio codec.

CONFLICT OF INTEREST

The author confirms that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

This work was supported by the Major National Science and Technology Special Projects (2010ZX03004-003-03), National Nature Science Foundation of China (No. 61231015, No. 61102127, No. 61201340, No.61201169,

No.61272278), The Natural Science Foundation of Jiangsu Province Youth Project (No. BK2012200), Foundation of Department of Education of Hubei Province (No. Q20132207).

REFERENCES

- [1] M. Gidlund and J. Ekling, "VoIP and IPTV distribution over wireless mesh networks in indoor environment," *IEEE Transactions on Consumer Electronics*, vol. 54, pp. 1665-1671, Nov. 2008.
- [2] K.-K. Tseng, Y.-C. Lai, and Y.-D. Lin, "Perceptual codec and interaction aware playout algorithms and quality measurements for VoIP systems," *IEEE Transactions on Consumer Electronics*, vol. 50, pp. 297-305, Feb. 2004.
- [3] A. Shallwani, and P. Kabal, "An adaptive playout algorithm with delay spike detection for real-time VoIP," *Electrical and Computer Engineering, IEEE CCECE 2003*, vol. 2, pp. 997-1000, May 2003.
- [4] A. Sat, and B. W. Wah, "Analyzing voice quality in popular VoIP applications," *IEEE Transactions on Multimedia*, vol.16, no. 1, pp. 46-59, Jan. 2009.
- [5] N. Aoki, "A VoIP packet loss concealment technique taking account of pitch variation in pitch waveform replication," *Electronics and Communications in Japan*, vol. 89, no. 3, pp. 1-9, Mar. 2006.
- [6] N. Aoki, "VoIP packet loss concealment based on two-side pitch waveform replication technique using steganography," *TENCON 2004 IEEE Region 10 Conference*, vol. 3, pp. 52-55, Nov. 2004.
- [7] V.P. Bhute, and U.N. Sharawankar, "Speech packet concealment techniques based on time-scale modification for VoIP," *International Conference on Computer Science and Information Technology*, pp. 825-828, Aug. 2008.
- [8] H. Sanneck, A. Stenger, K. B. Younes, and B. Girod, "A new technique for audio packet loss concealment," *Global Telecommunications Conference*, pp. 48-52, Nov. 1996.
- [9] E. Gündüzhan, and K. Momtahan, "A linear prediction based packet loss concealment algorithm for PCM coded speech," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 8, pp. 778-785, Nov. 2001.
- [10] J. Lindblom, and P. Hedelin, "Packet loss concealment based on sinusoidal modeling," *IEEE Workshop on Speech Coding*, pp. 65-67, Oct. 2002.
- [11] J.-H. Chen, "Packet loss concealment for predictive speech coding based on extrapolation of speech waveform," *ACSSC 2007 Conference Record of the Forty-First Asilomar Conference on Signal, Systems and Computers*, pp. 2088-2092, Nov. 2007.
- [12] K. Kondo, and K. Nakagawa, "A speech packet loss concealment method using linear prediction," *IEICE Transactions on Information and System*, vol. E89-D, no. 2, pp. 806-813, Feb. 2006.
- [13] J.-F. Wang, J.-C. Wang, J.-F. Yang, and J.-J. Wang, "A voicing driven packet loss recovery algorithm for analysis-by-synthesis predictive speech coders over Internet," *IEEE Transaction on Multimedia*, vol. 3, pp. 98-107, Mar. 2001.
- [14] C. A. Rødbro, M. N. Murthi, S. V. Andersen, and S. H. Jensen, "Hidden Markov model-based packet loss concealment for voice over IP," *IEEE Transaction on Audio Speech Language Process*, vol. 14, pp. 1096-1623, Sep. 2006.
- [15] S. R. Miralavi, S. G. Aidin, and T. A. Rahimi, "Kalman filter method for packet loss replacement in presence of background noise," *9th International Multi-Conference on Systems, Signals and Devices*, pp. 1-4, Mar. 2012.
- [16] Spall J.C., Cristion John A., "Model-free control of general discrete-time systems," *Decision and Control*, vol.3, pp. 15-17, Dec. 1993.

Received: September 22, 2014

Revised: November 03, 2014

Accepted: November 06, 2014

© Chuanhe et al.; Licensee Bentham Open.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.