International Journal of Scientific & Engineering Research, Volume 6, Issue 4, April-2015
ISSN 2229-5518

Comparative Study of Packet Loss Concealment Methods in Transmission of Voice over IP (VoIP)

Hamza KHEDDAR, Bachir BOUDRAA

Speech Communication and Signal Processing Laboratory, Faculty of Electronics and Computer science, University of Sciences and Technology Houari Boumediene (USTHB), P.O. Box 32, El-Alia, Bab-Ezzouar, Algiers, 16111, Algeria

hamza.kheddar@gmail.com , bboudraa@usthb.dz

Abstract —The work consists of developing two MELP coders operating respectively at 1.2 kbps and 2.4 kbps. The improvements of the coder operating at 2.4 kbps by implementing packet loss concealment (masking) techniques based on the receiver which consists of interleaving information frames. These techniques are compared and mentioned on [1]. For this purpose, we extended this comparison to a method previously developed in our laboratory, called multiple descriptions coding (MDC) described on [2], which use two MELP coders operating respectively at 2.4 kbps for the first description and 1.2 kbps for the second description. We used as evaluation technique a method called PESQ (Perceptual Evaluation of Speech Quality) standardized by ITU-T.

Keywords **— Interleaving, MELP Coding, MDC, PLC, PESQ.**

I. Introduction

In a VoIP system, at the receiver, some packet may be missing; this packet loss degrades the voice quality and results in ruptures in the conversation and a sense of hashing of speech. It is therefore essential to establish a mechanism for hiding the losses. Several algorithms for masking packet loss also called PLC (Packet Loss Concealment) are used both at the transmitter and/or at the receiver.

In the laboratory of coding at the institute of electronic, it was developed a MELP coder for VoIP using multiple description coding method called (MDC), to combat packet loss and increase the robustness of systems against these losses. This multi-description contains in a single packet both two MELP encoders. The first run at 2.4 kbps is used to obtain a good quality of voice after a good transmission. The second run at 1.2 kbps is used to recover any loss of packets. Our work is to improve the Codec MELP by the implementation of lost frames concealment techniques based on the receiver. These techniques consist on interleaving of information frames. We then present a comparative study between the best interleaving method and the MDC method. The comparative assessment was made using a method called PESQ (Perceptual Evaluation of Speech Quality).

II. MELP Coding

The MELP has now become the new military and Federal standard for speech at 2.4 kbps, replacing federal standards FS-1015 (LPC-10) and FS-1016 (CELP) speech that produce poor quality at this rate. Implementing a MELP coder involves four steps: *analysis*, *encoding*, *decoding* and *synthesis* [3] – [4].

A) MELP Encoder

In the synthesis MELP, LP (linear prediction) all-pole filter is excited by a signal built from periodic contributions and noise. At the encoder (Fig.1), the LP parameters are first determined. The residual is then obtained. The pitch is estimated from the low-pass filtering of the speech signal. The voicing strengths are evaluated based on the correlation maxima of the band-pass filtered signal. Voicing determines how the periodic parts and noise contribute to the excitement of the LP in specific frequency bands. Describe, in fact, the presence of periodicity in the function of the frequency signal. The Fourier coefficients define the spectral characteristics of the periodic excitation of LP. They are usually calculated from the FFT of the signal. Determining the gain can be performed either on the LP residual or directly on the speech signal, synchronously or with a fixed length window [5].

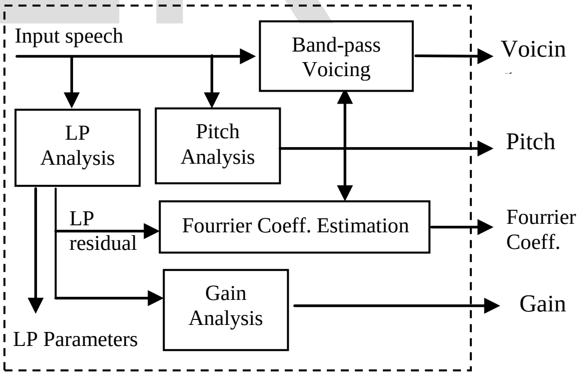

Fig.1 Basic schema of MELP Encoder.

B) MELP Decoder

At the decoder (Fig.2), the periodic part of the excitement is generated from the Fourier coefficients interpolated. Fourier synthesis is applied to spectra in which the Fourier coefficients are placed at the harmonic frequencies derived from the interpolated pitch. The sound of excitement is generated from white noise. The frequency bands of the periodic part of the signal and noise are shaped through time

International Journal of Scientific & Engineering Research, Volume 6, Issue 4, April-2015
ISSN 2229-5518

domain filtering according to the transmitted voicing information. The two components of the excitation are added and the signal is scaled by the encoded gain. Finally, the linear prediction synthesis is performed [5].

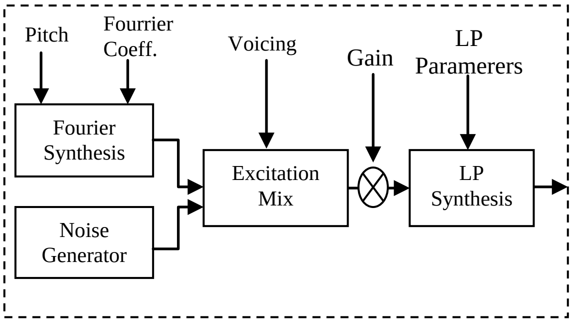

Fig.2 Block diagram of MELP decoder.

III. Interleaving

To achieve a voice in real-time high quality, packet loss concealment mechanism must be put in place. Several packet loss concealment algorithms PLC (Packet Loss Concealment) are used either at the transmitter or at the receiver [6]-[8]. Interlacing is an effective method to disperse packet loss bursts into a series of small losses. As a result, the errors will be produced on relatively short code words and the listener will be able to mentally interpolate small gaps. The intelligibility of speech is then preserved. This part focuses on the theory of some implemented interleaving methods.

1) *Convolutional interleavers*

A convolutional interleaver can be modeled as a shift register arrangement, each having a characteristic vector. In a convolutional interleaver of degree d, the input vector sequence is divided into d subsequences. Each sub-sequence consists of a different number of connected shift registers, which thus corresponds to a different delay according to the number of feature vectors that are stored there [9]. A convolutional interleaver of degree 4 is illustrated in fig.3. A convolutional interleaver size N ($d = \sqrt{N}$ subsequences) takes the form:

$$\pi_{conv}(i) = i - d(i \bmod d) \quad (1)$$

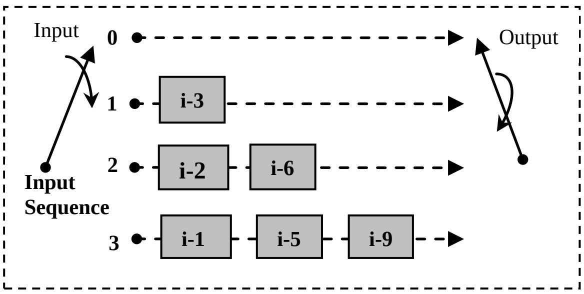

Fig.3 Convolutional interleaver of degree d = 4.

2) *Decorrelated convolutional interleaver*

The decorrelated convolutional interleaver introduced the same decorrelated structure of the convolutional interleaver described above. A decorrelated convolutional interleaver is formed by permuting the order in which the individual sub-sequences are accessible. For a decorrelated convolutional interleaver of size d, the order in which the sub-sequences are accessible is defined by the permutation P of the length d [9]. For example, a decorrelated interleaver of size 4, using the permutation $P = \{1\ 3\ 0\ 2\}$, is shown in fig.4. In the general case, at time index i, a feature vector will be delivered to subsequence $P_{(i \bmod d)}$, which has a delay of $d(P_{(i \bmod d)})$ frames. Thus:

$$\pi_{dec}(i) = i - d\left(P_{(i \bmod d)}\right) \quad (2)$$

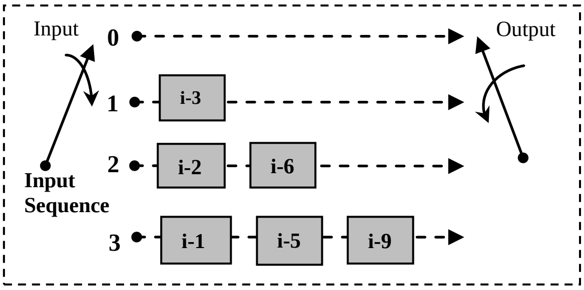

Fig.4 Decorrelated convolutional interleaver of size d = 4 and for a permutation P = {1 3 0 2}.

3) *Interleaving by grouping*

The grouping process and interleaving produces a vector of coefficients of $v = \{v_i | i = 1, \ldots, ML_{M-1}\}$. Fig.5 shows a simple example where $M = 3$ is the block index and $L = 2$ is the half of the length of the block of the analysis window. The coefficients are grouped and interleaved by using the following three steps: In (1), each line corresponds to a block and each block, the coefficients are grouped into frames. In (2), the frames of smallest scale (block 0) are interleaved in pairs with the immediate upper frame in the (block 1). This first step produces two new frames of interleaved coefficients. In (3), this two frames are interleaved with the frame of largest scale (block 2) in such a way that the resulting vector has alternatively a coefficients of each block: one of block 2, followed by one of block 1, followed by one of block 0, followed by one of block 2, and so on [10].

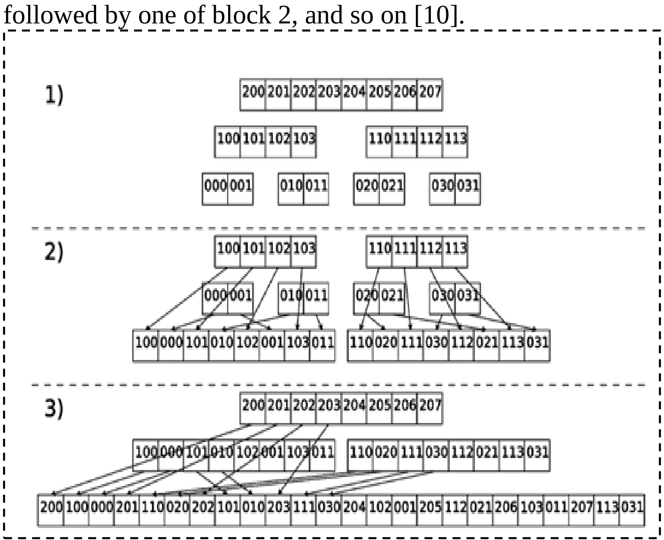

Fig.5 processus d'entrelacement avec M = 3, L=2.

International Journal of Scientific & Engineering Research, Volume 6, Issue 4, April-2015
ISSN 2229-5518

4) *Optimal spread block interleavers*

A block interleaver of degree d operates by re-arranging the transmission order of a $d \times d$ block of input vectors. Two block interleavers, $\pi_{bloc\,1}$ and $\pi_{bloc\,2}$, [8] are considered optimal in terms of maximizing their spread for given degree, and are given [9].

$$\pi_{bloc\,1}(id + j) = (d - 1 - j)d + i \quad ou \quad 0 \leq i, j \leq d - 1 \qquad (3)$$

$$\pi_{bloc\,2}(id + j) = jd + (d - 1 - i\ ) \quad ou \quad 0 \leq i, j \leq d - 1 \qquad (4)$$

The operation of these interleavers can be considered as a rotation of d × d of the feature vectors located in the buffer memory (buffer) either 90 ° clockwise or 90 ° anti-clockwise, as shown in fig.6.

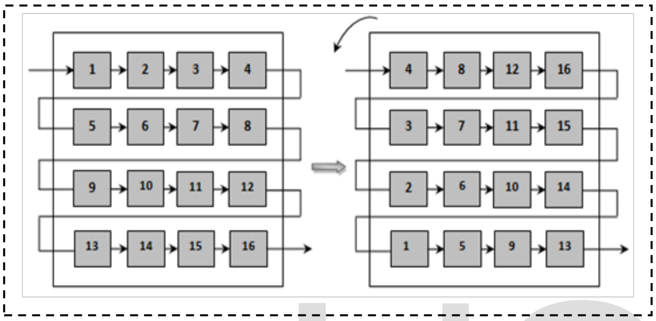

Fig.6 Rotation of buffer by 90° anti-clockwise [9].

IV. MULTIPLE DESCRIPTION CODING (MDC)

The Multiple Description Coding (MDC) is an interesting technique to fight against loss and transmission errors. In fact, we code the signal on two descriptions in the same packet. The first description uses a MELP coder and encodes the current frame **Fn** at 2.4 kbps while the second description uses another MELP encoder running at 1.2 kbps to encode the three frames **Fn+1**, **Fn+2**, **Fn+3**, following **Fn**. Obviously, **Fn** is used to reconstruct the signal with a good quality while **Fn+1**, **Fn+2** and **Fn+3** are used to recover the eventual loss packets. Indeed, the second description thus formed contributes to reconstruct the speech when one, two, three or even four successive packets are lost. This redundant information has not the same quality of the extracted one from **Fn** as it is roughly quantified. It only helps to reconstruct an intelligible speech when packets are lost. The packetization scheme is shown in figure 2. Notice that the MELP 2.4 operates on a frame of 22.5 ms, while the MELP 1.2 operations are achieved on a 67.5 ms frame [11]-[12].

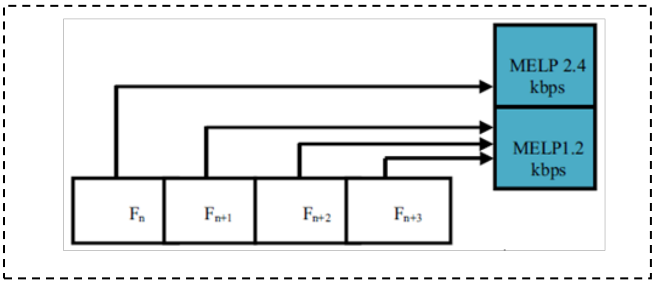

Fig. 7 Packetization using two descriptions

The MELP coder parameters are given in Tab.1. The two coders encode the fundamental frequency (pitch), the flag of the aperiodicity, the five bands of voicing, two gains corresponding to the energy of two half-frames, ten LPC coefficients converted into LSF and spectral amplitudes of ten harmonics of the pitch. A fine description at 2.4 kbps is required in order to provide good speech quality in no-error conditions. The configuration allows also a coarse description at 1.2 kbps with reasonable quality to recover until three successive packet losses for larger bursts. The packetisation scheme is shown in Fig.7. A packet will contain both mentioned MELP coders and will be coded using 135 bits (54 +81) corresponding to a rate of 6 kbps. Hence, formation of a packet requires the presence of four successive frames of 22.5 ms each. In a transmission without packet loss, only the first 54 bits will be used to decode the signal. Then, a packet is attributed to the current frame corresponding to the MELP 2.4 (Fig.7). This causes a delay of 22.5 ms. Note that forming and sending the first packet request a delay of 90 ms. Afterwards, every 22.5 ms a packet is sent [13].

Paramètres	**MELP à 2.4 kbps**		**MELP à 1.2 kbps**				
Samp. frequency	8kHz		8kHz				
Frame size	180 samples (22.5 ms)		3*180 samples (67.5 ms)				
debit by Frame	44,44 frame/s		14.8148 frames/s				
Mode of voicing	V	N/V	VVV	UVV VUV	VVU	UUV UVU VUU	UUU
10 LSFs	25	25	43	43	39	43	27
Pitch	7	7	12	12	12	12	12
10 Fourier amplitudes	8	-	8	8	8	8	-
5 Bandes de voisement	4	-	6	4	4	2	-
2 Gains	8	8	10	10	10	10	10
Flag	1	-	1	1	1	1	-
Protection	-	13	-	2	6	4	31
Synchronisation	1	1	1	1	1	1	1
Total bits/Frame	**54 bits**		**81bits**				
Total	**54*44,44= 2400 bps**		**81*14.8148 = 1200 bps**				

Tab.1: Bit allocation encoder MELP 2.4 kbps and 1.2 kbps [13].

International Journal of Scientific & Engineering Research, Volume 6, Issue 4, April-2015
ISSN 2229-5518

V. Comparative Study Between Interleaving and MDC

We simulated different packet loss to introduce degradations in the synthetic signal. These losses were simulated randomly by using the RAND function which follows a uniform distribution law. The packet loss rate is given by the following formula:

$$\text{Taux} = \frac{\text{nombre de trames perdues}}{\text{nombre de total de trames}} \times 100 \qquad (5)$$

1) Evaluation of MELP coders

The evaluation of the performance of the two MELP coders implemented separately were designed using the Recommendation P.862 of the ITU-T (International Telecommunication Union) [14] called PESQ (Perceptual Evaluation of Speech Quality). This method describes an objective method for predicting the subjective quality for telephony and for voice coders. It is intended to evaluate the influence of factors such as packet loss, the variable delay and distortion due to channel errors that is poorly evaluated by conventional methods. The PESQ is designed to compare a reference version (original) to that obtained by synthesizing this reference or after transmission and have been adversely affected. The results are shown in Tab. 2.

	MELP 2.4 Kbps	MELP 1.2 Kbps
PESQ	2.99	2.71

Tab.2: Results of objective tests of two MELP coders.

2) Comparison between Interleaving and MDC

The results given in Tab.3 and illustrated on Fig.8 show the loss rate of the frames based on PESQ evaluation for the three cases, which are: original MELP, after application of MDC compared with the decorrelated convolutional interleaving judged as the best interleaving method [1, 9].

The masking method titled MDC with raising average of 0.60, present itself as the best method of masking compared to interleaving techniques, in fact with the best method known decorrelated convolution interleaving which gives a raising average of 0.50. The method of decorrelated convolution interleaving, applied to male and female combined corpus, gives a low degradation of our signal loss rates under 5%. The intelligibility of the signal is preserved according to PESQ scale (PESQ >2) for both methods for rates approaching 30%.

Loss rate %	Original MELP (PESQ)	MELP with MDC (PESQ)	MELP with Interleaving (PESQ)	Variation de PESQ	
				MDC	Interlea.
0	2.92	2.92	2.92	0	0
5	2.73	2.85	2.7	0.12	0.03
10	2.58	2.81	2.65	0.23	0.07
12	2.43	2.74	2.57	0.31	0.14
15	2.11	2.61	2.5	0.5	0.39
18	1.69	2.45	2.36	0.69	0.67
20	1.44	2.38	2.27	0.94	0.83
25	1.25	2.3	2.23	1.05	0.98
30	1.12	2.11	2.05	0.99	0.93
Average				**0.6**	**0.5**
Standard deviation				**0.366**	**0.396**

Tab.3 Evolution of PESQ obtained by MELP before and after application of technical MDC and interleaving for different loss rates for men and women speakers.

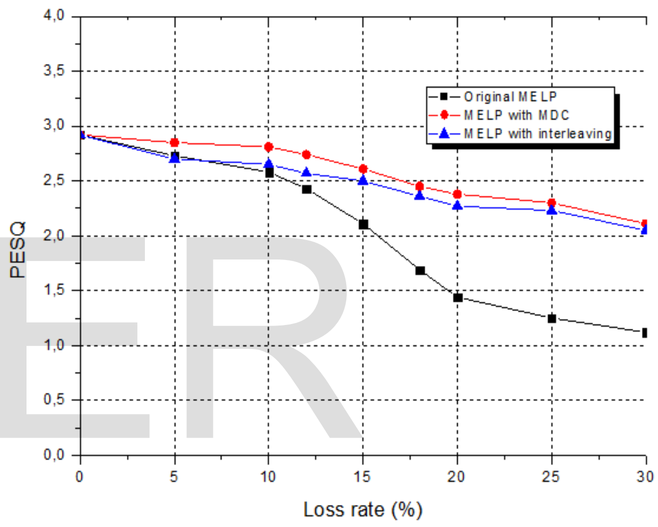

Fig.8 Evolution of PESQ obtained by MELP before and after application of technical MDC and interleaving for different loss rates for men and women speakers.

VI. Conclusion

Based on the performance Obtained using PESQ (Including average and standard deviation) for the MDC and decorrelated convolution interleaving. We can conclude that the performance of the MDC technique provides a significant improvement in perceptual quality compared to the decorrelated convolutif interleaving method essentially when the loss rate less than 12%, more than 12%, the two methods are almost equivalent to each other.

Références

[1] H .Kheddar, B .Boudraa," *Implementation of Interleaving Methods on MELP 2.4 Coder to Reduce Packet Loss in the Voice over IP (VoIP) Transmission", Int. Journal of Engineering Research and Applications* ISSN : 2248-9622, Vol. 5, Issue 3, (Part -4) March 2015, pp.85-8.

[2] S. Mohamed, "codage par description multiple en MELP pour la VoIP". Mémoire magister, laboratoire LCPT université USTHB, 2010.

International Journal of Scientific & Engineering Research, Volume 6, Issue 4, April-2015
ISSN 2229-5518

[3] G.Baudoin, J. Cernocký, P. Gournay, et G. Chollet, " *Codage de la parole à bas et très bas débit*". Département Signaux et Télécommunications, ESIEE BP 99 93162.
[4] J. Edward, Daniel and Keith A, "*Federal Standard 2.4 kbps MELP Over IP*",Roc. 43 rd IEEE Midwest Symp. on Circuits and Systems, Lansing MI, 0-7803-6475-9, pp.568-571, Aug 8-11,2000.
[5] J. Stachurski, A. McCree, V. Viswanathan, "*high quality MELP coding at bit-rates around 4 kb/s*''. 0-7803-5041-3/99, pp 485-488, IEEE 1999.
[6] C. Perkins, O. Hodson, et V. Hardman, "*A Survey of Packet Loss Recovery Techniques for Streaming Audio*". 0890-8044, pp.40-48, IEEE Network .September/October 1998.
[7] Teck-Kuen Chua and David C. Pheanis, "*QoS Evaluation of Sender-Based Loss-Recovery Techniques for VoIP*". 0890-8044, pp.14-22, IEEE Network. November/December 2006.
[8] B.P. Milner, A.B. James. "*Analysis and compensation of packet loss in distributed speech recognition using interleaving*", pp.1947- 1950, ISCA-SPEECH. September 1-4, 2003.
[9] B.P. Milner, A.B. James, "*A Comparison of Efficient Interleaver Designs for Real Time Distributed Speech Recognition"*, ASIDE, November 10-11. 2005.
[10] Gal Richard, E. Ravelli, Student et L. Daudet, Union of MDCT Bases for Audio Coding, *IEEE transactions on audio, speech, and language processing, vol. 16, pp.1361-1372. No. 8, November 2008.*
[11] McCree, K. Truong, E. B. George, T. P. Barnwell, V. Viswanathan. A 2.4 kbits/s MELP Coder Candidate for the New U.S. Federal Standard. In Proc. of IEEE-ICASSP, 1996.
[12] T. Wang, K. Koishida, V. Cuperman and A. Gersho. "A 1200 bps Speech Coder Based on MELP''. Proc. IEEE. Inter. Conf. Acoustics. Speech and Signal Processing, 2000.
[13] M.SAIDI, B.Boudraa, M.Bouzid, M.Boudraa," *Multiple Descriptions Coding in MELP Coder for Voice over IP"*,9th international multi-conference IEEE,2012.
[14] ITU-T, "Perceptual evaluation for speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs", Recommendation P.862, International Telecommunications Union, 2001.