



HAL
open science

Multiple description coding technique to improve the robustness of ACELP based coders AMR-WB

Hocine Chaouch, Fatiha Merazka, Philippe Marthon

► **To cite this version:**

Hocine Chaouch, Fatiha Merazka, Philippe Marthon. Multiple description coding technique to improve the robustness of ACELP based coders AMR-WB. *Speech Communication*, 2019, 108, pp.33-40. 10.1016/j.specom.2019.02.002 . hal-03498641

HAL Id: hal-03498641

<https://hal.science/hal-03498641v1>

Submitted on 21 Dec 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Open Archive Toulouse Archive Ouverte

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible

This is an author's version published in:

<http://oatao.univ-toulouse.fr/24830>

Official URL

DOI : <https://doi.org/10.1016/j.specom.2019.02.002>

To cite this version: Chaouch, Hocine and Merazka, Fatiha and Marthon, Philippe *Multiple description coding technique to improve the robustness of ACELP based coders AMR-WB*. (2019) *Speech Communication*, 108. 33-40. ISSN 0167-6393

Any correspondence concerning this service should be sent to the repository administrator: tech-oatao@listes-diff.inp-toulouse.fr

Multiple description coding technique to improve the robustness of ACELP based coders AMR-WB[☆]

Hocine Chaouch^{a,*}, Fatiha Merazka^a, Philippe Marthon^b

^a LISIC Laboratory, Telecommunications Department. USTHB University, P.O. Box 32 El Alia, Algiers, Algeria

^b ENSEEIHT Informatique, 2 Rue Camichel BP 7122, 31071 Toulouse Cedex 7, France

A B S T R A C T

Keywords:

VoIP
ITU-T G.722.2
Multiple description coding
Markov model
WB-PESQ

In this paper, a concealment method based on multiple-description coding (MDC) is presented, to improve speech quality deterioration caused by packet loss for algebraic code-excited linear prediction (ACELP) based coders. We apply to the ITU-T G.722.2 coder, a packet loss concealment (PLC) technique, which uses packetization schemes based on MDC. This latter is used with two new designed modes, which are modes 5 and 6 (18,25 and 19,85 kbps, respectively). We introduce our new second-order Markov chain model with four states in order to simulate network losses for different loss rates. The performance measures, with objective and subjective tests under various packet loss conditions, show a significant improvement of speech quality for ACELP based coders. The wideband perceptual evaluation of speech quality (WB-PESQ), enhanced modified bark spectral distortion (EMBSD), mean opinion score (MOS) tests and Multiple Stimuli with Hidden Reference and Anchor (MUSHRA) for speech extracted from TIMIT database confirm the efficiency of our proposed approach and show a considerable enhancement in speech quality compared to the embedded algorithm in the standard ITU-T G.722.2.

1. Introduction

Voice over Internet Protocol (VoIP) has gained a great popularity over the recent years, chiefly, due to its low cost and deployment easiness. However, the quality of service (QoS) has not yet reached a level equivalent to the one offered by traditional public switched telephone network (PSTN) (Goode, 2002).

VoIP uses packetized transmission of speech over the Internet (IP network) (Merazka, 2012) and therefore, at the receiver, some packets may be lost due to network delay, network congestion (jitter) or network errors. These lost packets deteriorate the speech quality and may cause conversation interruptions. Hence, it is necessary to employ a mechanism to recover the lost packets. Several packet loss recovery algorithms, also known as packet loss concealment (PLC) techniques, which can be either transmitter or receiver based, are used to replace these lost packets, (Perkins et al., 1998a; Kostas et al., 1998).

Algebraic code-excited linear prediction (ACELP) coders, such as ITU-T G.722.2, also known as adaptive multi-rate wideband (AMR-WB), are often used in VoIP systems because of their good speech quality in the absence of packet loss. However, their reliance on long-term prediction (LTP) causes propagation errors through speech frames making ACELP coders more sensitive to packet loss (ITU-T Rec., 2003a). This

latter fact will cause the quality of the reconstructed speech to degrade under packet loss conditions (Kim and Kleijn, 2004).

In literature, PLC techniques can be classified into repetition methods (Serizawa and Nozawa, 2002), interpolation/extrapolation methods (Perkins et al., 1998b) and more sophisticated regeneration methods based on a speech model (Sanneck, 1996).

In order to mitigate the effects of packet loss and transmission errors in VoIP, multiple description coding (MDC) is used. MDC divides data into distinct descriptors which depend on an acceptable decoding quality. In this case, the quality is increased by using more than two descriptors (Wah et al., 2000) and since the probability of losing all the descriptors is considered to be small, additional delay is not needed and more bandwidth is only necessary if the effective channel coding is required. While some practical MDC coders have been developed for image and video, relatively little attention has been given to MDC speech coding. Orozco et al. in Orozco et al. (2006) have made a comparison between code excited linear prediction (CELP) and sinusoidal coders where a packetization scheme based on MDC was applied to the sinusoidal coder is presented. The authors have applied their proposed MDC method in the lower bit rates. Also, the adaptive multi-rate (AMR) speech coding standard based on CELP speech was introduced by Yang et al. (2010). This strategy is based on error concealment which is applied to consecutive frame loss when transmission environment is not reliable and

[☆] Multiple description coding technique to improve the robustness of ACELP based coders AMR-WB

* Corresponding author.

E-mail addresses: chaouchhou@gmail.com (H. Chaouch), fmerazka@usthb.dz (F. Merazka), Philippe.Marthon@enseeiht.fr (P. Marthon).

the channel coding could not effectively control error occurrence. In [Zhipu et al. \(2005\)](#), authors have introduced multiple description source coding schemes to improve the statistical stability and performance of the estimation error covariance of Kalman filter with packet loss. Authors in [Li et al. \(2012a\)](#) have compared the performance of different MDC schemes for AMR-WB codec based on rate-distortion (R-D) theory while considering the parameter importance for high packet loss condition. Their proposed MDC scheme achieved substantial robustness in both low and high packet loss conditions. In [Li et al. \(2012b\)](#), the authors have proposed an analytical and an experimental comparison of forward error correction (FEC) and MDC performance for the AMR-WB codec. Considering the results of this comparison, the authors proposed an optimized approach to select the optimal packet loss recovery technique based on network conditions to achieve the best speech quality.

In this paper, we introduce and describe a new sender-based PLC method based on MDC into ACELP speech coder. In the previous works, the MDC approach has been used on narrow band coder at very low, low and medium bit rates ([Orozco et al., 2006](#)) and on AMR-WB at high rates for comparing MDC with FEC analytically and experimentally. Our MDC approach aims to improve the speech quality deterioration caused by packet loss for the AMR-WB coder at two designed high quality bit rates 18.25 kbps and 19.85 kbps over the technique embedded in the standard ITU-T G.722.2 ([Bessette, 2002](#); [Merazka and Fulvio, 2015](#)). Note that the suitable modes are selected according to the required transmission rate. In addition, a novel modeling packet loss as a Second-order Markov chain with four states is proposed and used instead of Gilbert model which is a First-order Markov chain with two states.

We compared the performance of the decoded speech obtained by our proposed PLC based MDC with the original G.722.2 codec using wideband perceptual evaluation of speech quality (WB-PESQ), enhanced modified bark spectral distortion (EMBSD), mean opinion score (MOS) and multiple stimuli with hidden reference and Anchor (MUSHRA) evaluation. The performance measures prove that our PLC approach based on MDC is better than the one embedded in the standard ITU-T G.722.2.

The reminder of this paper is structured as follows. In [Section 2](#), a brief overview of the AMR-WB G.722.2 is presented. In [Section 3](#), a novel packet loss model as a second-order Markov chain is introduced and described. A review on our concealment method based MDC is provided in [Section 4](#). In [Section 5](#), discussions on simulation results are given. Finally, a summary of the main contribution of this paper is presented in the [Section 6](#).

2. Overview of The AMR-WB G.722.2

The G.722.2 with AMR-WB algorithm is used as an internet wideband audio codec for VoIP with an audio band of 50-7000 Hz instead of the 200-3400 Hz band employed in classical telephony. When the bandwidth is increased the intelligibility and naturalness of speech are enhanced. The ITU-T G.722.2 AMR-WB codec is similar to 3GPP AMR-WB codec. The relative 3GPP proprieties are the TS 26.190 standards of the speech codec and the TS 26.194 The G.722.2 with AMR-WB algorithm is used as an internet wideband audio codec for VoIP with an audio band of 50-7000 Hz instead of the 200–3400 Hz band employed in classical telephony. When the bandwidth is increased the intelligibility and naturalness of speech are enhanced. The ITU-T G.722.2 AMR-WB codec is similar to 3GPP AMR-WB codec. The relative 3GPP proprieties are TS 26.190 standards of the speech codec and TS 26.194 for the Voice Activity Detector (VAD) ([3GPP T. S., 2001](#)) [3GPP T. S.](#)

G.722.2 depicts the precise mapping from input blocks of 320 speech samples in 16 bits uniform pulse code modulation (PCM) format to encoded blocks of 132, 177, 253, 285, 317, 365, 397, 461 and 477 bits and from encoded blocks of 132, 177, 253, 285, 317, 365, 397, 461 and 477 bits to output blocks of 320 reconstructed speech samples ([ITU-T Rec., 2003a](#)).

Table 1

G.722.2 - Bit allocation of the AMR-WB coding algorithm modes 5 and 6 for 20 ms frame.

18.25 kbps	VAD-flag					1
	ISP					46
	LTP-filtering	1	1	1	1	4
	Pitch delay	9	6	9	6	30
	Algebraic code	64	64	64	64	256
	Gain	7	7	7	7	28
Total						365
19.85 kbps	VAD-flag					1
	ISP					46
	LTP-filtering	1	1	1	1	4
	Pitch delay	9	6	9	6	30
	Algebraic code	64	64	64	64	288
	Gain	7	7	7	7	28
Total						397

The coding scheme for the multi-rate coding modes is also called algebraic code excited linear prediction coder, here in after referred to as ACELP. The multi-rate wideband ACELP coder is referred to as AMR-WB ([3GPP T. S., 2001](#)). The G.722.2 also uses an integrated voice activity detector (VAD) [3GPP T. S.](#) The G.722.2 also uses an integrated voice activity detection (VAD). The sampling rate is 16000 samples/s leading to a bit rate for the encoded bit stream of 6.60, 8.85, 12.65, 14.25, 15.85, 18.25, 19.85, 23.05 and 23.85 kbps which correspond, in practice, to modes 0, 1, 2, 3, 4, 5, 6, 7 and 8, respectively.

G.722.2 depicts the precise mapping from input blocks of 320 speech samples in 16 bits uniform pulse code modulation (PCM) format to encoded blocks of 132, 177, 253, 285, 317, 365, 397, 461 and 477 bits and from encoded blocks of 132, 177, 253, 285, 317, 365, 397, 461 and 477 bits to output blocks of 320 reconstructed speech samples ([ITU-T Rec., 2003a](#)).

In this paper, we are interested in modes 5 and 6 of the coder which correspond to bit rates 18.25 kbps and 19.85 kbps respectively. The parameters of these modes are given in [Table 1](#).

3. Modeling packet loss as a second-Order Markov chain

In this paper, we introduce a new packet loss model. Our scheme models lost packets as a second order Markov chain. Let X_t denotes the t -th packet outcome while $X_t = 1$ and $X_t = 0$ correspond to a packet loss and error-free transmission (successful), respectively. Our model can be seen as a second-order Markov chain. where each state of this model is represented by a couple (X_t, X_{t+1}) . We distinguish four probability transitions

$$p_{00} = P(X_{t+1} = 1 | X_t = 0, X_{t-1} = 0) \quad (1)$$

$$p_{01} = P(X_{t+1} = 1 | X_t = 0, X_{t-1} = 1) \quad (2)$$

$$p_{10} = P(X_{t+1} = 1 | X_t = 1, X_{t-1} = 0) \quad (3)$$

$$p_{11} = P(X_{t+1} = 1 | X_t = 1, X_{t-1} = 1) \quad (4)$$

The corresponding transition graph is shown in [Fig. 1](#). Our four state Markov chain model is irreducible and aperiodic, and thus, it is ergodic and convergent. The stationary probability distribution is given by

$$\pi_{00} = P((X_t, X_{t-1}) = (0, 0)) = \frac{1 - p_{01}}{p_{00}} \pi_{01} \quad (5)$$

$$\pi_{10} = \pi_{01} \quad (6)$$

$$\pi_{11} = \frac{p_{10}}{1 - p_{11}} \pi_{01} \quad (7)$$

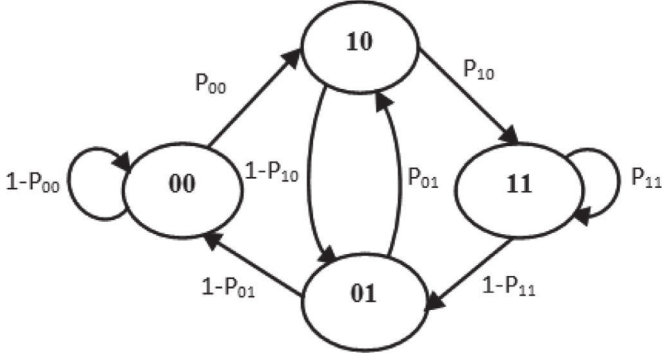


Fig. 1. Second-order Markov Model with four states (00: good, 10: breacking, 11: bad, 01: recovery).

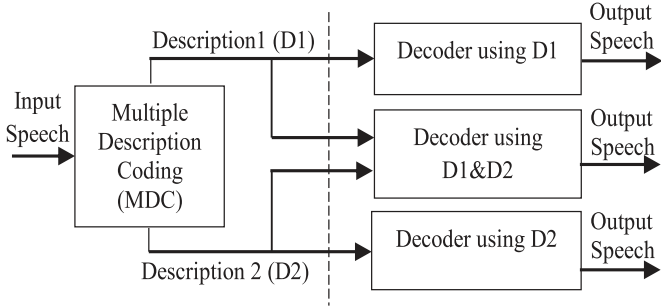


Fig. 2. Multiple descriptions coding basic scheme.

$$\pi_{01} = \frac{1}{a} \quad (8)$$

$$\text{with: } a = 2 + \frac{1-p_{01}}{p_{00}} + \frac{p_{10}}{1-p_{11}}$$

Note that our model is more general than the simple Gilbert model which is a First-order Markov chain with two states (Estrada et al., 2010; Bolot, 1993). However, when $p_{00} = p_{01}$ and $p_{10} = p_{11}$, our model is reduced to the simple two-state Gilbert model. Note that the advantage in our case is the fact that we can add a “breaking” state (1, 0) and “recovery state” (0, 1) unlike the Gilbert and Gilbert–Elliot models (Ellis et al., 2014; Rahl et al., 1986).

4. Concealment method based MDC

The MDC basic structure is illustrated in Fig. 2. The speech, after the encoder, is divided into two or several descriptions which are independently transmitted. Each description is separately decoded to decrease the quality reconstruction of the input speech. However, if two or more descriptors are available, they can be conjointly decoded for a higher-quality reconstruction of output speech (Lang et al., 2007; Choupani et al., 2012).

In this work, we have modified the source code of the original standard G.722.2 codec in order to obtain new bit rates which represent the descriptions for the MDC technique. To achieve this, we have used the database “DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus (TIMIT), training and test data” NIST (1990). The read speech constitutes the corpus of TIMIT database which was intended to deliver a speech data, in order to achieve acoustic phonetic knowledge and to develop, improve and evaluate an automatic speech recognition mechanism. The database was divided into two parts, one for training, and another for test. Given that our codec contains 10 codebooks of immittance spectral frequency (ISF) parameters, its realization required the use of a test core, which is composed of 29150 frames, each with a frame length of to 20 ms. Linde, Buzo and Gray (LBG) algorithm has been adopted in

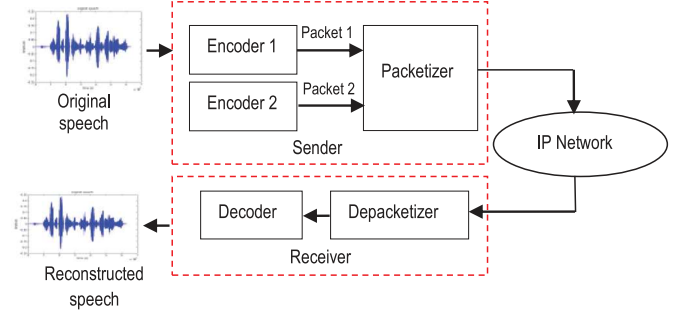


Fig. 3. Block diagram of VoIP transmission.

order to have the desired bit size of the codebook (Linde et al., 1980; Merazka, 2009). Note that, the LBG algorithm output data type is float which will be converted to integer and then used by our codec.

Recall that, the AMR-WB G.722.2 uses 6 parameters to represent speech. For the other parameters as algebraic coding, Pitch delay and Gain, the module of a mode has been replaced by a module of another different one. To be done, we have generated four new descriptions. For quantizing each parameter, our modified AMR-WB G.722.2 speech coder requires the bit allocation given in Table 2. Note that our purpose is to generate two new modes that will give the same mode as the original codec.

The quantization of ISF residual vector “ r ”, in our modified coder, is based on the use of split-multistage vector quantization (S-MSVQ) as shown in Table 3. The vector is divided into two sub-vectors “ $r_1(n)$ ” and “ $r_2(n)$ ” of dimensions 9 and 7, respectively. The quantization of these sub-vectors is performed in two stages

- The quantization of the bit rate 12.2 kbps (12,6 kbps and 7,25 kbps), for the two sub-vectors “ $r_1(n)$ ” and “ $r_2(n)$ ” in the first stage, is based on using “8” bits for each one. In the second stage, the quantization error vector is divided into “3” and “2” sub-vectors, respectively, i.e. the number of bits is (8,8,6,7,6,3,3) ((8,8,6,7,7,5,4) and (8,8,6,7,7,5,5) respectively) which gives “41” bits of ISF (rep “45” bits and “46” bits).
- The quantization of the bit rate 6.05 kbps for the sub-vectors “ $r_1(n)$ ” and “ $r_2(n)$ ” is based on the use of “8” bits and “6” bits, respectively. In the second stage, we divide the quantization error vectors into “2” and “1” sub-vectors, i.e. number of bits is (8,8,6,7,6,3,3) which gives “25” bits of ISF.

Thereby, we have generated two new modes (5 and 6) and for the sake of the synchronicity of the sender and the receiver, we have used two synchronous coders and added a packetizer module as shown in Fig. 3. We can select the suitable modes according to the required transmission rate.

Recall that, MDC with multiple descriptions can be used (Wang et al., 2005). In our work, two descriptions have been used. At the sender, the packetization is done by the original speech signal coding. In this case, we apply two bitrates, which represent two descriptors transmitted in the same packet. In the new designed mode 5 (18.25 kbps), for example, the first one uses the G.722.2 codec to encode the present frame at 12.2 kbps while the second one uses another G.722.2 to encode the following frame at 6.05 kbps (ITU-T Rec., 2003b). The two packets will then, be arranged as shown in Fig. 4.

At the receiver side, the depacketization is done. When the MDC with two descriptors is applied, the lost frame is substituted by the received one at the decoder.

Fig. 4 shows the packetization/depacketization scheme that represents an example of our proposed MDC approach, using two descriptors (12.2 kbps and 6.05 kbps), in order to construct a new mode 5 allowing of lost frames. In fact, when one or more successive packets are lost, the second descriptor allows the reconstruction of the synthesized (recon-

Table 2
Bit Allocation of the modified AMR-WB G.722.2.

New bit rates for Mode 5 design (18.25 kbps)		12.2 kbps					VAD-flag	1	
							ISP	45	
							LTP-filtering	1 1 1 1 4	
							Pitch delay	9 6 9 6 30	
							Algebraic code	64 64 64 64 144	
							Gain	7 7 7 7 28	
							Total	252	
New bit rates for Mode 6 design (19.85 kbps)		12.6 kbps					VAD-flag	1	
							ISP	41	
							LTP-filtering	1 1 1 1 4	
							Pitch delay	9 6 9 6 30	
							Algebraic code	36 36 36 36 144	
							Gain	7 7 7 7 24	
							Total	244	
			7.25 kbps					VAD-flag	1
								ISP	25
								Pitch delay	8 5 5 5 23
								Algebraic code	12 12 12 12 48
								Gain	6 6 6 6 24
								Total	121

Table 3
Quantization of ISP for 6.05, 7.25, 12.2 and 12.6 kbps.

12.2 kbit/s	1) unquantized 16-element-long ISP vector			2) Stage 1 (r_1) 8 bits			2) Stage 1 (r_2) 8 bits
	2) stage 1 (r_1) 8 bits		3) stage 2 ($r^{(2)}$ 1, 3-5) 7 bits	3) stage 2 ($r^{(2)}$ 1, 6-8) 6 bits	3) stage 2 ($r^{(2)}$ 2, 0-2) 6 bits	3) stage 2 ($r^{(2)}$ 2, 3-6) 3 bits	
6.05 kbps	1) unquantized 16-element-long ISP vector			2) stage 1 (r_1) 8 bits			
	2) stage 1 (r_1) 8 bits		3) stage 2 ($r^{(2)}$ 1, 0-4) 5 bits	3) stage 2 ($r^{(2)}$ 1, 5-8) 3 bits	3) stage 2 ($r^{(2)}$ 1, 0-6) 3 bits		
12.6 kbps	1) unquantized 16-element-long ISP vector			2) stage 1 (r_1) 8 bits			2) stage 1 (r_2) 8 bits
	2) stage 1 (r_1) 8 bits		3) stage 2 ($r^{(2)}$ 1, 0-2) 6 bits	3) stage 2 ($r^{(2)}$ 1, 3-5) 7 bits	3) stage 2 ($r^{(2)}$ 1, 6-8) 7 bits	3) stage 2 ($r^{(2)}$ 2, 0-2) 5 bits	3) stage 2 ($r^{(2)}$ 2, 3-6) 4 bits
7.25 kbps	1) unquantized 16-element-long ISP vector			2) stage 1 (r_1) 8 bits			2) stage 1 (r_2) 8 bits
	2) stage 1 (r_1) 8 bits		3) stage 2 ($r^{(2)}$ 1, 0-2) 6 bits	3) stage 2 ($r^{(2)}$ 1, 3-5) 7 bits	3) stage 2 ($r^{(2)}$ 1, 6-8) 7 bits	3) stage 2 ($r^{(2)}$ 2, 0-2) 5 bits	3) stage 2 ($r^{(2)}$ 2, 3-6) 5 bits

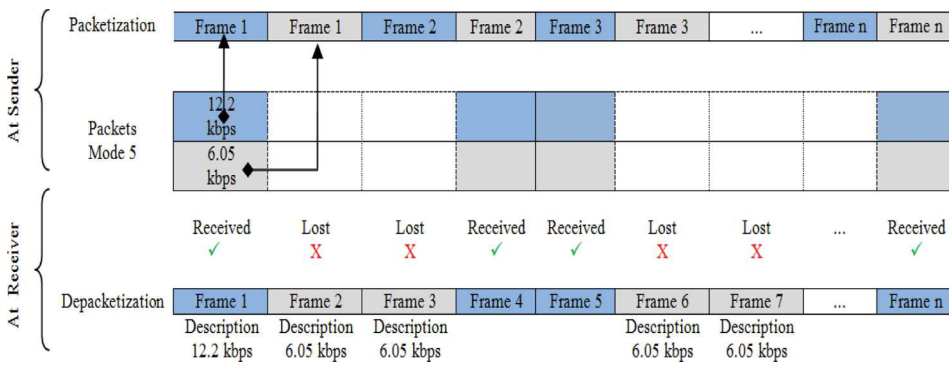


Fig. 4. Packetization/depacketization process based on two descriptions of MDC : mode 5.

structed) speech. The depacketization aims to replace the lost frames with the good ones of the second descriptor.

5. Simulations and discussions

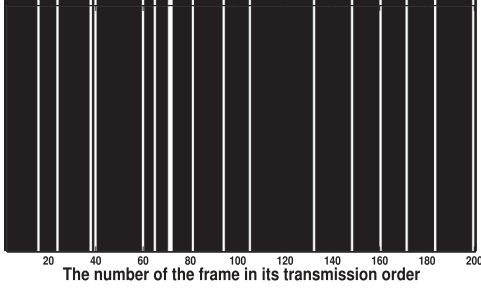
In our experiments, a wave file with 198 frames is used to test our proposed Second-order Markov model resilience against network loss, for a variety loss rates as given in Table 4.

Fig. 5 plots simulated frames for different loss rates corresponding to 11.11%, 18.68%, 32.32% and 42.42%, where the number of lost frames is represented in the figure with white vertical segments, while received frames are in black vertical segments. In Fig. 5, the x-axis represents the number of the frame in its transmission order. If the frame number k is lost, we place at the abscissa $x = k$, a white vertical segment, else, if the frame number k is well received, we place (at the abscissa $x = k$) a black vertical segment. The more white vertical segments there are, the more frames are lost, however, loss rate is important.

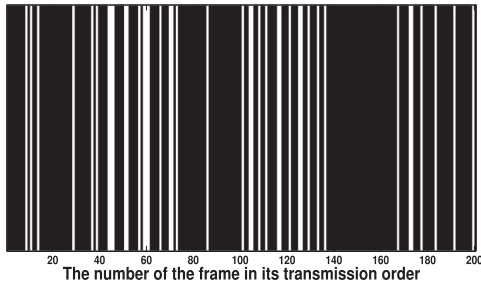
Table 4
Simulated loss rates with our proposed second order Markov model.

Rate (%)	Lost frames	P_{00}	P_{01}	P_{10}	P_{11}
00	00	00	00	00	00
11.11	22	0.10	0.15	0.10	0.15
18.68	37	0.20	0.30	0.20	0.30
32.32	64	0.30	0.30	0.30	0.40
42.42	84	0.30	0.50	0.30	0.50

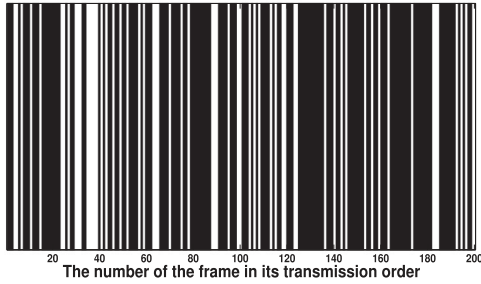
(a): with loss rate equal to 11.11%



(b): with loss rate equal to 18.68%



(c): with loss rate equal to 32.32%



(d): with loss rate equal to 42.42%

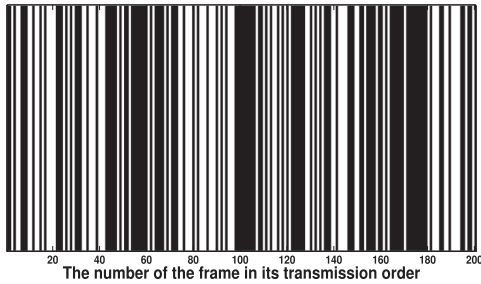


Fig. 5. Simulated frames for different loss rates : (a) 11.11%, (b) 18.68%, (c) 32.32% and (d) 42.42%.

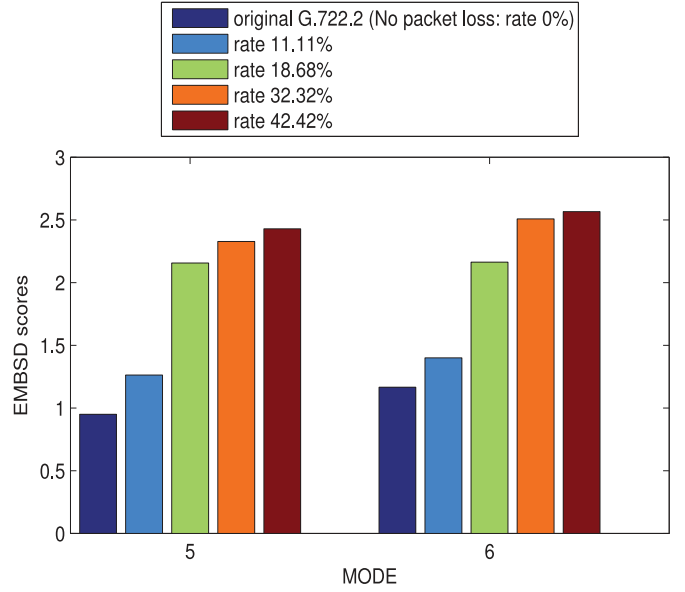


Fig. 6. EMBSD values of the AMR-WB without and with packet loss for modes 5 and 6.

For example, in Fig. 5(b) corresponding to loss rate equal to 18.68%, the frame number 60 is lost, while the frame number 180 is well received.

Our aim is to quantify the perceptual voice quality by employing objective and subjective quality estimation algorithms. For the objective ones, the WB-PESQ (ITU-T, 2005) and the EMBSD (Yang, 1999) have been used whereas for the subjective ones, we have used MOS (ITU-T Rec., 2006) and confirmed our results with MUSHRA tests (Recommendation ITU-R, 2003). Recall that

- The EMBSD values yield 0 for a comparison with the same file.
- The PESQ scores which lie in the range of -0.5 up to 4.5 yields a score of 4.5 for a comparison with the same file.
- The MOS scales lie in the range of 1 up to 5. The MOS scale is defined in ITU-T Rec. (1996) as 5 = Excellent, 4 = Good, 3 = Fair, 2 = Poor, 1 = Bad, and yields a score of 4.54 for a comparison with the same file.
- The MUSHRA is 0-100 scale for a comparison with the same file.

First, we evaluate the performance of the AMR-WB without and with packet loss for modes 5 and 6. We use for packet loss, the loss rates obtained with our proposed second order Markov model given in Table 4. The EMBSD, WB-PESQ and MOS measures are shown in Figs. 6–8 respectively. We can confirm from these figures, that the speech quality is generally good, for both modes 5 and 6, in the original codec and bad with packet loss.

A Second evaluation is conducted under our proposed MDC based concealment method. Firstly, we provide the performance evaluation of the four new bitrates which are given in Table 5.

From this table, we can say that the quality of the new descriptions is satisfactory as it varies between fair and good. Secondly, the use of the MDC yields the performance shown in Figs. 9–12. The Results are almost similar to the original speech, which indicates that the presented method has a good performance

Tests in Fig. 9, with EMBSD for different loss rates, show that the speech quality with MDC for the two modes remains always higher than the original ITU-T G.722.2 without MDC.

In Fig. 10, the WB-PESQ measurement confirms the efficiency of our concealment method based MCD. It shows that, for the same mode, the more the loss rate increases the more the quality decreases and deteriorates in the original coder. On the other hand, the use of MDC improves the quality compared to the original codec. So, the intelligibility of the

Table 5

Tests results with EMBSD, WB-PESQ, MOS and MUSHRA for the proposed concealment method based MDC.

	New bit rates for mode 5 design (18.25 kbps)		New bit rates for mode 6 design (19.85 kbps)	
	6.05 kbps	12.2 kbps	7.25 kbit/s	12.6 kbps
EMBSD	1.315	0.904	1.327	0.869
WB-PESQ	3.297	3.711	3.354	3.779
MOS	3.264	3.830	3.347	3.913
MUSHRA	60	84	61	89

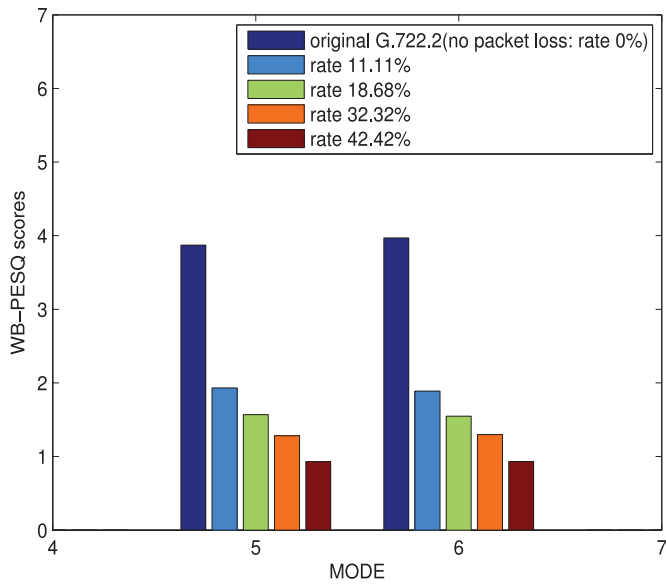


Fig. 7. WB-PESQ scores of the AMR-WB without and with packet loss for modes 5 and 6.

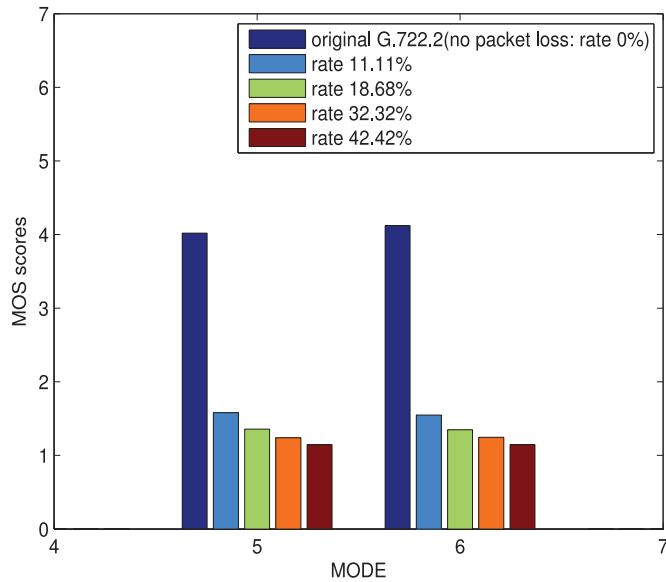


Fig. 8. MOS scores of the AMR-WB without and with packet loss for modes 5 and 6.

signal is preserved according to MOS scales depicted in Fig. 11, with scores varying between fair and good like the original codec without packet loss. Hence, the proposed MDC based concealment method is better than the embedded one in the standard ITU-T G.722.2 coder. It is

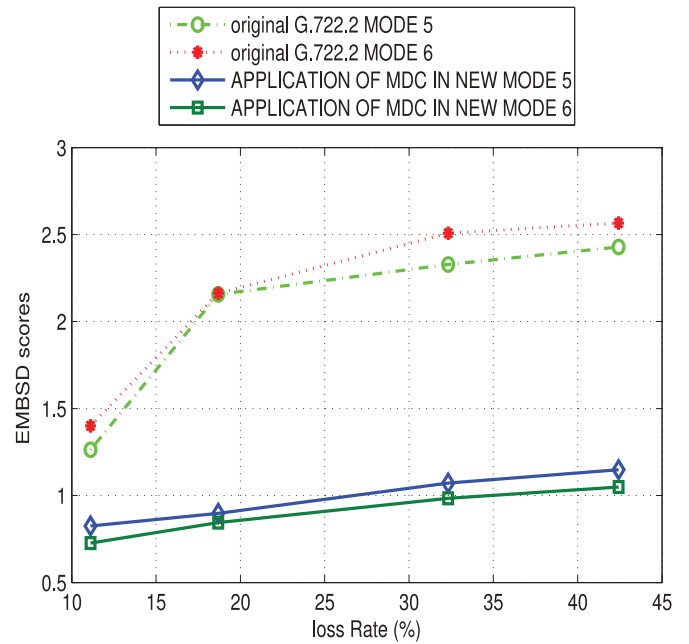


Fig. 9. EMBSD values for different loss rates comparing the original G.722.2 (modes 5 and 6) with our proposed concealment method based MCD (new modes 5 and 6).

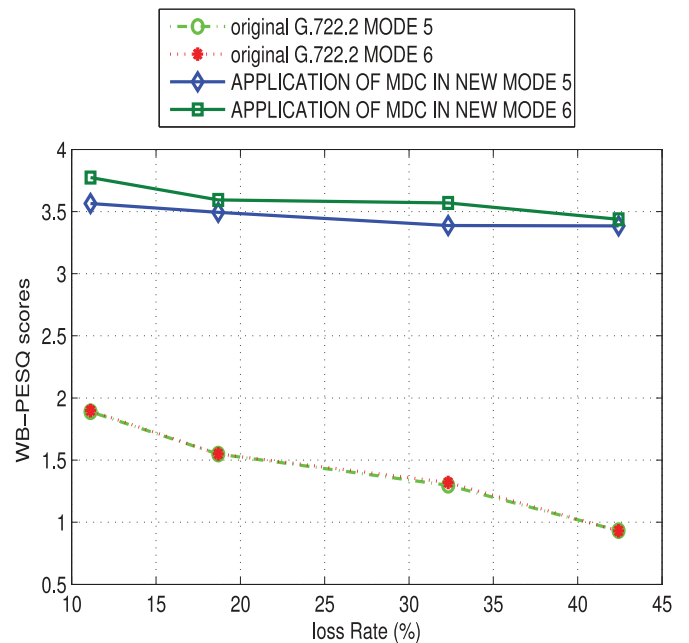


Fig. 10. WB-PESQ scores for different loss rates comparing the original G.722.2 (modes 5 and 6) with our proposed concealment method based MCD (new modes 5 and 6).

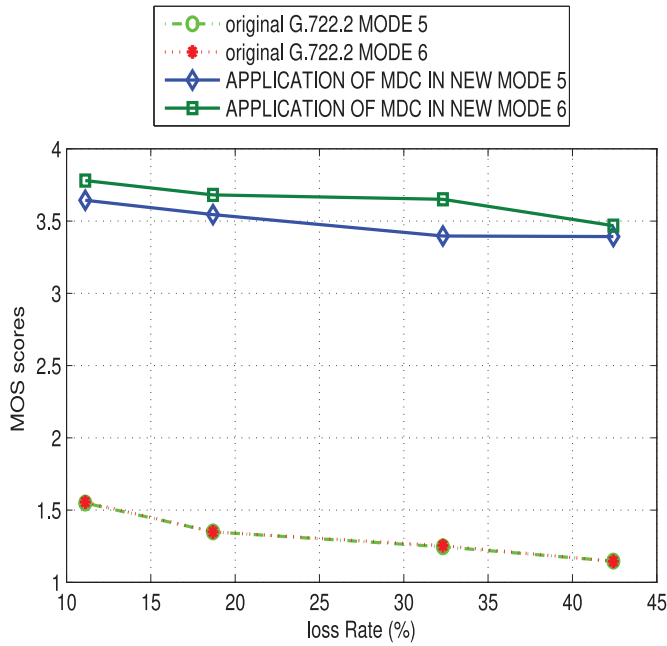


Fig. 11. MOS scores for different loss rates comparing the original G.722.2 (modes 5 and 6) with our proposed concealment method based MCD (new modes 5 and 6).

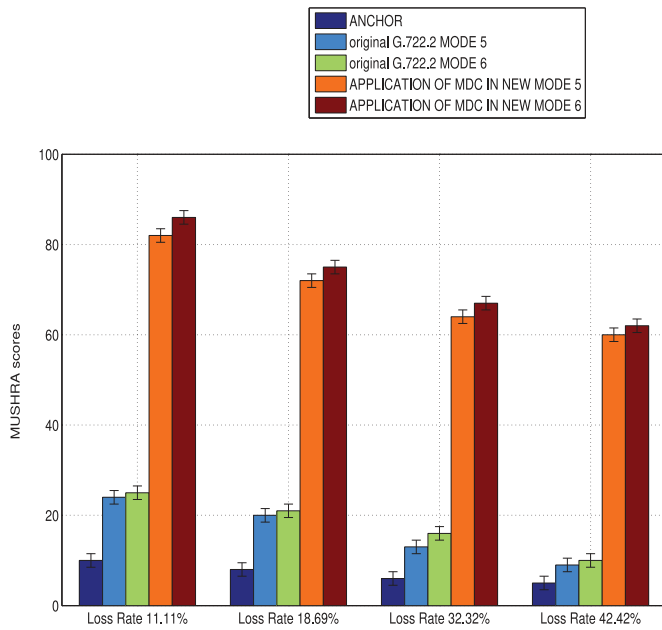


Fig. 12. MUSHRA scores for different loss rates comparing the original G.722.2 (modes 5 and 6) with our proposed concealment method based MCD (new modes 5 and 6).

a new approach, which provides an excellent speech quality and a high accuracy over lossy networks.

Subjective tests were also carried out to evaluate the performance of our proposed MDC scheme. The subjective test method used in the experiments is MUSHRA methodology. This test has the advantage of requiring less participants than subjective MOS test in order to obtain statistically significant results (20 listeners are enough) (Recommendation ITU-R, 2003). Listeners must compare the standard PLC algorithm embedded in ITU-T G. 722.2 and our proposed MDC method by listening and comparing them with the unprocessed sig-

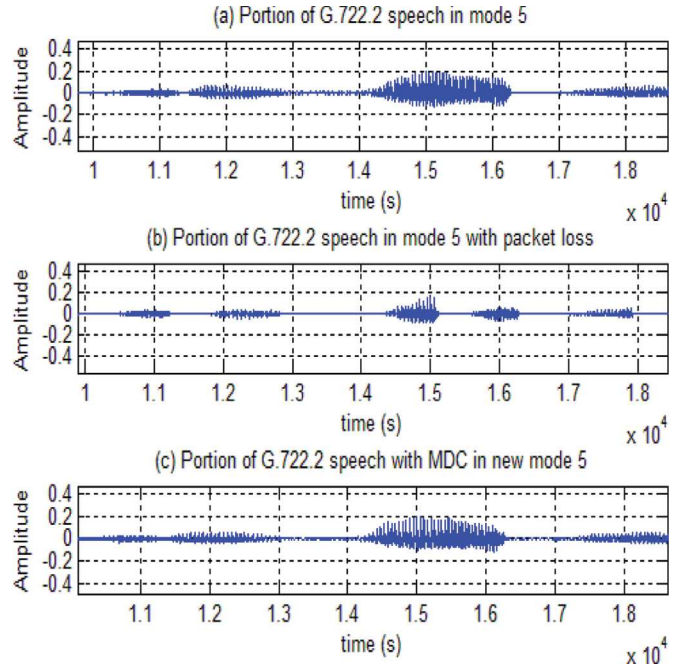


Fig. 13. Audigram portion mode 5 for loss rate 42.42%.

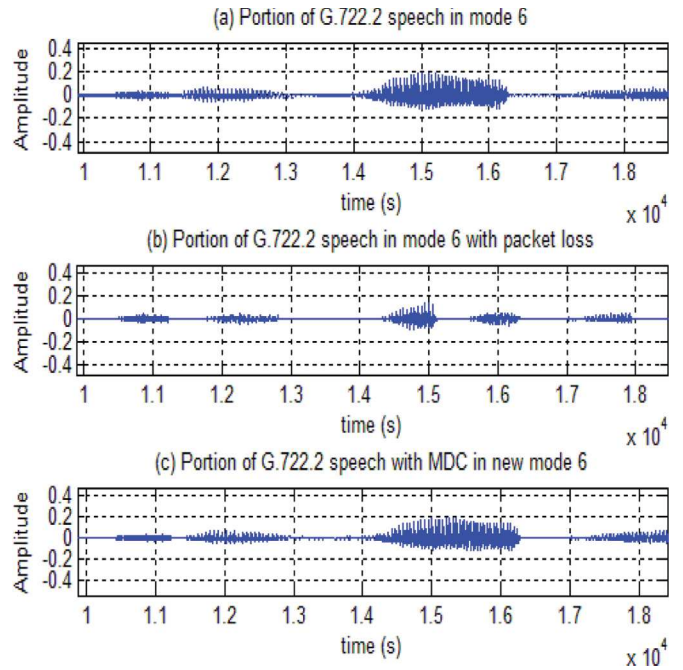


Fig. 14. Audigram portion mode 6 for loss rate 42.42%.

nal (reference) and the degraded signal with loss rate 50% (anchor). The confidence intervals have been set to 95%. The test set-up consisted of 14 sentences evaluated by 22 listeners. In our experiments, listeners gave scores according to quality of decoded speech by original and improved MDC algorithm. The test sentences were presented to listeners at a randomized order and repeated for four different loss rates 11.11%, 18.68%, 32.32% and 42.42%. The performance evaluations are presented in Fig. 12.

It can be seen, from Figs. 9–12, that the MDC based concealment method outperforms the embedded method in the original ITU-T G.722.2 coder in low and high loss rates. Clearly, the designed MDC

technique highly improves the intelligibility and naturalness of speech signal.

Figs. 13 and 14 show audiograms portions of speech, which prove and confirm the efficiency and the robustness of our approach.

6. Conclusion

In this paper, we have presented an MDC technique that properly ensures a good speech quality for any packet loss rate. We have employed our proposed Second order Markov model with four states (00: good, 10: bracking, 11: bad, 01: recovery) to simulate network loss for different loss rates. The main purpose was to realize global bit rates of 18.25 kbps and 19.85 kbps corresponding to modes 5 and 6 respectively. Based on WB-PESQ measurements, MOS scores, EMBSD tests and MUSHRA scores and under a variety of frame erasure conditions, we can clearly see that our proposed method significantly improves the speech quality compared to the embedded algorithm in the standard ITU-T G.722.2 coder. While the experiments have been performed on the G.722.2 speech codec modes 5 and 6 the proposed scheme is clearly applicable to other modes and could be extended to other CELP based speech coders as well.

As for future work, we have two main tracks. Firstly, we intend to apply our proposed method to the rest of the other modes (bit rates) while making more comparisons with other packet loss concealment methods. Secondly, we aim for comparing the performance of our proposed method to recent approaches such as hidden Markov model (HMM).

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.specom.2019.02.002](https://doi.org/10.1016/j.specom.2019.02.002).

References

- 3GPP T. S., AMR wideband speech codec. In: Voice Activity Detector VAD.
- 3GPP T. S., 2001. AMR wideband speech codec. Transcoding functions.
- Bessette, B., et al., 2002. The adaptive multi-rate wideband speech codec (AMR-WB). *Trans. Speech Audio Process.* 10 (8), 620–636.
- Bolot, J.C., 1993. End-to-end frame delay and loss behavior in the internet. In: *ACM SIGCOMM, France*, pp. 289–298.
- Choupani, R., Stephan, W., Mehmet, T., 2012. Unbalanced multiple description wavelet coding for scalable video transmission. *J. Electron. Imaging* 21 (4), 043006-1.
- Ellis, M., Pazaros, D.P., Kypraios, T., Perkins, C., 2014. A two-level markov model for packet loss in UDP/IP-based real-time video applications targeting residential users. *Elsevier Comput. Networks. J.* 70, 384–399.
- Estrada, L., Torres, D., Toral, H., 2010. Characterization and modeling of packet loss of a voIP communication. *Int. J. Electron. Commun. Eng.* 4 (6), 970–974.
- Goode, B., 2002. Voice over internet protocol (voIP). *IEEE Internat. Conf.* 90 (9), 1495–1517.
- ITU-T Rec., 2005. Wideband extension to recommendation p.862 for the assessment of wideband telephone networks and speech codecs. International Telecommunication Union, Geneva, Switzerland.
- ITU-T Rec., G., 2003. Wideband coding of speech at around 16 kbit/s using adaptive multi-rate wideband (AMR-WB). International Telecommunication Union, Geneva, Switzerland.
- ITU-T Rec., G., 2003b. Wideband coding of speech at around 16 kbit/s using adaptive multi-rate wideband (AMR-WB).
- ITU-T Rec., P., 1996. Methods for subjective determination of transmission quality.
- ITU-T Rec., P., 2006. Mean opinion score (MOS) terminology.
- Kim, M.Y., Kleijn, W.B., 2004. Comparison of transmitter-based packet-loss recovery techniques for voice transmission. In: *Eighth International Conference on Spoken Language Processing ICSLP, Jeju Island, Korea*, pp. 641–644.
- Kostas, T.J., Borella, M.S., Sidhu, I., Schuster, G.M., Grabiec, J., Mahler, J., 1998. Real-time voice over packet-switched networks. In: *IEEE Netw.* 12 (1), 18–27.
- Lang, Y., Shenghui, Z., Jingming, K., 2007. A multiple description speech coder based on AMR-WB in converged IP network. 5th International Congress on Image and Signal Processing, Chongqing, pp. 1699–1703.
- Li, Z., Xie, Y., Qi, J., Gao, L., 2012. A novel multiple description coding scheme based on AMR-WB in converged IP network. *5th International Congress on Image and Signal Processing, Chongqing*, pp. 1699–1703.
- Li, Z., Zhao, S., Bruhn, S., Wang, J., Kuang, J., 2012. Comparison and optimization of packet loss recovery methods based on AMR-WB for voIP. *Speech Commun.* 54 (8), 957–974.
- Linde, Y., Buzo, A., Gray, R.M., 1980. An algorithm for vector quantizer design. *IEEE Trans. Commun. COM.* 28, 84–95.
- Merazka, F., 2009. Enhanced differential split vector quantization of line spectrum pairs for CELP-type coders in packet networks. In: *World Congress on Engineering and Computer Science, San Francisco, USA* 1–4.
- Merazka, F., 2012. Intraframe quantization of speech line spectrum pairs for code-excited linear prediction based coders in packet networks. *Trans. Emerging Telecommun. Technol. J.* 23 (8), 789–804.
- Merazka, F., Fulvio, B., 2015. Dynamic forward error correction algorithm over IP network services for ITU-t g. 722.2 codec. In: *IEEE 10th International Conference, Internet Technology and Secured Transactions ICITST, London, UK* 369–372.
- NIST, 1990. Timit speech corpus.
- Orozco, E., Stephane, V., Ahmet, M.K., 2006. Multiple description coding for voice over IP using sinusoidal speech coding. In: *IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP, Toulouse, France*, pp. I-9–I-12.
- Perkins, C., Hodson, O., Hardman, V., 1998. A survey of packet loss recovery techniques for streaming audio. In: *IEEE Network.* 12 (1), 40–48.
- Perkins, C., Hodson, O., Hardman, V., 1998. A survey of packet loss recovery techniques for streaming audio. *IEEE Network.* 12 (5), 40–48.
- Rahl, L.R., Brown, P.F., Souza, P.V., Mercer, R.L., 1986. Maximum mutual information estimation of hidden Markov model parameters for speech recognition. In: *International Conference on Acoustic, Speech and Signal Processing* 49–52.
- Recommendation ITU-R, B., 2003. Method for subjective assessment of intermediate quality level of coding systems.
- Sanneck, H., et al., 1996. A new technique for audio packet loss concealment. *Global Telecommunications Conference, GLOBECOM'96. 'Communications: The Key to Global Prosperity IEEE* 1996.
- Serizawa, M., Nozawa, Y., 2002. A packet loss concealment method using pitch waveform repetition and internal state update on the decoded speech for the sub-band ADPCM wideband speech codec. In: *In Speech Coding, 2002, IEEE Workshop Proceedings. IEEE*, pp. 68–70.
- Wah, B.W., Xiao, S., Lin, D., 2000. A survey of error-concealment schemes for real-time audio and video transmissions over the internet. In: *IEEE International Symposium. Multimedia Software Engineering. Taipei, Taiwan* 17–24.
- Wang, Y., Amy, R.R., Shunan, L., 2005. Multiple description coding for video delivery. In: *IEEE proceeding* 93 (1), 57–70.
- Yang, J., Yu, S.S., Zhou, J., Gao, Y., 2010. A new error concealment method for consecutive frame loss based on CELP speech. *Comput. Electr. Eng.* 36 (5), 1014–1020.
- Yang, W., 1999. Enhanced modified bark spectral distortion (EMBSD): An objective speech quality measurement based on audible distortion and cognition model. PhD Dissertation, Temple University, USA.
- Zhipu, J., Vijay, G., Babak, H., Richard, M.M., 2005. State estimation utilizing multiple description coding over lossy networks. In: *IEEE of the 44th Conference on Decision and Control, and the European Control, Seville* 12–15.