Commun. Fac. Sci. Univ. Ank. Series A2-A3
V. 46. pp. 47-61 (2001)

CHANNEL AND SOURCE CONSIDERATIONS OF A MELP DERIVED VOCODER OPERATING AT REDUCED BIT RATES

H.G. İLK, S. TUĞAÇ

Department of Electronics Eng., Faculty of Eng., Ankara University, Ankara, TURKEY

(Received Sept. 26, 2001; Accepted Dec. 29, 2001)

ABSTRACT

Robust low bit rate speech coders are essential in commercial and military communication systems. They operate at fix bit rates and those bit rates can not be altered without major modifications in the vocoder design. In this paper we introduce a Scaled Speech Coder, which operates on time-scale modified input speech. The proposed method offers any bit rate from 2,400 bits/s to downwards without modifying the principle vocoder structure, which is the Mixed Excitation Linear Prediction (MELP) vocoder. We consider the application of transmitting MELP-encoded speech over noisy communication channels after time scale compression is applied. Computer simulation results, both source and channel, are presented in terms of objective speech quality metrics and informal subjective listening tests. A statistical tool called bootstrap is also used to determine the accuracy of these test results. Design parameters such as codec complexity and delay are also investigated.

KEYWORDS

Speech Coding, Channel Characterisation and Simulation, Low Bit Rate Vocoders, Time Scale Modification of Speech

1. INTRODUCTION

In recent years, low bit rate (<4.0 kb/s) speech coding techniques have been the focal point of considerable research activity as part of the quest for minimum bit rate, high quality speech. The applications of these efficient speech coding techniques are numerous, ranging from voice mail to mobile communications and multimedia communication systems. Recently several approaches have been attempted for encoding speech at lower bit rates. These coders can be divided into three major groups.

*1. **Prototype Interpolation Coders:*** Prototype Waveform Interpolation (PWI) [1] and its variants such as Characteristic Waveform Representation (CW) [2] and Time-Frequency Interpolation (TFI) [3]

2. Harmonic Coders: Sinusoidal Transform Coders (STC) [4] and Multi-band Excitation Vocoders (MBE) [5]
3. LP Based Vocoders: Split-band Linear Predictive Coder [6] and Mixed Excitation Linear Prediction (MELP) [7] vocoder, the new U.S. Federal Standard operating at 2.4 kb/s.

The MELP vocoder is based on the traditional linear predictive parametric model, but also includes mixed excitation. MELP vocoder has five main features [8].

i) A reliable pitch estimate: The MELP vocoder is based on representing short-term voiced speech as the summations of sinusoids. Therefore the new federal standard heavily relies on a robust pitch estimate. The pitch estimation procedure involves integer, fractional and final pitch calculation, pitch doubling check and average pitch update algorithms. In addition, all extracted speech parameters are interpolated, during synthesis, pitch synchronously in the decoder.

ii) Parameter Interpolation: The extracted speech parameters via encoder, i.e. Fourier magnitudes and pitch of the excitation signal and gain, line spectral frequencies (LSFs), jitter (Aperiodic Flag) and filter coefficients of the shaping filter are interpolated in the decoder to ensure smooth evolution in the characteristics of the synthesised speech.

iii) Mixed Excitation: The harmonic and noise speech components are synthesised separately. In order to separate the harmonic and noise components, five frequency bands are defined and each band is declared as Voiced/Unvoiced. The synthesised pulse and noise excitation are then filtered and summed to form the mixed excitation

iv) Bit Packing and Error Protection: Table 1 shows the bit allocation for the MELP vocoder, for both Voiced and Unvoiced modes. Bits, representing the extracted speech parameters are packed in the encoder. The transmission order for the MELP vocoder will be further described in the following sections. To improve performance in channel errors, the unused coder parameters for the unvoiced mode only are replaced with forward error correction (FEC). Three Hamming (7,4) codes and one Hamming (8,4) code are used. FEC is implemented when the Fourier magnitudes, bandpass voicing, and jitter information need not be transmitted. FEC replaces those 13 bits with the parity bits of the Hamming codes. These codes protect the first stage LSF index (7 bits) of the multi-stage vector quantizer (MSVQ) and both gain indices (5 bits for the first and 3 bits for the second gain).

Table 1. Bit allocation table for the MELP vocoder [8].

Parameters	Voiced	Unvoiced
LSF's	25	25
Fourier Magnitudes	8	-
Gain (2 per frame)	8	8
Pitch, overall voicing	7	7
Bandpass Voicing	4	-
Aperiodic Flag	1	-
Error Protection	-	13
Sync Bit	1	1
Total Bits/22.5 ms Frame	54	54

v) Bit Unpacking and Error Correction: The received bits are unpacked from the channel and assembled into codewords. If any erasure is detected in the current frame, by the Hamming code, by the pitch code (which contains the mode information - voiced, unvoiced or frame erasure), or directly signalled from the channel, then a frame repeat mechanism is implemented. That is, all of the parameters for the current frame are replaced with the parameters from the previous frame. There is no error correction for the voiced mode, except the special all-zero code for the first gain parameter is received. In this case, some errors in the second gain parameter can be detected and corrected which provides improved performance in channel errors.

In many applications it is desirable to transform a speech waveform into a signal which is more useful than the original. For example in time-scale modification; speech can be sped up in order to compress the words spoken into an allocated time interval or to quickly scan a passage.

There are numerous methods in both time and frequency domains, for the modification of speech waveforms. One frequency domain approach is based on the sinusoidal representation that explicitly estimates the amplitude and phase of the vocal cord excitation and vocal tract system function contributions to each sine wave [9]. This method is called Sinusoidal Analysis/Synthesis method, SASM. Another frequency domain approach manipulates an excitation by deconvolving the original speech with a vocal tract spectral envelope estimate [10]. Time expansion is achieved by doubling the unwrapped phase of the spectrum. This approach is called Speech Transformation Without Pitch Extraction, STWPE.

One important time domain modification algorithm is waveform similarity overlap-and-add (WSOLA) method, which ensures sufficient signal continuity that exists in the speech signal [11]. WSOLA algorithm performs better than other overlap-and-add algorithms, such as time domain pitch-synchronised overlap-and-add, TD-PSOLA [12], because it does not require a pitch estimate and ensures maximal similarity at the segment joints by using waveform similarity measures. In addition qualities such as naturalness, intelligibility and speaker dependent features (pitch and formant structure) are well preserved by the WSOLA algorithm.

In this paper a novel approach to vocoders, in order to reduce the bit rate required to transmit speech signal, is proposed. While traditional low bit rate vocoders code original input speech, the proposed procedure codes time-scale modified signal. This method is particularly useful when existing low bit rate speech coding algorithms are used as principal vocoders because time-scale modifications (compression and expansion) are performed as a prior and post-process at the transmitter and receiver respectively. The main contribution of this paper to speech coding applications may be viewed in two different aspects:

i) The proposed procedure offers a flexible bit rate switching method to reduce the bit rate of the principal vocoder, leading to a "desired" operating bit rate at the expense of increased complexity and delay and graceful degradation in speech quality.

ii) The proposed procedure may be used in situations where channel capacity limits the user for error correction. The spared bits can be used for error correction [13], leading to a robust speech coder against channel errors. The bit rate of the vocoder, employed by the proposed procedure, in this case remains unchanged.

2. SYSTEM DESCRIPTION

Traditional LP based vocoders use either a periodic pulse train or white noise as the excitation for an all pole synthesis model. These vocoders either operate at a fix bit rate or the bit rate varies depending on the channel or source conditions. In the former case, operating bit rate can not be altered without major modifications in the vocoder design.

As described in the "Introduction" section, MELP vocoder employs linear predictive coding, pitch estimation and Fourier magnitude modelling for the excitation in order to code speech signals efficiently. WSOLA algorithm, on the other hand, offers a powerful tool for removing redundancy from speech in time domain. Speech signals exhibit both short-term (i.e. from sample to sample) and long-term (i.e. pitch related) correlation. WSOLA algorithm, when used with modest compression factors, removes some of the long-term correlation from speech by using waveform similarity measures. The time compressed speech signal still exhibits strong short-term correlation and some pitch correlation and therefore LP coding techniques can be successfully applied after time scale modification. We therefore propose a MELP derived vocoder called Scaled Speech Coder (SSC) which takes advantage of both time and frequency domain analysis techniques. The motivation here is to merge two approaches, i.e. time-scale modification and low bit rate speech coding, in order to compress and code speech signals effectively. In addition, the proposed SSC procedure offers a novel way to change the operating bit rate of the principle vocoder without any modification in the design. Please also note that pitch and formant structures of the speech signal are not modified during time-scale modification [11].

In this method transmitter speeds up original input speech Signal A, by a factor of β using WSOLA [11]. This signal, Signal B in Figure 1, is then coded using the new US Federal Standard Mixed Excitation Linear Prediction (MELP) algorithm at 2.4 kb/s [7].

At the receiver side, coded signal is first decoded to obtain time-scale modified Signal C which is then slowed by a factor of 1/β in order to reconstruct the synthesised speech Signal D. Advantage stems from the fact that time-scale modified signal requires less coding time than that of the original input speech at the expense of increased complexity and delay. Other question of interest is the degraded voice quality. Figure 1 illustrates the overall block diagram of the Scaled Speech Coder.

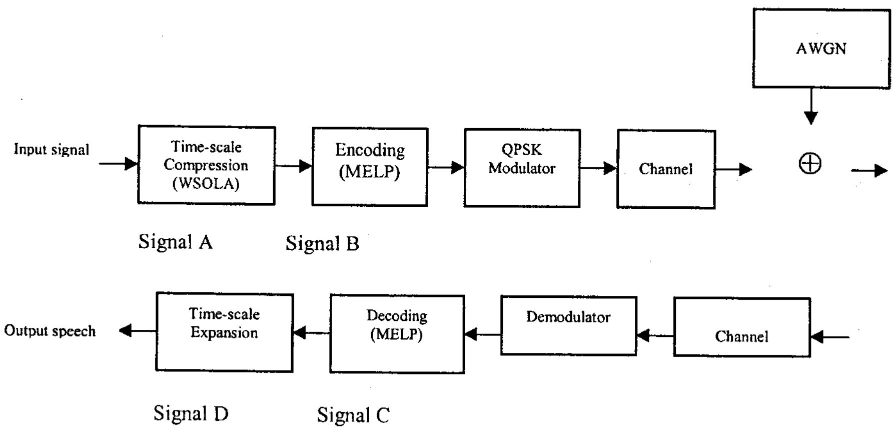

Figure 1. The overall block diagram of the Scaled Speech Coder

It is possible to obtain variants of Scaled Speech Coder by employing other time scale modification algorithms proposed in References [9,10,12] and/or low bit rate speech coders proposed in References [1,5,6]. WSOLA [11] and MELP [7] algorithms have been selected for their well-known performances.

3. SOURCE CODING SIMULATIONS

In order to assess the performance of the proposed procedure, different compression factors (β) of 0.5, 0.7 and 0.9 have been computer simulated, under benign input conditions, for source coding. Transmission channel characterisation and simulation are not considered initially, in order to isolate the distortion introduced by the channel conditions. The operating bit rates are 1200, 1680 and 2160 bits/s respectively, where the operating bit rate of the principle MELP vocoder is 2400 bits/s. Phonetically balanced sentences obtained from TIMIT database were used for test purposes. All speech files were converted to 8 kHz. Figure 2.a) illustrates the spectrogram of original speech utterance "Cash-mere", while Figures 2.b), c), and d) depict Signals B, C and D in Figure 1 in the frequency domain respectively, for β=0.5. Please note that time scale is different for Figures 2.b) and 2.c) than the original. Figure 3 on the other hand illustrates the synthesised speech spectrogram, obtained from MELP vocoder without any compression.

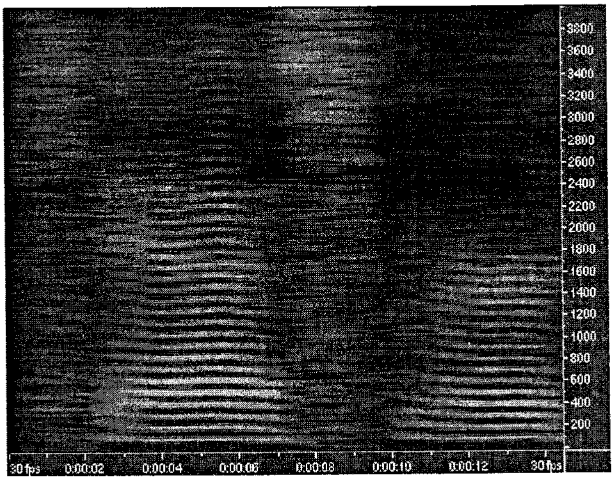

Figure 2.a) The spectrogram of original speech utterance "Cash-mere", Signal A in Figure 1

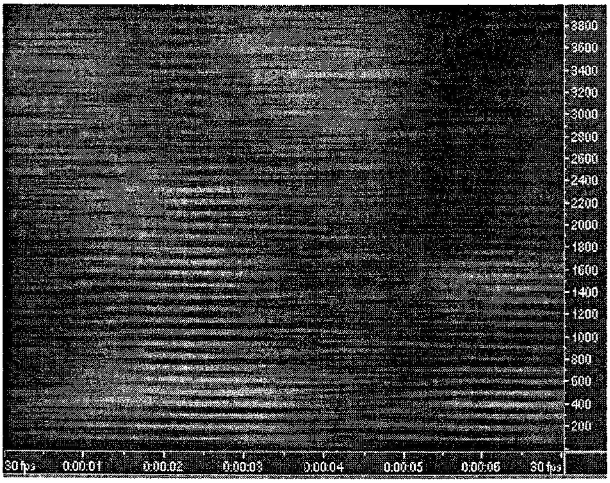

Figure 2.b) The spectrogram of time compressed speech utterance "Cash-mere", Signal B in Figure

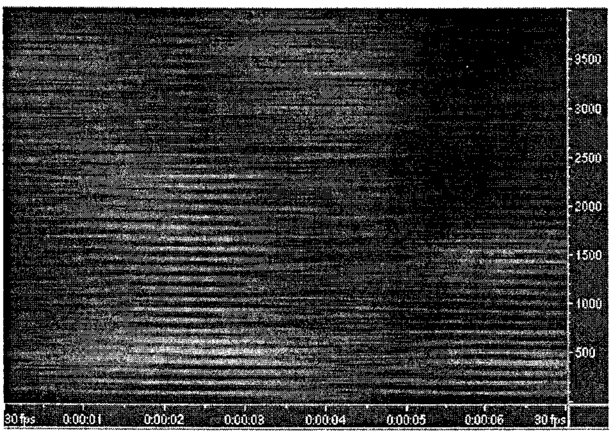

Figure 2.c) The spectrogram of time compressed and MELP decoded speech utterance "Cash-mere", Signal C in Figure 1.

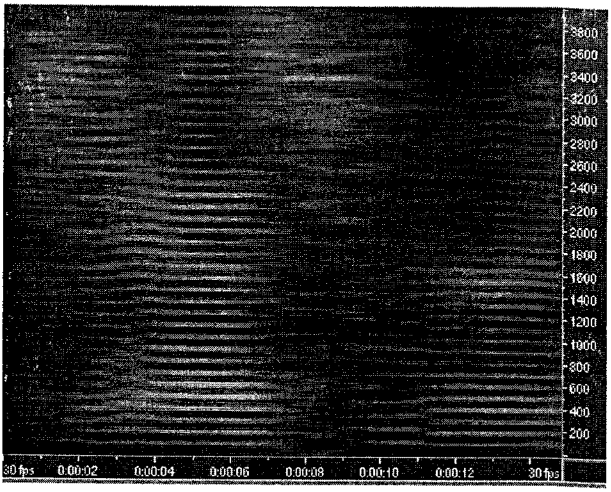

Figure 2.d) The spectrogram of time expanded, MELP decoded speech utterance "Cash-mere", Signal D in Figure 1.

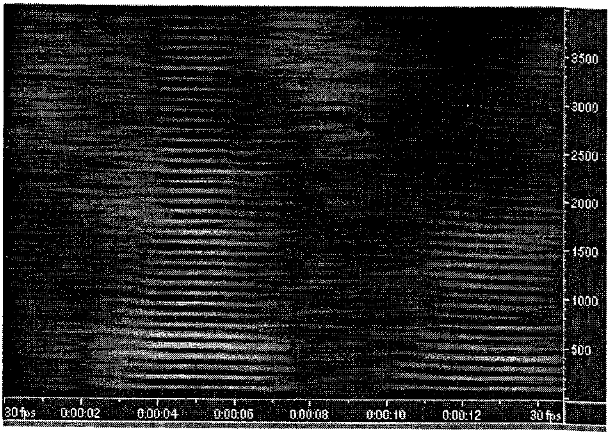

Figure 3) The spectrogram of Federal Standard MELP decoded speech utterance "Cash-mere"

Simulation results mainly concentrate on intelligibility and voice quality versus bit rate and delay. 25 subjects, of whom 8 of them are trained, were accessed for this study. Voice quality is assessed by using Mean Opinion (MOS) and Degradation Mean Opinion Score (DMOS) tests. During MOS tests, listeners were asked to rate the output speech quality according to absolute scale, ranging from "very bad" for grade 1 to "excellent" for grade 5. The main obstacle, during MOS tests, was that; ordinary subjects were not familiar with low bit rate vocoders and they were confused between harsh, muffled, buzzy and nasal quality of speech and noise added after coding. To overcome this limitation, DMOS test was also conducted. In this test, listeners were asked to rate the quality of time scaled and coded sentences relative to the quality of standard MELP vocoder output. Please note that this study is not concerned with the performance measures of the MELP vocoder. A detailed comparison of the MELP vocoder with other standard coders is addressed in Reference [14].

In general, 25 subjects are not sufficient to assess the quality of the decoded speech. Thus a statistical tool called bootstrap [15], in determining the accuracy of the MOS and DMOS test results of an unknown population, is used. The bootstrap is a general method for determining the accuracy of estimators (sample mean/median/correlation coefficient and so on) of an unknown parameter (population mean/median /correlation coefficient respectively) when the underlying distribution is unknown. 95% BCa (bias-corrected and accelerated) confidence intervals have been constructed for MOS and DMOS test results obtained from 25 listeners. BCa method is an improved version of the standard intervals, which attempts to improve a normal transformation for the non-normality of the estimator and to correct the bias resulting from the transformation. In this way, MOS and DMOS test results are presented not only in mean and standard deviation figures but they are also presented within a confidence interval.

When β=0.9, compressed and coded male speech scores somewhere between (3.42, 4.40) with 95% probability for MOS, while the MELP coded male speech scores somewhere between (3.74, 4.54). DMOS test results also indicate that there is no significant difference between these two decoded speech files, as tabulated in Table 2. On the other hand abrupt changes, especially for the lower bound of mean opinion and degradation mean opinion scores, are observed when β=0.5. These results indicate that compression factor β should lie somewhere in the range of 1.0 to 0.5 and preferably between 0.7 to 0.5 for the best compromise between coding efficiency and voice quality.

Table 2. Statistically assessed Mean Opinion and Degradation Mean Opinion Scores for each compression factor β.

Compression Ratio (β)	**95% BCa (Bias Corrected and Accelerated) Confidence Intervals**			
	MOS (Male)	**DMOS (Male)**	**MOS (Female)**	**DMOS (Female)**
0.5	[2.12 , 2.83]	[2.51 ,3.60]	[2.01 , 2.78]	[2.88 , 3.89]
0.7	[2.89 , 3.76]	[4.23, 4.66]	[2.79, 3.74]	[4.10 , 4.62]
0.9	[3.42 , 4.40]	[4.62, 4.95]	[3.10 , 4.14]	[4.47 , 4.91]
1.0 (MELP)	[3.74 , 4.54]	5	[3.45, 4.41]	5

4. CHANNEL CHARACTERISATION AND SIMULATION

The proposed SSC procedure provided high communication quality speech under error-free channel conditions, for compression ratios above 0.5. However, channels impaired by large amounts of Gaussian noise could be experienced in practical applications and therefore performance of the proposed method, under adverse channel conditions needs to be determined.

All of the simulation results, described in this paper, were obtained by using additive white Gaussian noise (AWGN) channel and quadrature phase shift keying (QPSK) was used for modulation purposes. The transmission order for the MELP vocoder is given in Table 3 [8]. The sync bit alternates between 0 and 1 from frame to frame.

Table 3. The transmission order for the 54 bits in each MELP frame [8].

Bit	Voiced	Unvoiced	Bit	Voiced	Unvoiced	Bit	Voiced	Unvoiced
1	G(2)-1	G(2)-1	19	**LSF(1)-7**	LSF(1)-7	37	G(1)-1	G(1)-1
2	BP-1	FEC(1)-1	20	LSF(4)-6	LSF(4)-6	38	BP-3	FEC(1)-3
3	P-1	P-1	21	**P-4**	P-4	39	BP-2	FEC(1)-2
4	LSF(2)-1	LSF(2)-1	22	**LSF(1)-6**	LSF(1)-6	40	LSF(2)-2	LSF(2)-2
5	LSF(3)-1	LSF(3)-1	23	**LSF(1)-5**	LSF(1)-5	41	LSF(3)-4	LSF(3)-4
6	**G(2)-4**	G(2)-4	24	**LSF(2)-6**	LSF(2)-6	42	**LSF(2)-3**	LSF(2)-3
7	**G(2)-5**	G(2)-5	25	**BP-4**	FEC(1)-4	43	LSF(3)-3	LSF(3)-3
8	LSF(3)-6	LSF(3)-6	26	**LSF(1)-4**	LSF(1)-4	44	LSF(3)-2	LSF(3)-2
9	**G(2)-2**	G(2)-2	27	**LSF(1)-3**	LSF(1)-3	45	LSF(4)-4	LSF(4)-4
10	**G(2)-3**	G(2)-3	28	**LSF(2)-5**	LSF(2)-5	46	LSF(4)-3	LSF(4)-3
11	**P-5**	P-5	29	**LSF(4)-5**	LSF(4)-5	47	**AF**	FEC(4)-3
12	LSF(3)-5	LSF(3)-5	30	FM-1	FEC(4)-1	48	LSF(4)-2	LSF(4)-2
13	**P-6**	P-6	31	**LSF(1)-2**	LSF(1)-2	49	FM-5	FEC(3)-3
14	**P-2**	P-2	32	**LSF(2)-4**	LSF(2)-4	50	FM-4	FEC(3)-2
15	**P-3**	P-3	33	FM-8	FEC(2)-3	51	FM-3	FEC(3)-1
16	LSF(4)-1	LSF(4)-1	34	FM-7	FEC(2)-2	52	FM-2	FEC(4)-2
17	**P-7**	P-7	35	FM-6	FEC(2)-1	53	G(1)-3	G(1)-3
18	**LSF(1)-1**	LSF(1)-1	36	G(1)-2	G(1)-2	54	**SYNC**	SYNC

NOTES : G = Gain
P = Pitch/Voicing
FEC = Forward Error Correction Parity Bits
Bit 1 = least significant bit of data set
Highlighted Bits = 24 Most Significant MELP Bits
BP = Bandpass Voicing
LSF = Line Spectral Frequencies
FM = Fourier Magnitudes
AF = Aperiodic Flag

Bit stream obtained from different compression factors (β) of 0.5, 0.7 and 0.9 have been computer simulated under different channel signal to noise ratios (SNR) to investigate the bit error ratio (BER) performance.

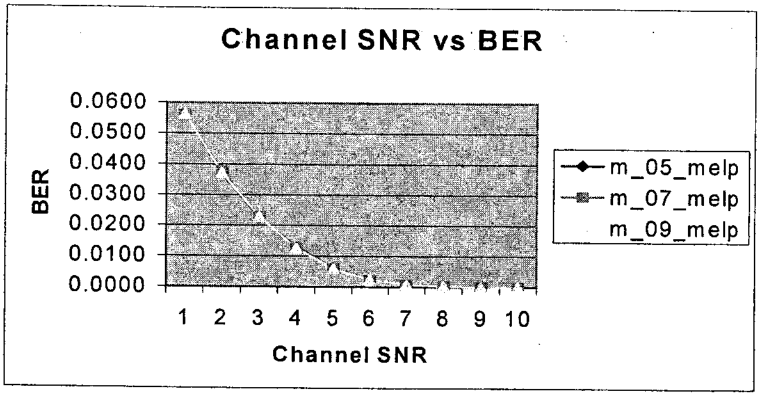

Figure 4. Channel SNR versus BER for different compression ratios

Table 4. The BER performances under different channel SNR values.

SNR (dB)	BER
1	0,0562
2	0,0371
3	0,0227
4	0,0128
5	0,0060
6	0,0024
7	7,82E-04
8	1,78E-04
9	2,99E-05
10	3,21E-06

The BER performances were found to be similar for different β values, as would be expected heuristically. Table 4 gives the BER performances, for β=0.5, under different channel SNR values. Please note that there is no error correction for the standard MELP vocoder, except for unvoiced frames, frame erasure and the special all-zero code for the first gain parameter, as described in detail in "Bit Unpacking and Error Correction" under "Introduction" section.

SNR between source coded speech files and files obtained after channel simulation are also measured according to Equation 1. These results, given in Table 5, indicate distortion added during channel transmission in objective quality metrics. These results are important because they verify that there is no distortion due to channel transmission when the channel SNR is equal or above 10 dB. The output speech quality rapidly degrades for channel SNRs below 5 dB.

$$SNR(s_{channel}(n), s_{source}(n)) = -10\log\frac{\sum_n (s_{source}(n) - s_{channel}(n))^2}{\sum_n s_{source}(n)^2} \quad (1)$$

Table 5. SNR measured between channel transmitted and source coded speech signals.

SNR(dB)	male_05	male_07	Male_09	fem_05	fem_07	fem_09
1	-4,9872	-3,6906	-3,7179	-3,3523	-3,8411	-3,9635
5	-3,0992	-2,8948	-2,2103	-1,5129	-3,2398	-1,6718
10	∞	∞	∞	∞	∞	∞

Finally, informal subjective listening tests are conducted in order to determine the effects of AWGN channel. Table 6 shows the subjective results with the following key; U = Unacceptable, P = Poor, F = Fair and G = Good. Subjective listening test results compare time-scale modified, coded and channel transmitted speech signal with respect to the original input speech file.

Table 6. Informal subjective listening test results

SNR(Db)	male_05	male_07	Male_09	fem_05	fem_07	fem_09
1	U	U	U	U	U	U
5	P	F	G	P	F	F
10	F	G	G	F	F	G

5. DESIGN ISSUES

A. Real Time Implementation

It is possible to real-time implement the proposed Scaled Speech Coder. MELP algorithm has been real-time implemented on both fixed and floating point DSPs [16] while Reference [17] describes the details of real-time implementation of WSOLA algorithm on floating point TMS 320-C31, 60 MHz, 45 MIPS processor. This processor was supported by 128 K RAM and 32 K ROM memory and two DSPs have been used, one for time compression and another for time expansion. This implementation has been tested for β values between 0.45 to 1.0.

B. Codec Delay

MELP vocoder operates on 22.5 ms frames and requires one additional frame for buffering, namely the overall delay is 45.5 ms. Please note this additional frame is not required due to codec complexity but it is required by the nature of the MELP algorithm. WSOLA algorithm delay is within the frame and therefore the overall delay for the proposed SSC algorithm is determined multiplying 45.5 ms by the expansion factor ($1/\beta$). Table 7 gives the overall delay with minimal buffering for each compression factor β.

Table 7. The overall delay with minimal buffering for each compression factor β

Compression Ratio (β)	Bit Rate (Bits/s)	Codec Delay (ms)
0.5	1,200	91
0.7	1,680	64.5
0.9	2,160	50.5
1.0 (MELP)	2,400	45.5

C. Full-Duplex Operation

WSOLA algorithm has been initially employed to convert a half-duplex system into a virtual full-duplex link in CRC, Communication Research Center, Ottawa, Canada [17]. In the case of Scaled Speech Coder full-duplex communication can be achieved when there is no additional delay introduced by the complexity requirements. If the Digital Signal Processor (or DSPs) employed for the real time implementation of the SSC are not sufficient and/or codes are not optimized for a real time implementation, i.e. time compression-encoding and decoding-time expansion operations can not be performed simultaneously, the "overall" delay would be the delay with minimal buffering listed in Table 7, plus the delay introduced by the complexity. In this case, if the "overall" delay is not acceptable, the Scaled Speech Coder can not be used in a full duplex communication.

6. CONCLUSIONS

A novel approach to speech coding was introduced in this paper for the compression of speech signals used in vocoders. While conventional methods code original input speech, the proposed algorithm codes time-scale modified signal. WSOLA algorithm was used for time-scale modification while the new Federal US Standard MELP vocoder was employed for coding purposes.

The simulation of transmitting MELP-encoded speech over noisy communication channels after time scale compression is also considered. The main contribution of this paper to speech coding applications may therefore be viewed in two different aspects:

i) The proposed procedure offers a flexible bit rate switching method to reduce the bit rate of the principal vocoder, leading to a "desired" operating bit rate at the expense of increased complexity and delay and graceful degradation in speech quality.

ii) The proposed procedure may be used in situations where channel capacity limits the user for error correction. The spared bits can then be used for error correction [13] leading to a robust speech coder against channel errors. The bit rate of the principal vocoder, employed by the procedure, in this case remains unchanged.

The proposed Scaled Speech Coder (SSC) procedure produces comparable communication quality speech at half the bit rate of the standard MELP vocoder for random bit errors under 1% during transmission. Synthesised speech quality degrades gracefully for decreasing values of the compression ratio β and it is similar to that of standard MELP vocoder. Best compromise between coding efficiency and voice quality is observed when compression ratio is set somewhere between 0.7-0.5. In addition, SSC procedure does not require modifications on the principle vocoder architecture because pitch and formant structure of the input speech signal are not modified after time compression is applied and therefore time compressed speech signal is directly applicable to the principle vocoder. WSOLA algorithm is cascaded to MELP vocoder in order to compress and expand speech signals in the encoder and decoder respectively.

Acknowledgements : The authors wish to thank Dr. Nur Serinken of the Communications Research Centre (CRC) Ottawa, Canada for his support. The authors are also grateful to the members of AUSPG, Ankara University Speech Processing Group, for evaluating decoded speech files and making constructive comments. Synthesised speech files can be found at the URL address, http://auspg.science.ankara.edu.tr.

REFERENCES

[1] W.B. Kleijn, "*Continuous Representations in Linear Predictive Coding*", Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing, pp. 201-204, 1991.

[2] W.B. Kleijn and J. Haagen, "*A Speech Coder Based on Decomposition of Characteristic Waveforms*" Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing, pp. 508-511, 1995.

[3] Y. Shoham, "*High Quality Speech Coding at 2.4 to 4.0 KBPS Based on Time-Frequency Interpolation*", Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing, pp. 167-170, 1993.

[4] R.J. McAulay and T.F. Quatieri, *Low Rate Speech Coding Based on the Sinusoidal Model.* Advances in Speech Signal Processing edited by S. Furui and M.M. Sondhi, pp. 165-207. Marcel Dekker, New York, 1992.

[5] D.W. Griffin and J.S. Lim, "*Multiband Excitation Vocoder*" IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. ASSP-36, pp. 1223-1235,1988.

[6] A. Kondoz, "*A 2.4-1.2 kb/s SB-LPC Speech Coder and Noise Pre-processor: The Turkish NATO Candidate*". IEEE National Signal Processing and Applications

Conference, Antalya - Turkey, pp. 539-544, 2000.
[7] L.M. Supplee, R.P. Cohn, J.S. Collura and A.V. McCree, "*MELP: The New Federal Standard at 2400 BPS*" Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing, pp. 1591-1594, 1997.
[8] Federal Information Processing Standards Publication (FIPS Draft Publication), "Analog to Digital Conversion of Voice by 2,400 Bit/second Mixed Excitation Linear Prediction (MELP)", pp. 1-57, 1998.
[9] T.F. Quatieri and R.J. McAulay, "*Speech Transformation Based on a Sinusoidal Representation*", Technical Report, Lincoln Laboratory, MIT, Lexington, Massachusetts, pp.1-56, 1986.
[10] S. Seneff, "*System to Independently Modify Excitation and/or Spectrum of Speech Waveform Without Explicit Pitch Extraction*", IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. ASSP-30, No.4, pp. 566-578, 1982.
[11] W. Verhelst and M. Roelands, "*An Overlap-add Technique Based on Waveform Similarity (WSOLA) For High Quality Time-Scale Modification of Speech*", Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing, pp. 554-557, 1993.
[12] E. Moulines and F. Chanpentier, "*Pitch Synchronous Waveform Processing Techniques for text-to-speech Synthesis Using Diphones*", Speech Communication, Vol.9, pp.453-467, 1990.
[13] D.J. Rahikka, J.S. Collura, T.E. Fuja and T. Fazel, "US Federal Standard MELP Vocoder Tactical Performance Enhancement via MAP Error Correction", MILCOM, Vol.2, pp. 1458-1462, 1999.
[14] M.A Kohler, "*A Comparison of the New 2400 bps MELP Federal Standard with Other Standard Coders*" Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing, pp. 1541-1544, 1997.
[15] B. Efron and R. Tibshirani, "*An Introduction to the Bootstrap*", Chapman and Hall Publications, London, pp. 153-165, 1993.
[16] E. Ertan, E. Aksu, H.G. Ilk, H. Karci, O. Karpat, T. Kolcak, L. Sendur, M. Demirekler and E. Cetin, "*Implementation of an Enhanced Fixed Point Variable Bit Rate MELP Vocoder on TMS320C549*", Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing, pp. 2295-2298, 1999.
[17] N. Serinken, B. Gagnon and O. Erogul, "*Full Duplex Speech for HF Radio Systems*" IEE HF Radio Systems and Techniques Conference Publications, pp. 281-284, 1997.