# MELP: THE NEW FEDERAL STANDARD AT 2400 BPS

Lynn M. Supplee, Ronald P. Cohn, John S. Collura
U.S. Department of Defense, 9800 Savage Road, Ft. Meade, Maryland 20755 USA

Alan V. McCree
Corporate Research & Development, Texas Instruments, Dallas, Texas 75265 USA

## ABSTRACT

This paper describes the new U. S. Federal Standard at 2400 bps. The Mixed Excitation Linear Prediction (MELP) coder was chosen by the DoD Digital Voice Processing Consortium to replace the existing 2400 bps Federal Standard FS 1015 (LPC-10). This new standard provides equal or improved performance over the 4800 bps Federal Standard FS 1016 (CELP) at a rate equivalent to LPC-10. The MELP coder is based on the traditional LPC model, but includes additional features to improve its performance.

## 1. INTRODUCTION

In 1993, the DoD Digital Voice Processing Consortium (DDVPC) began the process to select a new U. S. Federal Standard at 2400 bps [1]. After compiling all of the user requirements, the DDVPC developed a set of minimum requirements [2] and designed a test plan [3, 4] to select the new standard based on these requirements. Formal testing began in September 1995 and a selection was made in March 1996. The Mixed Excitation Linear Prediction (MELP) coder, which was developed by Texas Instruments and Atlanta Signal Processors, Inc. [5], and was based on work performed in a Ph.D. thesis at Georgia Tech [6], was chosen by the DDVPC to replace the current 2400 bps Federal Standard FS 1015 (LPC-10) [7]. Having selected MELP and decided that the new standard would be bitstream defined, the DDVPC then initiated the formal process to define the new Federal Standard. This process began with the writing of the standards document and is estimated to take approximately fifteen months to complete. The new standard is expected to be completed in the fall of 1997. This paper will describe the proposed MELP Federal Standard.

## 2. SCOPE OF THE NEW STANDARD

The new 2400 bps standard defines a set of requirements for the conversion from analog voice to a 2400 bps digitized form for digital transmission by a method know as MELP. This standard has been defined to facilitate interoperability between communication facilities and systems for the Federal Government while providing increased performance over LPC-10. The new standard is expected to be used by all Federal departments and agencies, both civilian and military, in the design and procurement of all communications equipment employing 2400 bps digitized voice. In addition to the requirements, the new 2400 bps standard will include an algorithm description of the MELP coder.

## 3. ALGORITHM DESCRIPTION

### 3.1 Coder Overview

MELP is based on the traditional LPC model, with additional features included for improved performance. These additional features are mixed excitation, aperiodic pulses, adaptive spectral enhancement, pulse dispersion filtering, and Fourier magnitude modeling. The addition of these features allows the coder to better match the characteristics of the input speech.

### 3.2 Encoder

The first step in the encoding process is to filter out any low frequency noise by use of a 4th order Cherbychev type II high pass filter, with a cutoff frequency of 60 Hz and a stopband rejection of 30 dB. This filtered speech, from now on referred to as the input speech, is again filtered using a 1 kHz 6th order Butterworth lowpass filter, in order to perform the initial pitch search for the pitch estimation. The pitch estimate is the pitch lag between 40 and 160 samples for which the normalized autocorrelation function is maximized.

The next step is to perform the bandpass voicing analysis. The speech is filtered into five frequency bands, using 6th order Butterworth filters with passbands of 0-500, 500-1000, 1000-2000, 2000-3000, and 3000-4000 Hz. The lowest band is used to perform fractional pitch analysis. The bandpass strength for the low band is based on the normalized correlation corresponding to the fractional pitch value that was calculated. If the low band voicing strength is less than 0.5 then the aperiodic flag is set to one, otherwise it is set to zero. For the remaining bands, the bandpass voicing strength is based on the fractional pitch correlations for the bandpass signal and the time envelope of the bandpass signal, where the correlation of the time envelope is first decremented by 0.1 to compensate for a bias.

Now that the bandpass analysis is complete, a 10th order linear prediction analysis is performed on the input speech using a 200 point (25 ms) hamming window centered on the last sample of the current frame. The autocorrelation analysis is implemented using the Levinson-Durbin recursion. In addition each of the prediction coefficients, $a_i$, $i = 1, 2, ..., 10$, is multiplied by $0.994^i$, the bandwidth expansion coefficient. The LPC residual signal can then be calculated by filtering the input speech signal with the prediction filter whose coefficients were determined by the linear prediction analysis. Next a peakiness value is calculated over 160 samples of the residual signal. If the peakiness value exceeds 1.34, then the lowest band voicing strength is forced to 1.0. If the value exceeds 1.6, then the lowest three bandpass voicing strengths are forced to 1.0.

At this point the final pitch estimate is calculated using the lowpass filtered residual signal. First an integer pitch search is performed over lags from 5 samples shorter to 5 samples longer than the fractional pitch. A fractional pitch refinement is then

Copyright 1997 IEEE

made around the optimum integer pitch lag. If the resulting pitch correlation exceeds or is equal to 0.6, a pitch doubling procedure is performed using a set of primary thresholds, otherwise a fractional pitch refinement is performed using the input speech signal. If the refined pitch correlation is less than 0.55 then the pitch for the current frame is replaced by the long-term average pitch, otherwise a pitch doubling check procedure using a set of secondary thresholds is used on the input speech signal.

Once the final pitch is determined then the gain is estimated. The input gain is measured twice per frame using a pitch adaptive window length. The estimated gain is the RMS value of the input signal over the calculated time window, measured in dB. Next the long-term average pitch value is updated with a simple smoothing procedure that uses the pitch correlation and the second gain term as thresholds.

The next step is to quantize the LPC coefficients, pitch, gain, and bandpass voicing (see section 4.1). Before writing out the bits to the bitstream the Fourier magnitudes are determined and quantized. The magnitudes of the first ten pitch harmonics are measured from the LPC residual signal generated by the quantized LPC coefficients. The analysis uses a 512-point FFT of a 200-sample window. The frame of the residual is generated using the quantized LPC coefficients. Then a Hamming window is applied, the signal is zero padded to 512 points, an FFT is taken, and the harmonics are found with a spectral peak-picking algorithm. The magnitudes are then quantized (see section 4.1).

If the frame is unvoiced, the bits are then protected using Hamming codes (See section 4.2) and packed into the bitstream for transmission. Otherwise, if the frame is voiced the bits are just packed into the bitstream.

### 3.3 Decoder

The decoder starts by unpacking the bits and assembling them into the parameter codewords. The method for decoding the parameters is different depending on whether the mode is voiced or unvoiced. The pitch is decoded first since it contains the mode information. If the pitch code is all zero or has only one bit set then the mode is unvoiced and the decoder performs error correction and uses defaults for some of the parameters. If two bits are set in the pitch code then a frame erasure is indicated. Any other pitch code means the mode is voiced and the parameters are decoded. If a frame erasure has occurred then a frame repeat mechanism is implemented.

After decoding the parameters, the decoder looks at noise attenuation. The noise estimator is updated and any gain attenuation is applied, to both gain values, if the input signal is determined to be quiet. The attenuation is a simplified, frequency invariant case of the Smoothed Spectral Subtraction noise suppression method [8]. Noise estimation and gain modification are disabled for repeated frames.

All of the synthesis parameters are then interpolated pitch-synchronously. This includes the LSF's, log speech gain, pitch, jitter, Fourier magnitudes, pulse and noise coefficients for mixed excitation, and spectral tilt coefficient for adaptive spectral enhancement filter. Normally, all of these parameters are linearly interpolated between the past and current frame values. Two exceptions to this rule are when there is an offset with a high pitch frequency and when the second subframe gain is more than 6 dB greater than the previous gain.

The next step is to generate the mixed excitation. The excitation is generated as the sum of the filtered pulse and noise excitations. The pulse excitation is calculated using an inverse DFT of one pitch period in length. The noise excitation is generated by a uniform random number generator and then normalized. These excitations are then filtered and added together.

An adaptive spectral enhancement filter is now applied to the mixed excitation. This filter is a 10th order pole-zero filter, with an additional 1st order spectral tilt compensation. Its coefficients are calculated by bandwidth expansion on the interpolated LPC filter coefficients and adapt based on the signal-to-noise ratio.

The next step is to perform the LPC synthesis. The LPC synthesis uses a direct form LPC filter, with the coefficients corresponding to the interpolated LSF's. The gain is now applied to the synthesized speech. The gain scaling factor is computed for each pitch period. This scale factor is linearly interpolated to prevent discontinuities in the synthesized speech. After applying the gain, the pulse dispersion filter is applied. This filter is a 65th order FIR filter derived from a spectrally-flattened triangle pulse. Finally, some buffering is performed since the synthesizer produces a full period of synthesized speech.

## 4. REQUIREMENTS

In order for an algorithm to be considered compliant with the new 2400 bps Federal Standard it must meet the following set of requirements. This set of requirements was defined by the DDVPC to insure that all implementations of the new Federal Standard provide the same performance as MELP and are interoperable with each other.

### 4.1 Parameter Quantization and Encoding

The new standard uses a number of different coding tables for efficient parameter quantization. Inclusion of these quantization techniques in MELP has shown to improve the performance of MELP while still maintaining a low bit rate. A complete listing of the coding/lookup tables will be included in the Federal Standard document, but due to their size only a description of the quantization techniques used will be provided in this paper. The parameters that are quantized and transmitted are the final pitch, the bandpass voicing strengths, two gain values, the linear prediction coefficients, the Fourier Magnitudes and the aperiodic flag.

#### *4.1.1 Pitch Quantization*

The final pitch and the low band voicing strength are quantized jointly using 7 bits. If the low band voicing strength is less than or equal to 0.6, then the frame is unvoiced and an all-zero code is sent. Otherwise, the log of the final pitch is quantized using a 99-level uniform quantizer ranging from log20 to log160. The resulting index is then mapped to a 7 bit codeword using a lookup table.

#### *4.1.2 Bandpass Voicing*

When the low band voicing strength is greater than 0.6, the remaining four strengths are each quantized to one if their value exceeds 0.6, else they are quantized to zero. One exception is if the quantized values for the four remaining bands are 0001,respectively, then the high band is set to zero. If the low band voicing strength is less than or equal to 0.6, then the remaining four strengths are set to zero.

#### *4.1.3 Gain*

Two gain parameters are transmitted for each frame. The second gain (G2) is quantized into 5 bits using a 32-level uniform quantizer ranging from 10.0 to 77.0 dB. The quantized index is the transmitted codeword. The first gain (G1) is quantized to 3 bits using the following adaptive algorithm.

Copyright 1997 IEEE

```
if (|G2-G2previous|< 5.0 and |G1-0.5*(G2+G2previous)| < 3.0)
    quantizer_index = 0
else
    gain_max = max(G2previous, G2) + 6.0
    gain_min = min(G2previous, G2) - 6.0
    if (gain_min < 10.0) gain_min = 10.0
    if (gain_max > 77.0) gain_max = 77.0
    quantizer_index = G1 quantized with a 7-level uniform
              quantizer ranging from gain_min to gain_max
endif
```

The resulting quantized index plus one is the transmitted codeword.

#### *4.1.4 Linear Prediction Coefficients*

The prediction coefficients are converted into line spectrum frequencies (LSF's) [9] and then forced into ascending order with a minimum separation of 50 Hz. The resulting LSF vectors are then quantized using a multi-stage vector quantizer (MSVQ). The MSVQ codebook consists of four stages whose indices have 7, 6, 6, and 6 bits, respectively. The quantized LSF vector is the sum of the vectors selected by the search process, with one vector selected from each stage. The MSVQ search finds the codebook vector which minimizes the square of the weighted Euclidean distance between the unquantized and quantized LSF vectors. The search procedure is an M-best approximation to a full search, in which the M=8 best code vectors from each stage are saved for use with the next stage. The indices of those four vectors are transmitted.

#### *4.1.5 Fourier Magnitudes*

The ten Fourier magnitudes are coded with an 8-bit vector quantizer. The codebook is searched using a perceptually weighted Euclidean distance, with fixed weights that emphasize low frequencies over higher frequencies. The index of the code vector is transmitted.

#### *4.1.6 Aperiodic Flag*

The aperiodic flag is a single bit and is transmitted as is.

### 4.2 Error Protection

Forward error correction (FEC) is implemented in the unvoiced mode only (when the low band voicing strength is less than or equal to 0.6). The parameters that are not transmitted in the unvoiced mode are the Fourier magnitudes, bandpass voicing and the aperiodic flag. FEC replaces these 13 bits with parity bits from three Hamming (7,4) codes and one Hamming (8,4) code. The (8,4) code is applied to the 4 most significant bits (MSB's) of the first MSVQ index, and the 4 parity bits are written over the bandpass voicing (FEC(1)). The remaining 3 bits of the first MSVQ index, along with a reserved bit (set to zero), are covered by a (7,4) code with the resulting 3 parity bits written to the MSB's of the Fourier series VQ index (FEC(2)). The 4 MSB's of the second gain codeword are protected with 3 parity bits which are written to the next 3 bits of the Fourier series (FEC(3)). Finally, the least significant bit (LSB) of the second gain codeword and the 3 bits of the first gain codeword are protected with 3 parity bits written to the 2 LSB's of the Fourier series and the aperiodic flag (FEC(4)).

### 4.3 Transmission Format

The transmission rate of the new standard shall be 2400 bps ±0.01 percent. Since all frames contain 54 bits, the coder's frame length is 22.5 ms ±0.01 percent. The allocation of the 54 bits in a MELP frame is shown in Table 1.

**Table 1: Bit Allocation**

| PARAMETERS | VOICED | UNVOICED |
|---|---|---|
| LSF parameters | 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* |

Table 2 provides the bit transmission order for each 54-bit MELP frame for both voiced and unvoiced modes. This order matches the LPC-10 Federal Standard with respect to the positioning of the frame's 24 MSB's.

### 4.4 Performance Verification

Once an implementation of MELP is completed, verification of its performance against the reference MELP C simulation must be performed. Verification of the new algorithm will be accomplished by formal quality and intelligibility testing. For intelligibility testing the Diagnostic Rhyme Test (DRT) will be used and for quality testing a forced choice A-B comparison test will be employed. The tests will be performed on the same set of conditions that were used in the 2400 bps selection test (quiet, quiet-vinson, HMMWV, M2 bradley, CH47 helicopter, office, F-15 eagle, E3A AWACS, P3C orion, MCE, auto, single tandem, double tandem, 1% BER, and 0.5% BLER). In addition to testing the algorithm's performance on its own, an interoperable performance test will performed with the reference MELP coder and the implementation, using both of the above quality and intelligibility tests. Results from both the quality and intelligibility tests will have to meet individual condition thresholds as well as overall quality and intelligibility thresholds, as defined by the DDVPC, in order for the implementation to be considered Federal Standard compliant.

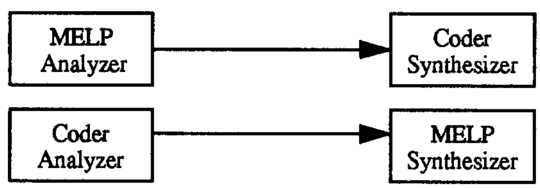


**Figure 1: Interoperability Testing**

## 5. PERFORMANCE

In the tests performed during the 2400 bps coder selection, MELP performed equal to or better than the Federal Standard CELP algorithm. In most cases it outperformed CELP by a considerable amount. For more information on MELP's

Copyright 1997 IEEE

**Table 2: Bit Transmission Order**

| 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; See section 3.2 for details
Bit 1 = least significant bit of data set
BP = Bandpass Voicing
LSF = Line Spectral Frequencies
FM = Fourier Magnitudes
AF = Aperiodic Flag

performance please refer to the following paper [10].

## 6. CONCLUSION

The new 2400 bps Federal Standard will be based on MELP. MELP provides equivalent or better performance than the 4800 bps Federal Standard CELP coder at a lower bit rate. The new standard will be bitstream defined and will require implementers to verify their implementations in order for them to be considered Federal Standard compliant. Coding tables for quantization and some filter coefficients will be supplied in the standard. MELP meets the user requirements for a new 2400 bps coder.

## 7. ACKNOWLEDGEMENTS

The authors would like to thank Thomas Tremain for his vision and drive that made this new standard a reality. We wish to give special recognition to Kwan Truong, Stephen McGrath, and Thomas Barnwell, from ASPI, and E. Bryan George, Vishu Viswanathan, Wai-Ming Lai, and Wilf LeBlanc, from TI, for their outstanding contributions in the development of the MELP coder. We also would like to thank the DDVPC for all the effort involved in managing the 2400 bps selection test.

## 8. REFERENCES

[1] T. E. Tremain, M. A. Kohler, T. G. Champion, "Philosophy and Goals of the DoD 2400 bps Vocoder Selection Process," *Proceedings of ICASSP*, IEEE, pp. 1137-1140, 1996

[2] M. Bielefeld, "1995 Terms of Reference (TOR) for the Selection of a New Government Standard Voice coder at 2400 bps," MP-8932D-95-05, The MITRE Corporation, 1995

[3] M. Bielefeld, L. M. Supplee, "Developing a Test Program for the DoD 2400 bps Vocoder Selection Process," *Proceedings of ICASSP*, IEEE, pp. 1141-1144,1996

[4] M. Bielefeld, "1995 Test and Evaluation Plan (TEP) for the Selection of a New Government Standard Voice coder at 2400 bps," MTR MTR95W0000136, The MITRE Corporation, 1995

[5] A. McCree, K. Truong, E. B. George, T. P. Barnwell, V. Viswanathan, "A 2.4 kbit/s MELP Coder Candidate for the New U. S. Federal Standard," *Proceedings of ICASSP*, IEEE, pp. 200-203, May 1996

[6] A. McCree, T. P. Barnwell III, "A Mixed Excitation LPC Vocoder Model for Low Bit Rate Speech Coding," *IEEE Trans. Speech and Audio Processing*, vol. 3, pp. 242-250, July 1995

[7] T. E. Tremain, "The Government Standard Linear Predictive Coding Algorithm: LPC-10," Speech Technology, pp. 40-49, April 1982

[8] L. Arslan, A. McCree, and V. Viswanathan, "New Methods for Adaptive Noise Suppression," *Proceedings of ICASSP*, IEEE, pp. 812-815, 1995

[9] P. Kabal and R. P. Ramachandran, "The Computation of Line Spectral Frequencies Using Chebyshev Polynomials," *IEEE Transactions on ASSP*, vol. ASSP-34, No. 6, pp. 1419-1426, December 1986

[10] M. A. Kohler, "A Comparison of the New 2400 bps MELP Federal Standard with Other Standard Coders," *Proceedings of ICASSP*, IEEE, April 1997

Copyright 1997 IEEE