Real Time Implementation of 600 bps MELP Vocoder

Ranjith.T [1], Subha. P. E [2] and T.R. Ramamohan [3]
Central Research Laboratory, Bharat Electronics Limited, Bangalore-560013

ranjith.thachan@gmail.com [1], pe.subha@yahoo.in [2] and tr.ramamohan@yahoo.in [3]

***Abstract* - The Federal Standard MELP (Mixed Excitation Linear Prediction) speech coder is known to provide good quality decoded speech at 2400 bps[1], [5]. For use in narrow band HF channels it is often desirable to have speech coders which work at much lower bit rate. This paper presents a 600 bps MELP vocoder implementation that takes advantage of inherent inter-frame redundancy of the MELP parameters [2]. Multi-frame based vector quantization technique is used to encode the MELP parameters to affect bit rate reduction. The output speech quality of the 600bps-vocoder was found to be as good as that of federal standard 2400 bps LPC (Linear Prediction Coding). The algorithm was ported on to a fixed point DSP (Digital Signal Processor) and stage by stage optimization was performed to meet the real time requirements.**

I. INTRODUCTION

The MELP coder was designed to overcome some of the key limitations of the LPC vocoder. It utilizes a more sophisticated speech production model, with additional parameters to capture the underlying signal dynamics with improved accuracy. Originally developed by McCree as a Ph. D. thesis, it was later refined and submitted as candidate for U. S Federal Standard voice coding at 2.4 kbps [5].

After a short description of the standard MELP and its low-bit rate version, the DSP implementation aspects and optimization results are reported.

II. MELP 2400 VOCODER PARAMETERS

MELP vocoder is basically an improvisation upon the highly simplistic LPC speech coder. The vocoder has incorporated several techniques such as vector quantization, speech enhancement and superior analysis/synthesis methods. The essential idea behind the MELP speech model is the generation of a mixed excitation signal as an input to the synthesis filter, where the "mixing" refers to the combination of a filtered periodic pulse sequence with a filtered noise sequence[1], [5]. In addition to the traditional LPC parameters such as gain, pitch and spectral features, MELP speech model consists of Aperiodic-flag, Fourier magnitude and Band- pass voicing strength parameters.

The mixed excitation is implemented using a five band mixing model. The primary effect of this multi-band excitation is to reduce the buzz usually associated with LPC10 vocoders. The parameters estimated for this purpose are encoded as band-pass voicing strengths [1].

The shape of the excitation pulses are captured by means of Fourier magnitudes which are computed over the prediction-error signal that comes out of LP analysis [5]. The objective is to create, on the decoder side, a periodic sequence as close as possible to the original excitation signal. Another set of parameters called aperiodic pulses are most often used during transition regions between voiced and unvoiced segments of speech signal. This indicates the synthesizer to generate a jittery excitation [5].

The Line Spectral Features (LSFs) which are compact representations of the LPC spectrum are quantized through a four stage vector quantization algorithm [4].

In addition to the encoded parameters, MELP speech model makes use of adaptive spectral enhancement filter based on the poles of the LPC vocal tract filter [2]. It helps in enhancing the formant structure of the synthetic speech.

III. MELP 600 VOCODER PARAMETERS

The inherent inter-frame redundancy of the MELP parameters can be exploited to design lower bit rate coders. The low-bit rate version consists of the Federal standard MELP front end, a block buffer for accumulating multiple frames of MELP parameters, and individual block vector quantizers for MELP parameters. A diligent study of the parameter entropy [2] results in determining a good choice of block length for quantization at low bit rates. The 600 bps system uses a 27.5ms frame length (220 samples, at a sampling rate of 8000Hz) and a block buffer (multi-frame) of four frames. This yields a total of 66bits per multi-frame (110 ms), or 600 bits per second. Bit allocation is listed in Table-2. The information provided in the paper on 600 bps MELP vocoder [2] served as guidelines for deciding the bit rate allocation. The multi-frame quantization methodology for each parameter set is given as follows.

A. Spectral Feature Quantization.

The low-rate quantization of the spectrum quantizes four frames of LSFs in sequence using a multi-stage vector quantization process. The first two stages use 10 bits each while the remaining two use 9 bits each for quantization. Thus, the four frames of spectra are quantized to only 38 bits.

B. Band Pass Voicing Quantization

The band-pass voicing strengths determine the voiced/unvoiced nature of the five bands of the MELP speech model. The least significant (v/u) bit is encoded along with pitch while the upper four bits are encoded separately. It was observed that certain patterns of v/u combinations come more often than the others. The probability density is tabulated as shown in Table-1. This indicates that, for individual frame, without much distortion in quality, a level-4 quantiser is realizable. The, case of all band un-voicing is taken care of by pitch encoding. Because, the first band voicing information is encoded along with pitch. Further reduction in bit rate can be obtained by taking advantage of frame to frame redundancy of the voicing decisions. The 600 bps MELP vocoder uses a six bit code-book to quantize the most probable voicing transitions that occur over a four-frame block.

C. Energy and Pitch Quantization

The energy parameters of MELP exhibit a considerable amount of frame-to-frame correlation. A sequence of energy values from four frames are grouped together and quantized to 12 bits.

The refined pitch values of successive frames of the standard MELP also exhibits a significant frame to frame correlation suggesting that bit rate reduction is possible by the vector quantization of the parameters of nearby frames. Vector quantization was performed over pitch values of four frames to yield a bit allocation of 10 bits.

BPV decisions	Probability
Prob(u,u,u,u,u)	0.1245
Prob(v,u,u,u,u)	0.1668
Prob(v,v,u,u,u)	0.1424
Prob(v,v,v,u ,u)	0.1320
Prob(v,v,v,v,v)	0.2858
Prob(remaining)	0.1485

Table 1- v/u probability distribution

Speech Parameters	Bits
Aperiodic flag	0
Band-Pass voicing	6
Energy	12
Fourier Magnitudes	0
Pitch	10
Spectrum	10+ 10+ 9+ 9
Total	66

Table 2 – Bit rate allocation of 600 bps MELP vocoder.

D. Aperiodic flag and Fourier magnitude Quantization

The effect on voice quality, with aperiodic flag indicator removed was observed to be very little. So was the case when the Fourier magnitude value was kept constant. Therefore, it was decided to quantize the more important parameters better, at the cost of these parameters.

IV. CODEBOOK TRAINING

The LBG Vector quantization method [3] was used throughout to create the codebooks for the gain and pitch parameters. TIMIT database, which contains speech from 630 speakers from major eight major dialects of American English, each speaking ten phonetically rich sentences. In addition to that we included several speech files recorded from Indian speakers. A major chunk of such recorded speech files were provided by IIT Bombay. The remaining speech files were recorded in our Laboratory. Multi-stage vector quantizer [4] with LBG vector quantization technique at the core was used for LSF codebook generation. M-Best search method was used for encoding of the LSF vectors.

V. VOCODER PERFORMANCE

Objective measures as well as subjective listening were used to judge the quality of the vocoder output speech. A standard scale for speech quality assessment, MOS (Mean Opinion score) provided by the ITU-T standard P.862 PESQ (Perceptual Evaluation of Speech Quality) was used for objective performance evaluation. The PESQ provides a score between -0.5 to 4.5. A PESQ MOS measure of 4.5 indicates there is no perceivable difference between reference and degraded signal, and a value of -0.5, indicates there is absolutely no correlation between the two. The average PESQ MOS measures for MELP 2400, LPC 2400 and MELP 600 are shown in table 3. The MOS measure for MELP2400 is far superior for obvious reasons, compared to the other two standards. Interestingly, the 600bps version of MELP shows a MOS score comparable to that of LPC at 2400 bps.

Speech intelligibility was tested using subjective listening based on MRT (Modified Rhyme Test). The MRT list consists of 50 sets of single-syllable rhyming words that differ in one consonant. Vocoder intelligibility is measured by average scores obtained for correctly recognized words over a number of listeners.

Vocoder type	Average PESQ-MOS Value
MELP 2400	3.35
LPC 2400	2.72
MELP 600	2.65

Table 3 - Vocoder performance comparison

The average percentage of intelligibility of MELP 2400 was found to be 93%. For, MELP 600 it was 82%. Whereas, that of LPC- 2400 was 80.5%.

VI. Hardware implementation

The 2400 bps MELP source code was developed by NSA, Microsoft, ASPI, Texas Instruments, and ATT [7]. This source code together with the information given in the paper on 600 bps MELP vocoder [2] served as a reference for coming up with our version of the 600 bps system.

A. Offline testing

The fixed point source code was first modified for MELP 600 on Microsoft's Visual C++ platform. Functionality verification was done and, it was migrated to Analog Devices' DSP Integrated Development Environment, known as VDSP++. The algorithm was simulated offline with single processor simulator to see that the functionality is consistent.

B. Experimental Set up

The verified algorithm was ported on to a Black fin 533[9] based proprietary board. The board was part of an HF radio. The DSP hardware design runs the on chip core at 405 MHz while the off-chip accesses are limited to 150MHz. An external memory chip (SDRAM) was also provided to meet the additional memory requirements. The non-real time implementation was functionally verified and subjected to profiling.

C. Optimization and Real time implementation

The optimization was necessary to meet the real time requirement of completing all computation processes within frame duration. The main functions involved are analysis, parameter encoding, parameter decoding and synthesis. The multi-frame structure of 600 bps MELP system consists of four frames with each frame duration of 27.5 ms. The analysis () function is computed every frame, whereas the encoding is done only after the analysis of four frames. Considering the multi-frame structure and the large codebook sizes, encoding takes much of the processor cycles. Though, the process has to be done only once every four frame, to avoid any transmission delay, it should be completed within frame duration. So is the case with decoding and synthesis. However, these functions are not as much computationally intensive. The entire computation processes took close to 100ms, which was not acceptable considering the available frame duration. This was in spite of the level- 3 optimization provided by the compiler. The fixed point source code at the MELP front end was also thoroughly optimized at the C –Level. However, they are highly inefficient, as the DSP compilers have not evolved enough to exploit the tailored resources [8].

For bringing in further optimization assembly intrinsic functions (provided by Analog Devices) was used, wherever possible. In addition to that, computationally expensive functions were identified and were replaced with custom assembly codes.

X - No of clock cycles consumed by original code (before optimization)

Y - No of clock cycles consumed by optimized code.

CRR - Clock rate reduction.

$$CRR = \frac{X - Y}{X} x100 \ \%$$

Memory optimization techniques [8] such as data placement and caching were also used to bring down the processing time. It was not possible to accommodate the entire data inside the internal RAMs of the DSP. Therefore, less frequently accessed data was kept in SDRAM which was comparatively slower. Frequently accessed functions were cached.

The optimization results are given in table-4. The number of cycles taken by the original code, without optimization is listed in the fist column. The number of cycles consumed by the optimized code is listed in the second column. The third column shows the clock rate reduction achieved.

Function	Original Code: X-Clock Cycles	Optimized code: Y-Clock cycles	Clock Rate Reduction *(CRR)*
Analysis()	7,348,868	486,035	93.38%
Parameter encode()	80,254,544	6,277,542	92.18%
Parameter decode()	66,344	37,575	43.36%
Synthesis()	3,525,548	1,392,284	60.5%

Table 4 – Cycles reduction

Most of the cycles reduction was affected by using intrinsic assembly routines for commonly used functions. Further optimization was achieved by writing the codes of the computationally intensive functions in assembly, with the architecture of the processor in view. For instance, the parameter encode module was found to be taking large number of cycles, therefore the search function inside that was coded in assembly. It resulted in substantial reduction in number of cycles taken for computation.

VII. CONCLUSION

The inter-frame redundancy of 2400 bps MELP parameters was successfully exploited to derive a 600 bps version. The algorithm was ported onto a Blackfin ADSP-BF533 (fixed point processor). Thorough optimization of the code was carried out to affect real-time implementation. The vocoder was integrated with HF radios and tested in lab environment. The output speech was fairly intelligible with a MOS measure of 2.65 on the average.

REFERENCES

[1]. McCree. A. V, Lynn M. Supplee, Ronald. P. Cohn, John. S. Collura (1997), "MELP: The New Federal Standard at 2400 bps", *IEEE ICASSP pp. 1591-1594.*

[2]. Mark. W. Chamberlain "A 600bps MELP Vocoder for use on HF Channels", *IEEE Military Communications conferences, 2001.*

[3]. Y. Linde, A. Buzo, and R.M.Gray, "An Algorithm for Vector Quantization Design," *IEEE Transaction communication, pp 702-710, January 1980.*

[4]. W. P. Leblanc, B. Bhattacharya, S. A. Mahmoud and V.Cuperman, "Efficient Search and Design Procedures for Robust Multi-Stage VQ of LPC Parameters for 4kb/s Speech Coding", *IEEE Trans. Speech & Audio Processing, Vol. 1. Pp.272-285, October 1993.*

[5]. Wai C. chu, "Speech Coding Algorithm", *Wiley Publication*.

[6]. Modified rhyming Test: http://www.icphs2007.de/conference/Papers/1210/1210.pdf

[7]. MELP source code link: http://health.tau.ac.il/Communication%20Disorders/noam/speech/melp/Download/Download.htm

[8]. Benix Samuel and Ashok Jhunjhunwala "Optimal Porting of Embedded Software on DSPs" *Proceedings of the Fourteenth National Conference on Communications- 2008.*

[9] .Analog devices, Inc."*ADSP-BF533 Blackfin Processor Hardware Reference*" Revision 3.1 June 2005 Part Number 82-000555-01.

[10]. Analog devices, Inc. "*Blackfin Processor Instruction set reference*" revision 3.0, July 2004 Part Number 82-000410-14.