

The 1.2Kbps/2.4Kbps MELP Speech Coding Suite with Integrated Noise Pre-Processing

John S. Collura, Diane F. Brandt, Douglas J. Rahikka

National Security Agency
9800 Savage Rd, STE 6516, Ft. Meade, MD 20755-6516, USA
Phone: (301) 688-0269 Fax: (301) 688-0289 E-mail: jscollu@alpha.ncsc.mil

ABSTRACT - Recent advances in speech enhancement and noise pre-processing algorithms have dramatically improved the quality and intelligibility of speech signals, both in the presence of acoustic noise as well as in benign environments. The use of speech enhancement in combination with voice coding algorithms and applied to governmental wireless communications systems is an area of increasing importance.

In March 1996, the US government selected the 2.4Kbps MELP speech coding algorithm for applications which require high quality/intelligibility narrow band speech coding. In March of 1999, a companion 1.2Kbps MELP speech coding algorithm was unveiled. This additional MELP rate was added to satisfy disadvantaged transmission requirements such as survivable SATCOMs or high latitude HF radio links. The combined 1.2Kbps/2.4Kbps MELP algorithm incorporates an integrated speech enhancement algorithm as a front-end process to provide superior performance in harsh acoustic noise environments.

This paper introduces the integrated speech enhancement 1.2Kbps/2.4Kbps MELP speech coding algorithmic suite, and presents the first available test data for the combined algorithm. Finally, discussion of the test conditions, results and related issues will be presented.

I. INTRODUCTION

Legacy tactical secure voice communications use either the 2.4Kbps Linear Predictive Coding (LPC10e) algorithm or the 16Kbps Continuously Variable Slope Delta Modulation (CVSD) algorithm for speech compression. These algorithms were considered the state of the art in narrow band speech coding when they were introduced 20 to 30 years ago. In fact, it's a testimony to their effectiveness that until 1996 there were no alternative narrow band speech coding algorithms available to the tactical user.

In March of 1996, the U.S. government DoD Digital Voice Processing Consortium (DDVPC) announced the selection of the 2.4Kbps Mixed Excitation Linear Prediction (MELP) voice coding algorithm as the next standard for narrow band secure voice coding products and applications [1]. The selection of the MELP voice-coding algorithm represented a dramatic improvement in both speech quality and intelligibility at the 2.4Kbps data rate.

One of the driving forces behind the selection of the 2.4Kbps MELP algorithm was operation in harsh acoustic noise environments such as HMMWV's, helicopters and tanks. In these harsh acoustic environments, the MELP algorithm, while providing superior performance over legacy systems, exhibited somewhat degraded performance, thus demonstrating a need for improvement. Such improvement was addressed by the development and ongoing integration of a speech enhancement algorithm into the front end of the MELP speech coding algorithm.

There are a number of important transmission channels which don't support *robust* speech coding transmission at or above 2.4Kbps. These channels include survivable SATCOMs, high latitude HF links and covert operations, etc. To accommodate these requirements, a 1.2Kbps MELP algorithm was developed which shares the 2.4Kbps core algorithmic paradigm, but with an alternate quantization of the parameters. This paper introduces the resulting combined 1.2Kbps/2.4Kbps MELP algorithm suite including an integrated speech enhancement front end. The 1.2Kbps/2.4Kbps MELP algorithm coupled with the speech enhancement algorithm will be referred to as the MELPn algorithm.

The limited performance of the current generation voice coding systems (including the MELP algorithm) in harsh acoustic noise environments has given rise to the idea of enhancing the voice signal prior to compression. This paper discusses the combination of one such noise pre-processing algorithm developed at AT&T Labs Research in conjunction with the US government. Performance data presented in this paper will include quality and intelligibility testing, with and without noise pre-processing, and under two main scenarios. The first tested the system performance in the quiet environment, while the second tested the performance in the HMMWV and CH-47 harsh acoustic noise environments. The HMMWV is a heavy-duty four-wheeled drive vehicle used for troop transport. Due to the low gear ratios, and four-wheeled drive operations, the acoustic character of the HMMWV background noise is dominated by non-stationary low frequency rumbling with an average (over six speakers) speech SNR of approximately 12.6 dB. The CH-47 is a turbine driven tandem rotor heavy lift helicopter. This acoustic environment is characterized by both the beat

frequency of the rotors (quasi-stationary noise components) and the Gaussian type noise generated by the turbine operation. The CH-47 has an average (over six speakers) speech SNR of approximately 13.75 dB.

II. SYSTEM CONFIGURATION AND DESCRIPTION

A. *Speech Enhancement Algorithm*

Current generation noise pre-processing algorithms have generally been effective in noisy environments at the expense of introducing objectionable structured musical artifacts to the enhanced signal. An additional drawback was the need to manually switch the noise pre-processing algorithm depending upon the nature of the acoustic environment. This switching was needed due to the degradation inherent with the application of noise pre-processing algorithms to benign acoustic environments. These older noise pre-processing algorithms have tended to improve the quality but degrade the intelligibility of the noisy speech signal.

The AT&T speech enhancement/noise pre-processing algorithm is the culmination of several years of research (for a detailed description, see the following papers, [2][3][4]). This speech enhancement algorithm is based upon the following heuristic operational description. First, the speech signal is divided into time slices of 32 ms in length, and an FFT is applied to provide access to the spectral information contained in the signal. Next, an estimation algorithm is used to model the noise for frames in which speech is absent. This part of the algorithm uses a voice activity detector to enable the algorithm to distinguish between frames which are composed of speech + noise, and those which are noise only. A model for the noise is then maintained based upon those frames, which are noise only. The algorithm minimizes the mean squared error, (MMSE) of the log spectral amplitudes. It tracks the probability of speech presence and applies an additional gain factor based upon these probabilities. These probabilities are also used to update the noise power spectral density during speech. Once all of the bins have been accounted for, the resulting magnitude spectrum is then put through an Inverse FFT to recover the enhanced speech. This last step is of course followed by a synchronized overlap and add, (SOLA) technique, which helps to eliminate many of the artifacts. This enhanced speech is then used as the input to the speech coding algorithm, in this case the MELP algorithm. Work is proceeding on the integration of the speech enhancement algorithm into the front end of the MELP algorithm resulting in a new algorithmic designation of MELPn for the combined algorithm. The AT&T speech enhancement algorithms used for the bulk of this paper are versions 5 & 7, and will be designated as NPP5 and NPP7 in this paper. The

main difference between these versions lies in the frame update rate. NPP5 operates on a 32ms frame while NPP7 was made to synchronize with the MELP voice coder basic frame rate of 22.5ms. To accommodate the smaller frame size, the parameters of NPP7 were retuned to maintain system performance. This tuning process resulted in a higher level of overall performance.

Initial quality and intelligibility testing of the AT&T speech enhancement algorithm used in conjunction with the MELP voice coders indicates that a fundamental breakthrough has been achieved in the operation of noise pre-processing algorithms. Testing shows that speech quality and intelligibility improve in both harsh acoustic noise environments as well as the benign acoustic environments. The need for an algorithmic switch for the noise pre-processor will be eliminated once a more extensive test verifies these initial test results.

B. *2.4Kbps MELP Speech Coding Algorithm*

The 2.4Kbps MELP speech coding algorithm was selected by the US Government in March of 1996 to be the primary voice coding algorithm for seamlessly interoperable narrow band secure voice communications for strategic, tactical and SATCOMs applications. The selection of 2.4Kbps MELP as the standard voice coding algorithm provides the first enabling technology for seamless interoperability across the domains, regardless of the points of presence of the users in the communications networks. In fact, the end users probably will not be aware of the intermediate routing (SATCOM links, Internet Protocols, etc...) used to establish the end-to-end communications link.

The 2.4Kbps MELP algorithm divides the 8KHz sampled speech signal into 22.5ms frames for analysis. Table 1 provides a breakdown of the parameters used by the MELP algorithm with the number of bits per frame needed to quantize each parameter. A complete description of the MELP algorithmic paradigm can be found in the MELP FIPS draft standards publication at <http://www.plh.af.mil/ddvpc/frontpage.html>.

PARAMETER:	NUMBER OF BITS/FRAME
LSF's	25 Bit Multistage VQ
10 Fourier Magnitudes	8 Bits VQ
Pitch	7 Bits
Bandpass Voicing	4 Bits
Gain	2 x 4 Bits
Aperiodic Pulse	1 Bit
Sync	1 Bit

PARAMETER:	NUMBER OF BITS/3 FRAMES
LSF's	43 Bit Multistage VQ
10 Fourier Magnitudes	8 Bits VQ
Pitch + Global Voicing	12 Bits
Bandpass Voicing	6 Bits
Gain	10 Bits
Aperiodic Pulse	1 Bit
Sync	1 Bit

C. 1.2Kbps MELP Speech Coding Algorithm

The 1.2Kbps MELP speech coding algorithm was developed under contract with SignalCom Inc. in response to requirements by the user community (HF, Survivable SATCOMs, Covert Users, NATO, etc...) for a high performance lower bit rate algorithm. At the time that these requirements were under consideration, the 2.4Kbps MELP algorithm had recently been selected as the interoperable standard. It was subsequently decided that a 1.2Kbps version of MELP would make a needed addition to the MELP technology base. A single MELP algorithmic suite with two distinct quantization modes allows for efficient inclusion of both rates in secure voice equipments. Modern modem technology allows for seamless adaptation of the transmission rate to accommodate variations in channel availability or error rates.

Full disclosure of the algorithmic details will follow once research is concluded on the final 1.2Kbps algorithm. In lieu of full disclosure, a brief description of one mode of operation is provided here. The 1.2Kbps MELP algorithm divides the 8Kbps sampled speech signal into groups of three 22.5ms frames into a 67.5ms super frame for analysis. Depending upon the type of speech present in the signal, inter-frame redundancy can be exploited to efficiently quantize the parameters. For instance in the case where all three frames in the super frame are classified as voiced, the parameters are quantized as described in Table 2.

III. TESTING

A. Test Description

The object of this work was to measure any performance gains or degradations for communications equipment when the MELP voice coding algorithm is preceded by a speech enhancement algorithm. To do this objectively, the test included the most benign acoustic environment possible, that of the quiet background, and two of the harsher acoustic noise environments, the HMMWV, and the CH47 helicopter.

Two subjective measures were used to evaluate coder performance. The Diagnostic Rhyme Test (DRT) [5] was used to measure speech intelligibility, and the Diagnostic Acceptability Measure (DAM) [6] was used to measure speech quality. Both of the tests have been used extensively in previous DDVPC coder selection efforts. All coders were evaluated with six speakers: three males and three females.

The DRT is a two choice intelligibility test based upon the principle that the intelligibility relevant information in speech is carried by a small number of distinctive features. The DRT was designed to measure how well information as to the states of six binary distinctive features (voicing, nasality, sustension, sibilant, graveness, and compactness) have been preserved by the communications system under test. The DRT uses a suite of 96 rhyming word pairs (192 items per speaker), in which the initial consonants of the two words of each pair differ only with respect to one of the distinctive features. The listener must select which of the two rhyming words were spoken. With a carefully selected and monitored panel of eight listeners, the DRT has extremely high resolving power and test-retest reliability.

The DAM is a proprietary test developed and administered by Dynastat, Inc. in Austin, Texas. The DAM requires the listeners to judge the detectability of a diversity of elementary and complex perceptual qualities of the signal itself, and of the background environment. The qualities which are evaluated have been experimentally shown to

determine a listener's judgements of speech acceptability. The DAM thus provides multiple direct and indirect estimates of a communication system. The DAM is designed for use with small (12-16) crews of listeners, who are rigorously screened, trained, calibrated, and monitored to ensure their collective response very closely approximates that of the typical or normative listener.

B. Test Results

Several tests were run to exercise the performance of the MELP algorithms both with and without the speech enhancement front end. Additionally, the performance of NPP5 without speech coding was contrasted against that of the 128Kbps PCM source material for both quiet and HMMWV acoustic environments. Tables 3 through 5 below present performance data for three acoustic noise environments quiet, HMMWV, and the CH47 "Chinook" heavy lift helicopter. It should be noted that the data for Tables 3 through 5 reflects the latest available version of the speech enhancement software for the condition tested. Reporting of the 1.2Kbps MELP coder results reflects the integration of version 5 of the AT&T speech enhancement (noise pre-processing) algorithm with the voice coding software. At the time of the test, the parameters of the integrated speech enhancement algorithm had not been re-optimized for the 22.5ms frame size. Reporting of the 2.4Kbps MELP coder results reflects speech enhancement algorithm version 7, (fully optimized with a 22.5ms lower delay) which was implemented in a separate pre-processing stage for the 2.4Kbps testing.

TABLE 3: QUIET ENVIRONMENT		
TEST CONDITION	DRT/S.E.	DAM/S.E.
Source Material	97.8/0.23	84.1/2.1
Noise Pre-Processing Only	96.7/0.52	85.1/1.3
2.4Kbps MELP w/ NPP7	93.8/0.65	66.3/0.9
1.2Kbps MELP w/ NPP5	92.6/0.65	63.4/0.9
2.4Kbps MELP Coding Only	93.0/0.89	64.9/1.0

TABLE 4: HMMWV ENVIRONMENT		
TEST CONDITION	DRT/S.E.	DAM/S.E.
Source Material	91.0/0.37	45.0/1.2
Noise Pre-Processing Only	80.6/0.66	54.6/1.2
2.4Kbps MELP w/ NPP7	74.4/0.83	52.6/0.9
1.2Kbps MELP w/ NPP5	67.8/0.77	52.1/0.6

2.4Kbps MELP Coding Only	67.3/0.8	38.9/1.1
--------------------------	----------	----------

TABLE 5: CH47 HELICOPTER ENVIRONMENT	
TEST CONDITION	DRT/S.E.
2.4Kbps MELP + NPP7	76.9/0.77
1.2Kbps MELP w/ NPP5	69.1/0.81
2.4Kbps MELP Coding Only	66.7/0.61

TABLE 6: MELP + NOISE PRE-PROCESSING (HISTORY IN HMMWV ENVIRONMENT)		
TEST CONDITION (date tested)	DRT/S.E.	DAM/S.E.
Original 2.4Kbps MELP (Mar 96)	63.1/0.72	N/A
Updated MELP (Jan 98)	67.3/0.8	38.9/1.1
MELP + NPP version 3 (Jun 98)	72.3/0.92	48.4/0.6
MELP + NPP version 5 (Oct 98)	72.0/0.64	52.0/0.9
MELP + NPP version 7 (Jan 99)	74.4/0.83	52.6/0.8

C. Interpretation of Test Results

Tables 3 and 4 cover the same test conditions, but in different acoustic noise environments, quiet and HMMWV respectively. The first 2 test conditions for Tables 3 & 4 provide a reference for the other conditions by contrasting the performance of the original source material with that of NPP5 without the additional influence of speech coding. Then, the data for the 2.4Kbps MELP algorithm with and without NPP7 is presented. Finally, the 1.2Kbps MELP coder with integrated NPP5 is presented.

Table 3 presents data for testing performed in the benign quiet acoustic noise environment. This data indicates that noise pre-processing introduces minimal degradation with respect to both speech intelligibility and quality. When the performance of the 2.4Kbps MELP algorithm is compared with that of the 2.4Kbps MELPn algorithm, the difference is borderline positive for the intelligibility measurement with a clear benefit for the quality measurement. Of particular interest here is the marginal difference for both quality and intelligibility for the 1.2Kbps MELPn algorithm when compared with that of both the original 2.4Kbps MELP and 2.4Kbps MELPn versions.

Table 4 presents data for testing performed in the HMMWV acoustic noise environment. Most notable in this data is the simultaneous decrease in the intelligibility and increase in quality scores for the NPP5 algorithm without speech coding as contrasted against the unprocessed source

material. Both the 1.2Kbps and 2.4Kbps MELPn algorithms demonstrate higher overall voice quality than the unprocessed source material, and though the intelligibility scores are considerably degraded for 2.4kbps MELPn when compared with the unprocessed source material, they are considerably higher than that of the original 2.4Kbps MELP algorithm. Likewise, the performance of the 1.2Kbps MELPn algorithm out-performs the original 2.4Kbps MELP algorithm for both quality and intelligibility, although the intelligibility scores are within the standard error of the test.

Table 5 provides intelligibility scores for the CH47 helicopter environment. This information provides a simple verification that the performance gains exhibited in both the HMMWV and quiet environments were applicable to other acoustic environments. This corroboration of the test results demonstrates a consistency of operation regardless of the type of acoustic noise. Most notable in this data is the superior performance of the 1.2Kbps MELPn over original 2.4Kbps MELP.

Table 6 demonstrates the improvements which have been achieved with the MELP algorithm in the HMMWV acoustic noise environment from the initial selection in the spring of 1996 to the current MELPn algorithm using NPP7. These test scores demonstrate an intelligibility improvement of 7 to 10 points and a quality improvement of up to 14 points. Given standard errors of between 0.6 to 1.2 points, a DAM or DRT test improvement from one condition to another of 10 points is nothing less than phenomenal! These results are further corroborated by the CH47 helicopter scores. The important thing to note here is that the noise pre-processing algorithms operate well in both stationary and non-stationary types of acoustic background noise.

The development of the 1.2Kbps MELP algorithm and the integration of speech enhancement as a component of MELP establishes the baseline MELP technological suite. Additional components of this suite are either in development, or are under consideration. These include an 8Kbps MELP algorithm, a parametric 1.2/2.4 transcoding algorithm, and a 600bps MELP algorithm. The aim of providing such a comprehensive MELP voice coding suite is to make MELP available to the widest possible group of customers while enabling the concept of seamlessly interoperable government wide secure voice communications.

In the near future a full verification test will be run which compares the performance of the integrated MELPn (NPP7) algorithm with that of the original MELP algorithm as sanctioned by the DDVPC. The 1.2Kbps MELPn test data will set the performance standard which will be required of any subsequent DDVPC sanctioned implementation.

IV. ACKNOWLEDGMENTS

We wish to thank Dr. Richard Cox and Dr. Rainer Martin of AT&T Labs Research for their work on the Speech Enhancement algorithm and Dr. Vladimir Cuperman and Dr. Allen Gersho and the staff at SignalCom Inc. for their work on the 1.2Kbps MELP algorithm.

V. REFERENCES

- [1] Supplee Lynn M., Cohn Ronald P., Collura John S., McCree Alan V., "MELP: The New Federal Standard at 2400 bps", IEEE ICASSP-97 Conference, Munich Germany, pp.1591-1594.
- [2] Malah David, Cox Richard V., Accardi Anthony J., "Tracking Speech Presence Uncertainty to Improve Speech Enhancement in Non-Stationary Noise Environments." IEEE ICASSP-99 Conference, Phoenix AZ.
- [3] Accardi Anthony J. and Cox Richard V., "A Modular Approach to Speech Enhancement with an Application to Speech Coding". IEEE ICASSP-99 Conference, Phoenix AZ.
- [4] Martin Rainer, Cox Richard V., "New Speech Enhancement Techniques for Low Bit Rate Speech Coding." IEEE Speech Coding Workshop-99, Porvoo Finland.
- [5] Voiers William D., "Diagnostic Evaluation of Speech Intelligibility." In M.E. Hawley, Ed, Speech Intelligibility and Speaker Recognition (Dowden, Hutchinson, and Ross; Stroudsburg, PA 1977)
- [6] Voiers William D., "Diagnostic Acceptability Measure (DAM): A Method for Measuring the Acceptability of Speech over Communications Systems." Dynastat, Inc.; Austin Texas