# PROGRESS TOWARDS A NEW GOVERNMENT STANDARD 2400 BPS VOICE CODER

M. A. Kohler, L. M. Supplee, T. E. Tremain

U. S. Department of Defense, 9800 Savage Road, Ft. Meade, MD 20755

## ABSTRACT

In order to support the need for higher quality low rate voice communications for government, industry, and military customers, the United States Government is conducting a search for a new voice compression algorithm at 2400 bits per second (bps). The United States Department of Defense Digital Voice Processing Consortium (DDVPC), consisting of members from civilian and military branches of the U.S. government, is directing the testing and evaluation of several candidate 2400 bps algorithms. The goal of the DDVPC is to select a new algorithm which meets or exceeds the published requirements by mid 1996. The selected algorithm, to become the new standard, should be implementable in a small, low powered device by 1997. This paper describes the status of the testing and evaluation process from its beginning in early 1993 through the end of 1994.

## 1. INTRODUCTION

Since the late 1970s, the US Government has used the Linear Predictive Coder (LPC-10) for speech coding at 2400 bps [1]. This algorithm was officially accepted as Federal Standard 1015 in 1984 (now known as FIPS Pub. 137). In the mid 1980s, LPC-10e (an enhanced version) was introduced which improved the quality of LPC-10 while maintaining bitstream compatibility with Fed-Std 1015. The Federal Standard 1015 has been successfully implemented in a number of government communication systems such as STU-II, STU-III, ANDVT, MINTERM, as well as NATO systems. Recently the U.S. Government has recognized a need for a new and improved speech coder at this rate and is now conducting research to support a documented need, *Defense Information Systems Agency (DISA) Secure Voice Goal Architecture and Transition Strategy,* for a high quality 2400 bps speech compression algorithm [2]. One notable objective of the DISA Goal Architecture is to provide interoperability between traditionally "narrow-band" and "wide-band" systems. A new 2400 bps standard could bridge this gap, offering end-to-end digital connectivity between wireline, radio, satellite, and other tactical and strategic communications systems.

## 2. TERMS OF REFERENCE

The DDVPC developed a document containing the terms of reference defining the requirements of the new algorithm [3]. Quality, complexity, delay, and other characteristics are described with the required measurements and the objective measurements. Table 1 lists a subset of the parameters contained in the terms of reference document.

**Table 1: Abbreviated Terms of Reference**

| Parameter | Requirement (wrt FS1016) | Objective (wrt FS1016) |
|---|---|---|
| Error free & noise free environment | Equivalent/better: intelligibility and quality | Better: intelligibility and quality |
| Speech in office background noise | Equivalent/better: intelligibility and quality | Better: intelligibility and quality |
| 16 kbps CVSD -> Candidate | Equivalent/better: intelligibility and quality | Better: intelligibility and quality |
| Speech in other background noise, tandems, and channel errors | N/A | ~Equivalent or better: intelligibility and quality |
| One way delay | ≤ 180ms | ≤ 90ms |
| Processing Power | Coder in 1 DSP, ≤ 80MHz | Coder in 1 DSP... variable complexity |
| Memory | ≤ 4Mb | ≤ 2Mb |

## 3. TESTING

### 3.1 Test Schedule

In May 1993 the DDVPC held an informal workshop to discuss 2400 bps speech compression algorithms. Quality and intelligibility tests were performed that spring on four coders to measure their ability to meet the terms of reference. In February 1994 a second workshop was held, and a second informal test was conducted during that summer on eight coders. A subset of the results from the two tests are presented below. A third workshop was held in November of 1994, and the final test will take place during the fall of 1995. Selection of the algorithm to become the standard will be completed during the first part of 1996. The selection will be based on the results of the final test and the merits of each coder. A coder's merits depend upon the algorithm's memory usage and processor requirements, as well as the developer's property rights issues.

### 3.2 1993 Test

The 1993 test consisted of quality and intelligibility tests performed in five conditions. These conditions were chosen based on availability and relevance to the civilian and military services. In addition to a quiet environment, jeep and tank acoustic backgrounds were tested as well as a CVSD tandem and microphone shaping. A comparison of the results of tests conducted in both 1993 and 1994 are shown in Figures 2, 4, and 10.

### 3.3 1994 Test

The testing in 1994 was more extensive than in the previous year. The DDVPC's members selected a variety of conditions for testing that reflected the needs of all military and civilian branches. Several new conditions were recorded in modern environments to accommodate the requirements. In addition to the quiet condition, several acoustic background conditions were tested. These conditions primarily consisted of vehicle noise from equipment currently being used by the U.S. armed forces. The conditions and the tests that were performed in these environments are shown in Table 2.

The acoustic noise conditions are: modern office, high mobility multipurpose wheeled vehicle (HMMWV), M2 Bradley fighting vehicle, mobile command environment (MCE), Plymouth Reliant automobile, and CH47 helicopter. In addition to the acoustic noise environments, a tandem both ways with CVSD, and random bit errors at 2% were tested.

**Table 2: 1994 Test Conditions**

| Condition | MOS | DMOS | DAM | DRT | NTRT |
|---|---|---|---|---|---|
| Quiet | ✔ | | ✔ | ✔ | ✔ |
| Modern Office | ✔ | | | | |
| HMMWV | | ✔ | ✔ | ✔ | |
| M2 | | ✔ | | ✔ | |
| MCE | ✔ | | | | |
| Staff Car | | ✔ | | | |
| CH47 | | | | ✔ | |
| CVSD->Coder | ✔ | ✔ | ✔ | ✔ | |
| Coder->CVSD | ✔ | | ✔ | ✔ | |
| Quiet, 2% BER | | | ✔ | ✔ | |

Both quality and intelligibility tests were conducted. A test to measure how well a listener can recognize the talker, was also performed. The diagnostic acceptability measure (DAM), the mean opinion score (MOS), and the degradation mean opinion score (DMOS) were used to measure quality, and the diagnostic rhyme test (DRT) was used to measure intelligibility. The NRL Talker Recognizability Test (NTRT), currently being developed at the Naval Research Laboratory in Washington, DC, was used to measure speaker recognizability.

### 3.4 Results of 1993 and 1994 Tests

Each of the plots in Figures 1-12 show the quality and intelligibility scores of the candidate algorithms and FS1016 CELP in a different test condition. Development of the NTRT is not yet complete so speaker recognizability test scores are not available. In each figure, the "O"s show the scores for the conditions tested in 1993, and the "X"s show the scores for 1994 test. (Each set of scores from each individual condition has been sorted from lowest to highest prior to being plotted so that the order of the candidates changes for each figure, i.e. the first point on the plot does not necessarily correspond to the same candidate for each figure.) The "*" shows the score for the FS1016 CELP in that condition. The top line on the plot corresponds to the highest scoring candidate algorithm in the 1994 test and the bottom line corresponds to the lowest scoring candidate algorithm in the 1994 test. The middle line marks the CELP equivalency threshold. Any coder scoring above this line is considered statistically equivalent to CELP based on a Newman-Keuls analysis at 95% confidence. In Figure 5, the top 1994 score is the CELP equivalency threshold.

The 1994 test produced encouraging results showing that the requirements for the new coder can be met. As Table 1 shows, equivalency in quality and intelligibility with FS1016 CELP in quiet, office, and CVSD tandem is a requirement. Figures 1-4 show that several candidate algorithms meet or exceed the FS1016 CELP threshold in both quality and intelligibility in the quiet condition. (Note: Two DAM scores are shown due to testing with two different sets of talkers). Figure 6 shows that four coders meet or exceed the FS1016 CELP threshold in the office environment for quality, and figures 10-12 show that the CVSD->Coder requirement can be achieved. The candidate algorithms still require improvement in the harsher environments and errors, though, as Figures 5, 7, 8, and 9 illustrate.

### 3.5 1995 Test

The 1995 test will consist of two phases. Each algorithm will be tested using real time equipment for both phases. Candidate algorithms not achieving the requirements stated in the terms of reference, which will be part of the phase one testing, will not continue on to phase two. Phase 1 will consist of quality and intelligibility testing on a subset of the conditions. Phase 2 will also contain recognizability (described earlier) and communicability testing. The communicability test's measurements will be based on four different scenarios, one each from the Army, Air Force, Navy, and Department of Defense. The scenarios simulate a real-world situation where two communicators, connected by a transmission link, may or may not be experiencing the same conditions. This test is being developed by ARCON for RL/ERT at Hanscom AFB in Massachusetts.

In 1995, the following fifteen conditions will be tested:

1. Quiet with dynamic microphone
2. Quiet with H250 microphone
3. HMMWV
4. M2 Bradley fighting vehicle
5. CH47 Helicopter
6. Modern office
7. F-15 Eagle airplane
8. E3A AWACS airplane
9. P3C Orion airplane
10. MCE
11. Plymouth Reliant automobile
12. 1% bit errors
13. 2% block errors

14. CVSD->Coder
15. CVSD->Coder->CVSD.

In addition, all other requirements and objectives stated in the terms of reference will be either implicitly or explicitly tested.

## 4. CANDIDATE ALGORITHMS

Eight candidate algorithms were evaluated during the 1994 test. Four of the algorithms were variations of the multiband excitation algorithm, and two of the algorithms are variations of the mixed excitation linear prediction algorithm. A sinusoidal transform coder and a waveform interpolation coder are also current candidates.

### 4.1 Waveform Interpolation (WI)

The WI algorithm is based on the representation of the speech signal by two waveforms. One represents the slowly evolving characteristics of the input waveform while the other represents the rapidly evolving characteristics. Quantization of each of the waveforms is performed separately. Synthesis of the speech signal is performed by combining the two components and interpolating linearly. As the bit rate increases, the reconstructed speech signal converges to the original signal [4,5,6].

### 4.2 Sinusoidal Transform Coder (STC)

The STC algorithm uses a sinusoidal model with amplitudes, frequencies, and phases derived from a high resolution analysis of the short term Fourier transform. A harmonic set of frequencies is used as a replacement for the periodicity of the input speech. Pitch, voicing, and sine wave amplitudes are transmitted to the receiver. Conventional methods are used to code the pitch and voicing, and the sine wave amplitudes are coded by fitting a set of cepstral coefficients to an envelope of the amplitude [7].

### 4.3 Mixed Excitation Linear Prediction (MELP)

MELP is based on the traditional LPC model of exciting a pole filter with either a periodic impulse train to represent voiced speech or white noise to represent unvoiced speech. An improvement on the quality is achieved by using mixed pulse and noise excitation, periodic and aperiodic pulses, a pulse dispersion filter, and adaptive spectral enhancement [8].

### 4.4 Multiband Excitation (MBE)

The MBE model for speech assumes that both voiced and unvoiced excitation can exist at the same time in the same analysis frame but in different frequency bands. The speech spectrum is split into non-overlapping bands and each band is modeled as being either voiced or unvoiced. The voiced bands are synthesized using sinusoidal oscillators and the unvoiced bands are synthesized using bandpass filtered noise [9].

## 5. CONCLUSION

In general, scores have improved since this effort's commencement. The number of candidates has increased providing a broader algorithmic base and more competition. The viability of achieving the requirements stated in the Terms of Reference has been proven by the 1994 test, and it is hoped that continued algorithmic improvements prior to the 1995 test will provide even higher performance. Since the requirements of the new 2400 bps standard can be met, the chosen algorithm will provide even better communications than expected. In addition, the new standard will usher in a new era in secure communications for the U.S. Government by allowing end-to-end digital interoperability over almost all communications links in the traditional narrowband and wideband architectures.

## 6. ACKNOWLEDGEMENTS

The DDVPC membership consists of representatives from the following government organizations: Army JIEO, Ft. Monmouth, NJ; Naval Research Laboratory, Washington, DC; Rome Laboratories ERT, Hanscom AFB, MA; Department of Defense, Ft. Meade, MD; and Army CECOM, Ft. Monmouth, NJ.

## 7. REFERENCES

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

[2] "Secure Voice Goal Architecture and Transition Strategy," GTE Report No. 92-2479112-D, September 1992.

[3] M. Bielefeld, "Terms of Reference for the New Government Standard Voice Coder at 2400 bps", TL-8340-94-150, The MITRE Corporation, 1994.

[4] W. Bastiaan Kleijn, "Speech Coding Below 4KB/S using Waveform Interpolation", *Proceedings Globecom*, pp. 1879-1883, 1991.

[5] W.Bastiaan Kleijn and Jesper Haagen, "A General Waveform-Interpolation Structure for Speech Coding", *Proceedings EUSIPCO*, September 1994.

[6] W. Bastiaan Kleijn and Jesper Haagen, "Transformation and Decomposition of the Speech Signal for Coding", IEEE Signal Processing Letters, Vol. 1, No. 9, September 1994.

[7] R.J. McAulay & T.F. Quatieri, "The Application of Subband Coding to Improve Quality and Robustness of the Sinusoidal Transform Coder", *Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. II-439 - II-446, April 1993.

[8] A.V. McCree and T.P. Barnwell, "Implementation and Evaluation of a 2400 bps Mixed Excitation LPC Vocoder", *Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. II-159 - II-162, April 1993.

[9] Keith A. Teague, Bryce Leach, and Walter Andrews, "Development of a High-Quality MBE Based Vocoder for Implementation 2400 bps", *Proceedings of the IEEE Wichita Conference on Communications, Networking, and Signal Processing*, April, 1994.

[10] Vanoy Welch, Tina Kohler, Thomas Tremain, "The New Government Standard 2400 bps Speech Coder", *Proceedings of the 1994 DSPx Exposition and Symposium*, pp. 462-466, June 1994.

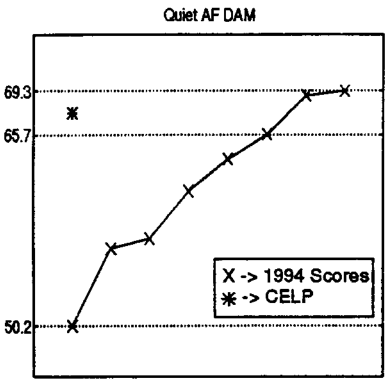

Figure 1

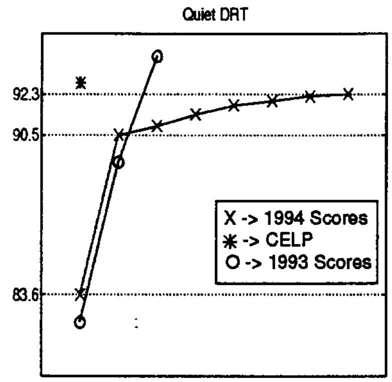

Figure 2

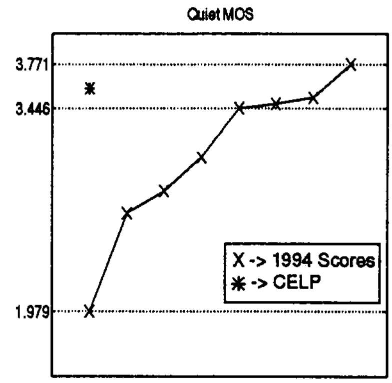

Figure 3

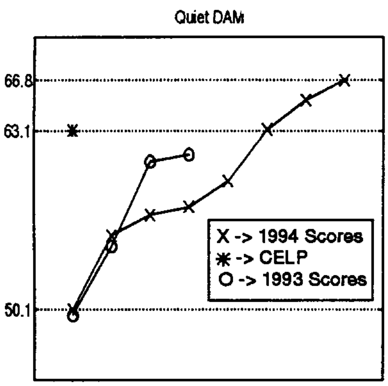

Figure 4

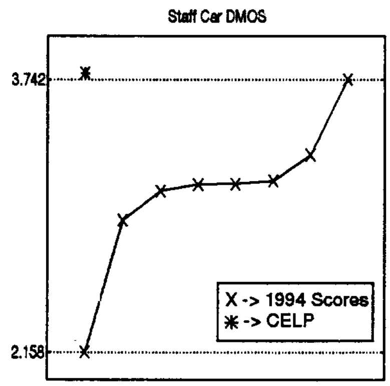

Figure 5

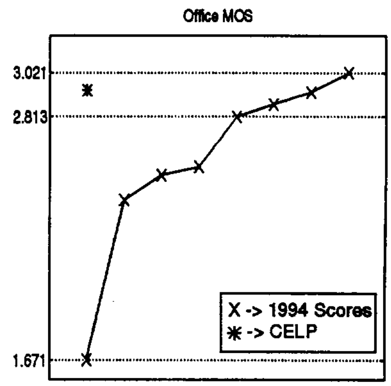

Figure 6

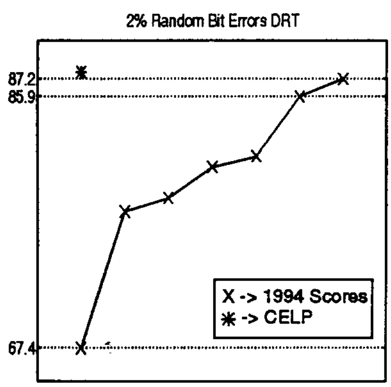

Figure 7

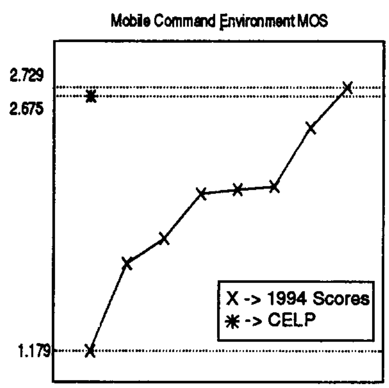

Figure 8

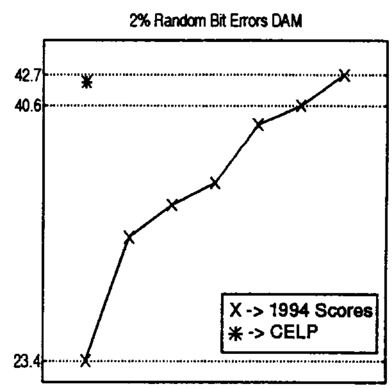

Figure 9

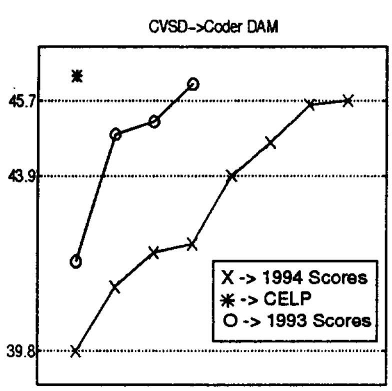

Figure 10

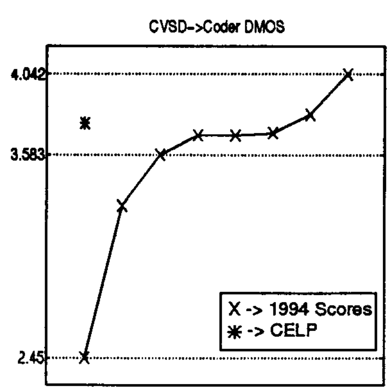

Figure 11

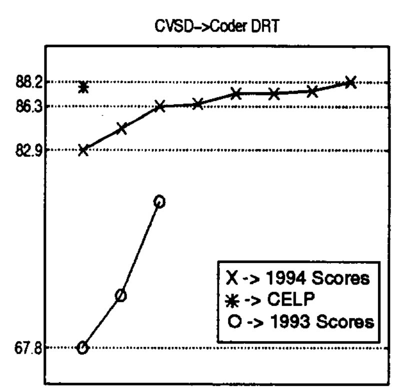

Figure 12