SOUTHERN METHODIST UNIVERSITY

# Voice Encoding Methods for Digital Wireless Communications Systems


BY

Bryan Douglas

Street address

city

state, zip

e-mail address

Student ID xxx-xx-xxxx


EE6302 Section 324, Fall 1997

# Table of Contents

**Abstract**

Wireless communications operators see phenomenal growth in consumer demand for high quality and low cost services. Since the physical spectrum for wireless services is limited, operators and equipment suppliers continually find ways to optimise bandwidth efficiency. Digital communications technology provides an efficiency advantage over analog wireless communications; multiplexing and filtering is easier, components are cheaper, encryption is more secure and network management is easier. Additionally, digital technology provides more value added services to customers (security, text and voice messages together, etc.).

Today wireless communication is primarily voice. The operator meets the increasing need for services by combining digital technology and special encoding techniques for voice. These encoders ("vocoders") take advantage of predictable elements in human speech. Several low data rate encoders are described here with an assessment of their subjective quality.

Test methods to determine voice quality are necessarily subjective. Different approaches attempt to reduce the variation in subjective testing, and a comparison of two methods resulted in governmental acceptance of the Mean Opinion Score, or MOS. The MOS is the most widely accepted test method.

The most efficient vocoders have acceptable quality levels and have data rates between 2 and 8 kbit/s. Higher data rate encoders (8-13 kbit/s) have improved quality while 32 kbit/s coders have excellent quality (but use more network resources. The operator must engineer the proper balance between cost, quality and available resources to provide the optimum solution to the customer.

**Background and Introduction to Encoding**

***Properties of Speech***

The two types of speech sounds, voiced and unvoiced, produce different sounds and spectra due to their differences in sound formation. With voiced speech, air pressure from the lungs forces normally closed vocal cords to open and vibrate. The vibrational frequencies (pitch) vary from about 50 to 400 Hz (depending on the person's age and sex) and forms resonance in the vocal track at odd harmonics. These resonance peaks are called formants and can be seen in the voiced speech figures 1 and 2 below [1].

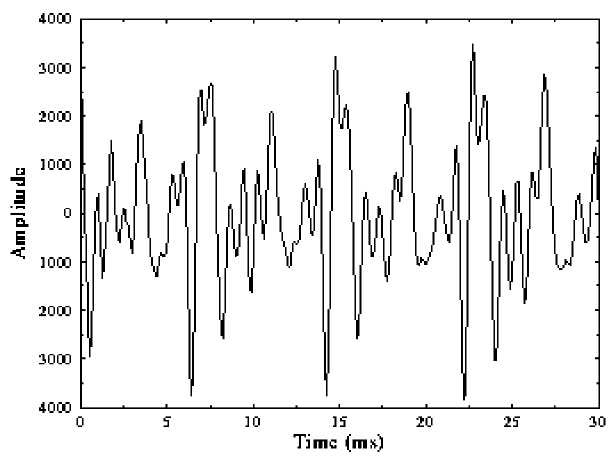


**Figure 1: Voiced Speech Sample [2]**

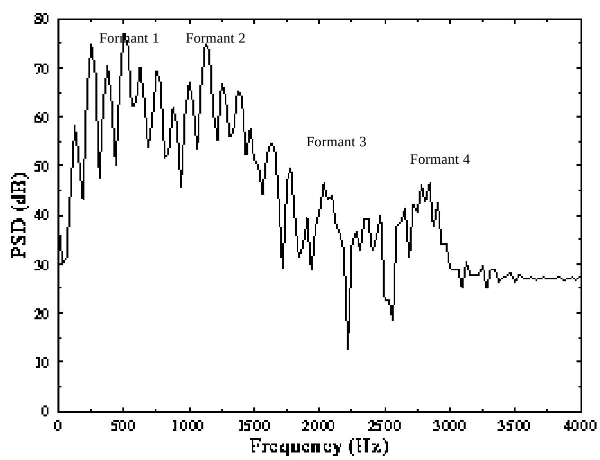


**Figure 2: Power Spectral Density, Voiced Speech [3]**

Unvoiced sounds, called fricatives (e.g., s, f, sh) are formed by forcing air through an opening (hence the term, derived from the word "friction"). Fricatives do not vibrate the vocal cords and therefore do not produce as much periodicity as seen in the formant structure in voiced speech; unvoiced sounds appear more noise-like (see figures 3 and 4 below). Time domain samples lose periodicity and the power spectral density does not display the clear resonant peaks that are found in voiced sounds

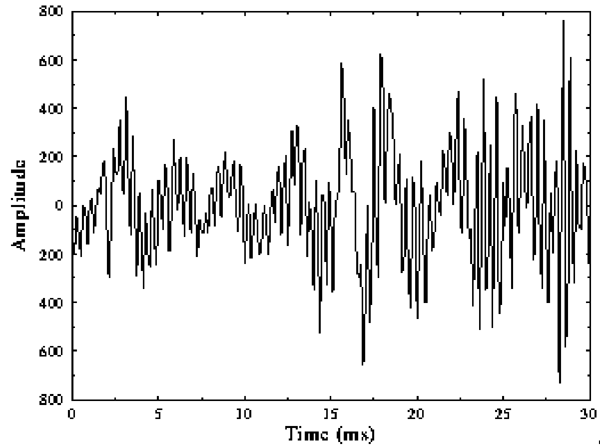


**Figure 3 Unvoiced Speech Sample [4]**

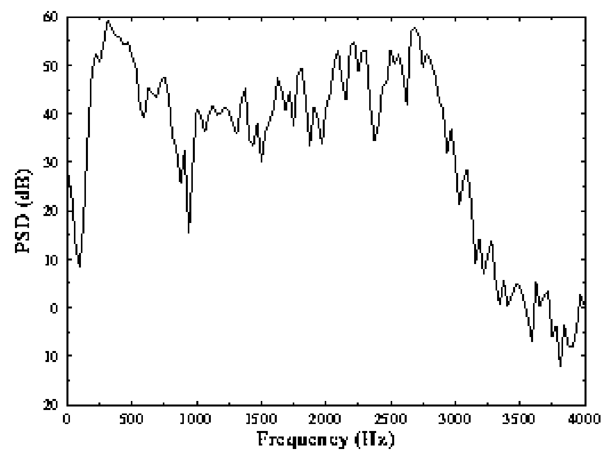


**Figure 4: Power Spectral Density, Unvoiced Speech [5]**

The spectrum for speech (combined voiced and unvoiced sounds) has a total bandwidth of approximately 7000 Hz with an average energy at about 3000 Hz. The auditory canal optimizes speech detection by acting as a resonant cavity at this average frequency. Note that the power of speech spectra and the periodic nature of formants drastically diminish above 3500 Hz. Speech encoding algorithms can be less complex than general encoding by concentrating (through filters) on this region. Furthermore, since line quality telecommunications employ filters that pass frequencies up to only 3000-4000 Hz, high frequencies produced by fricatives are removed. A caller will often have to spell or otherwise distinguish these sounds to be understood (e.g., “F as in Frank”).

### *General Encoding of Arbitrary Waveforms*

Waveform encoders typically use Time Domain or Frequency Domain coding and attempt to accurately reproduce the original signal. These general encoders do not assume any previous knowledge about the signal. The decoder output waveform is very similar to the signal input to the coder. Examples of these general encoders include Uniform Binary Coding for music Compact Disks and Pulse Code Modulation for telecommunications.

Pulse Code Modulation (PCM) is a general encoder used in standard voice grade circuits. The PCM encodes into eight bit words Pulse Amplitude Modulated (PAM) signals that have been samples at the Nyquist rate for the voice channel (8000 samples per second, or twice the channel bandwidth). The PCM signal therefore requires a 64 Kb/s transmission channel. However, this is not feasible over communication channels where bandwidth is a premium. It is also inefficient when the communication is primarily voice that exhibits a certain amount of predictability as seen in the periodic structure from formants. The increasing use of limited transmission media such as radio and satellite links and limited voice storage resources require more efficient coding methods. Special encoders have been designed that assume the input signal is voice only. These vocoders use speech production models to reproduce only the intelligible quality of the original signal waveform. The most popular vocoders used in digital communications are presented below.

### Types of Voice Encoders

The channel vocoder uses a bank of filters or digital signal processors to divide the signal into several sub-bands. After rectification the signal envelope is detected with bandpass filters, sampled, and transmitted. (The power levels are transmitted together with a signal that represents a model of the vocal tract.) Reception is basically the same process in reverse. These vocoders typically operate between 1 and 2 kbit/s. Even though these coders are efficient, they produce a synthetic quality and therefore are not generally used in commercial systems.

Since speech signal information is primarily contained in the formants, a vocoder that can predict the position and bandwidths of the formants could achieve high quality at very low bit rates. A formant vocoder transmits the location and amplitude of the spectral peaks (see figure 2) instead of the entire spectrum. These typically operate in the range of 1000 bit/s. Formant vocoders are not very popular because the formants are difficult to predict.

### *Linear Predictive Encoder (LPC)*

Linear Predictive Encoders are the most popular today and are used mainly in digital Personal Communications Services. The LPC algorithm assumes that each speech sample is a linear combination of previous samples. Speech is sampled, stored and analysed. Coefficients, calculated from the sample are transmitted and processed in the receiver. With long term correlation from samples, the receiver accurately processes and categorises voiced and unvoiced sounds. The LPC family use pulses from an excitation pulse generator to drive filters whose coefficients are set to match the speech sample. The excitation pulse generator differentiates the various types of LP coders discussed below [6]. LP filters are simple to implement and simulate filtering and acoustic pulses produced in the mouth and throat. An LPC coder is shown in figure 5.

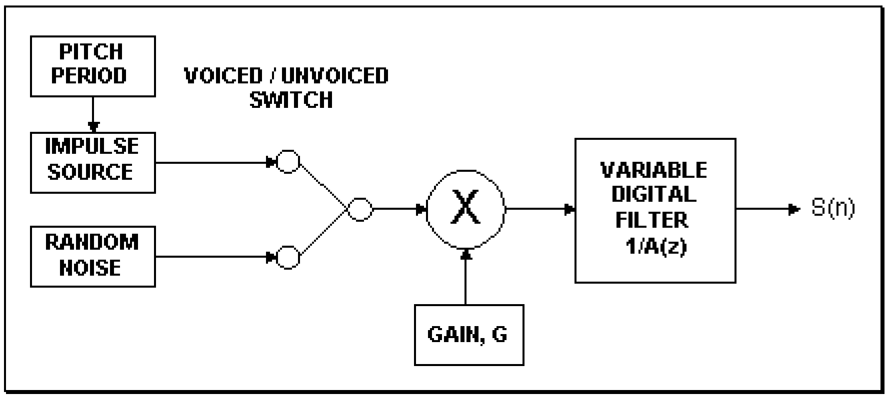


**Figure 5: Linear Predictive Model [7]**

***Regular Pulse Excited (RPE) Coder***

The RPE analyses the signal to determine if it is voiced or unvoiced. After determining the period for voiced sounds, the periodicity is encoded and the coefficient is transmitted. When the signal changes from voiced to unvoiced, a code is transmitted that stops the receiver from generating periodic pulses and starts generating random pulses to correspond to the noise like nature of fricatives. The RPE is used in the GSM full rate vocoder (See figure 6) and its U.S. version, PCS 1900.

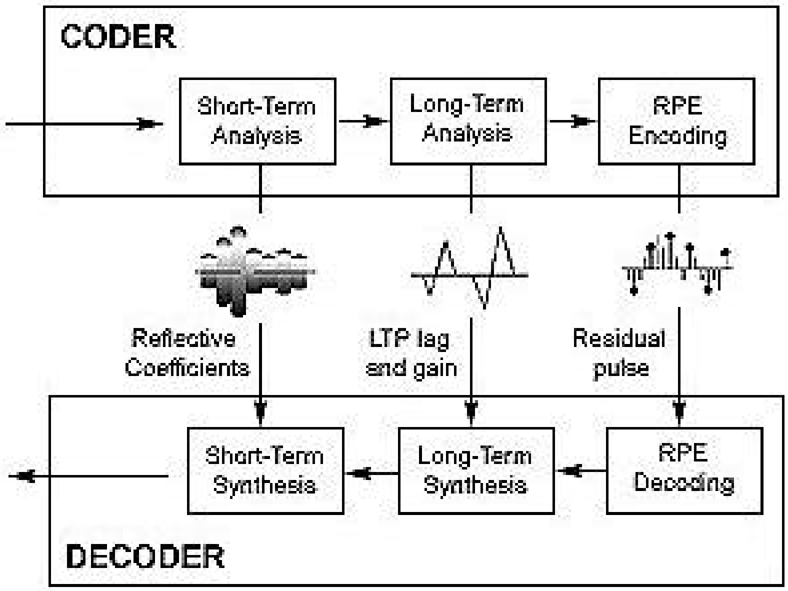


**Figure 6: GSM RPE Coder and Decoder [8]**

The GSM vocoder on the network side is in the Transcoder and Rate Adapter Unit (TRAU) [9] which transcodes data from 16 kbit/s to 64 kbit/s. Phase one of the GSM specification defines full rate coders; phase two improves capacity by supporting half rate CELP coders at comparable quality (but requires more processing capability).

### *Code Book Excited (CELP)*

This coder is optimised by using a code book (look up table) to find the best match for the signal. This method reduces the processing complexity and the required data transmission rate. Most digital cellular systems use CELP (or CELP based) coders. Improved CELP models include the Vector Sum Excited (VSELP) and the Algebraic Code Excited (ACELP).

The VSELP, used in D-AMPS (North American digital cellular, IS-54), GSM half rate and PDC (Japan), simplifies the code book arrangement so that frequently occurring speech combinations are organised close together. ACELP coders do not require fixed code books at both the transmitter and receiver but optimise codes by using a series of nested loops [10]. These coders are used in the GSM Enhanced Full Rate (EFR), D-AMPS Enhanced and U.S. PCS 1900 EFR systems.

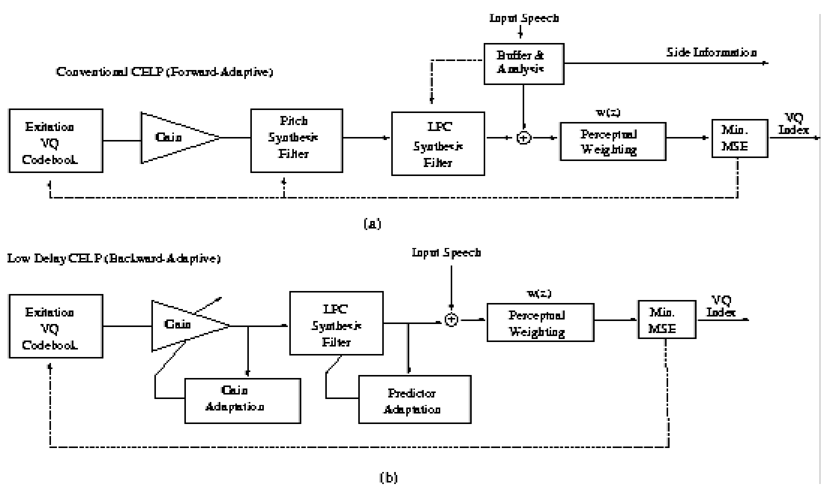


**Figure 7: "(a) Forward (b) Backward adaptive system for CELP" [11]**

### Vocoder Quality Measurements

Bellamy [12] presents several points to rate vocoder quality:

1. Cost/complexity
2. Voice Quality
3. Data Rate
4. Transparency for non-voice signals
5. Tolerance of transmission errors
6. Effects of tandem encodings
7. Coding formats
8. Signal processing requirements.

Cole, et.al. [13] suggest that the most important quality measures are voice quality, data rate, communication delay and coding algorithm complexity. While all of these can easily be measured and analysed, voice quality remains subjective. Several tests accepted by the industry adequately measure voice quality, but these all depend on trained listeners and results can vary depending on the speaker, hardware platform, listening groups and test data. Consequently data from one test group should not be compared with that from another. The Dynamic Rhyme Test (DRT), Diagnostic Acceptability Measure (DAM) and Mean Opinion Score (MOS) are commonly used tests.

The DRT is an ANSI (S32.2-1989) speech intelligibility test that provides a score of communications systems analysing the listener's ability to distinguish between 96 pairs of monosyllable rhyming words [14]. The DAM and MOS are speech acceptability (quality) measures [15].

#### *Comparison of Quality Measures*

The United States Department of Defense (DoD) Digital Voice Processor Consortium (DDVPC) used the DRT and MOS to evaluate 2400 bit/s coders based on test methods used in the industry (see table 1 below) [16].

| Sponsor | Year | System Selecting For | Test Used |
|---|---|---|---|
| DDVPC | 1989 | Improved 2400 bit/s Coder | DRT\DAM |
| | 1990 | 4800 bit/s Federal Std. 1016 | DRT\DAM\T-files |
| APCO | 1992 | Digital Land Mobile Radio | MOS\DMOS |
| TIA 45.3 | 1991 | Digital Full Rate Selection (IS-54) | MOS |
| | 1992 | Half Rate Coder Technology Evaluation | MOS |
| | 1993 | Half Rate Digital Cellular Selection | MOS |
| TIA 45.5 | 1992 | Wideband Spread Spectrum Cellular | MOS |
| Immarsat/Ausat | 1990 | Marine Satellite Voice Coder | MOS |
| CCITT | 1991 | European GSM | MOS |

**Table 1: Evaluation Methods in Digital Speech Communication Systems [17]**

Traditionally the DoD used the DAM as a quality measure, but in this evaluation they analysed and compared the DAM and MOS. Therefore the test objective was twofold: to evaluate 2400 bit/s coders and to compare the MOS to the DAM. While the DoD goals and environment differ from commercial digital communications (noise level in jeeps, etc.), the study concludes that the MOS is a cost effective and accurate (but subjective) measure of voice quality [18].

The first experiment compared MOS to MOS (to establish a control) using several different coders. There were 40 listeners (20 male, 20 female), who were not familiar with the technology. Two separate and independent laboratories conducted the MOS experiments using the standard MOS score:

| Score: | Rating: |
|---|---|
| 5 | Excellent |
| 4 | Good |
| 3 | Fair |
| 2 | Poor |
| 1 | Unsatisfactory |

**Table 2: Mean Opinion Score Quality Ratings**

The independent labs confirmed the validity of the MOS and determined that the results would not exclude this test from further analysis. Next the research compared the MOS to the DAM. Most of the coders tested produced similar results for both the MOS and the DAM; both tests have similar resolution in determining quality between coders. However, the rank order differed between some of the coders. For example, the MOS and the DAM reversed the rank of the STC48 and the CELP coders. The MOS evaluation agreed with an independent test (APCO-25) [19]. Therefore the researchers concluded that the MOS test was the preferred method. The coders tested and their average MOS scores are tabulated below. The results presented here are from experiment 1 only since experiment 2 was determined to be statistically equivalent. Furthermore, the results from noisy environments (inside military vehicles) are not presented since they would not give relevance to commercial vocoder implementation.  Note that the CSVD (both the 32 and 16 kbit/s versions) scored unexpectedly poorly; these were discounted because a faulty filter was detected in the processor.  Therefore the CVSD results here should not be used for any analysis.

| | | MOS Test Results | | | | | DAM Mean CAE Scores | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | MPR Test Lab | | COMSAT Test Lab | | | | |
| **Coder** | **Rate:** | **Rank:** | **MOS:** | **s.e.** | **MOS:** | **s.e.** | **Rank:** | **CAE:** | **s.d.** |
| ADPCM | 32 kbit/s | 1 | 3.59 | 0.05 | 3.90 | 0.06 | 1 | 69.0 | 1.12 |
| VSELP | 8 kbit/s | 2 | 3.16 | 0.06 | 3.53 | 0.06 | 2 | 66.5 | 1.28 |
| CELP | 4.8 kbit/s | 5 | 2.70 | 0.06 | 2.98 | 0.06 | 3 | 59.9 | 1.19 |
| LPC-10e | 2.4 kbit/s | 10 | 1.98 | 0.05 | 2.13 | 0.05 | 10 | 49.4 | 1.10 |
| STC48 | 4.8 kbit/s | 3 | 2.83 | 0.06 | 3.08 | 0.06 | 5 | 56.2 | 1.97 |
| MBE48 | 4.8 kbit/s | 4 | 2.76 | 0.06 | 2.99 | 0.06 | 4 | 57.8 | 1.16 |
| CVSD32 | 32 kbit/s | 8 | 2.45 | 0.05 | 2.59 | 0.05 | 9 | 50.0 | 1.20 |
| CVSD16 | 16 kbit/s | 11 | 1.90 | 0.04 | 1.93 | 0.04 | 11 | 43.2 | 0.71 |
| STC24A | 2.4 kbit/s | 7 | 2.48 | 0.05 | 2.59 | 0.05 | 7 | 51.9 | 1.48 |
| STC24B | 2.4 kbit/s | 6 | 2.57 | 0.06 | 2.72 | 0.06 | 6 | 53.2 | 1.37 |
| MBE24 | 2.4 kbit/s | 9 | 2.15 | 0.05 | 2.34 | 0.05 | 8 | 51.0 | 1.36 |

**Table 3: DDVDC Test Results [20]**

***Vocoder Comparison***

The results described above compare vocoder performance and test methods for quality measurement.  The International Telecommunications Union, Radio Sector (ITU-R) has also tested vocoders and reports a proposal for test methods results in reference [21].  The objective for the ITU-R study group was to provide a method for subjective comparison of vocoder quality.  Additionally the ITU-R has performed tests on vocoders (using the MOS) to compare quality of commercial systems that are already deployed.  The results (see table 4) from the ITU rate vocoder performance by providing MOS scores together with a measure of Forward Error Correction (FEC), coder delay and processing requirements (platform dependent).  The working group rated systems used for several different digital communications systems.  Specifically, they evaluated digital cellular (TDMA and CDMA), digital cordless used in Europe and Japan, and dispatch systems (for government and emergency operations) in North America, Europe and Japan.  The digital cellular systems all have acceptable quality with MOS scores between 3.3 and 4.1.  Digital cordless systems are rated highly at 4.0 and dispatch systems are acceptable at 3.2 to 3.98.

| System Type: | System Name: | Ref.: | Codec Type: | Codec Rate (kbit/s): | Forward Error Correction (kbit/s): | Codec Algorithmic Delay (mS): | Est. Quality (MOS): | Est. Processing: |
|---|---|---|---|---|---|---|---|---|
| Cellular TDMA | GSM/DCS/PCS Full Rate | ETSI/ETS 300580 ANSI J-STD-007 | RPE-LTP | 13 | 9.8 | 40 | 3.6-3.8 | 2.5 Mips |
| Cellular TDMA | GSM/DCS/PCS Half Rate | ETSI/ETS 300581 | VSELP | 5.6 | 5.8 | 40 | 3.5-3.7 | 17.5 Mips |
| Cellular TDMA | GSM EFR US PCS1900 EFR | ETSI/ETS 300723 ANSI J-STD-007A | ACELP | 12.2 | 10.6 | 40 | 4.1 | 15.4 WMops |
| Cellular TDMA | D-AMPS Full Rate | TIA/EIA IS-85 | VSELP | 8.0 | 5.0 | 28 | 3.7 | 22 Wmops |
| Cellular TDMA | D-AMPS Enhanced | TIA/EIA IS-641 | ACELP | 7.4 | 5.6 | 25 | 4.1 | 14 Wmops |
| Cellular TDMA | PDC Full Rate | RCR-STD-27 | VSELP | 6.7 | 4.5 | Unavailable | 3.40 | 7.8 Mops |
| Cellular TDMA | PDC Half Rate | RCR-STD-27 | PSI-CELP | 3.45 | 2.15 | Unavailable | 3.34 | 18.7 Mops |
| | | | | | | | | |
| Cellular CDMA | Composite CDMA/TDMA | TIA/EIA IS-661 | CELP-Like | 7.2 | 3.2 | 26 | >4.0 | 11 Mips |
| Cellular CDMA | CDMA | TIA/EIA IS-96 | CELP | 8/4/2/0.8 | 19.2 | 27.5 | 3.3 | 22 Mops |
| Cellular CDMA | CDMA | TIA/EIA IS-127 | RCELP | 8/4/0.8 | 19.2 | 30 | 4.1 | 20 Mops |
| Cellular CDMA | W-CDMA | USA Unavailable | ADPCM | 32 | 0 | Unavailable | Unavailable | Unavailable |
| | | | | | | | | |
| Digital Cordless | CT2 | ITU-T-G.726 | ADPCM | 32 | 0 | 0.25 | 4.0 | 10 Mips |
| Digital Cordless | DECT | ITU-T-G.726 | ADPCM | 32 | 0 | 0.25 | 4.0 | 10 Mips |
| Digital Cordless | PHS | ITU-T-G.726 | ADPCM | 32 | 0 | Unavailable | Unavailable | 1.0 Mops |
| | | | | | | | | |
| Dispatch | Project 25 U.S. | EIA/TIA IS-102.BABA | IMBE | 4.4 | 2.8 | 80 | 3.4 | 6.9 Mips |
| Dispatch | TETRA Europe | ETSI/ETS 300395 | ACELP | 4.567 | 2.633 | Unavailable | 3.3-3.5 | 15 Mips |
| Dispatch | IDRA Japan | RCR STD-32A | CSELP | 4.72 | 2.766 | Unavailable | 3.2 | 7.0 Mops |
| Dispatch | IDRA Japan | RCR STD-32A | VSELP | 4.2 | 3.177 | Unavailable | 3.20/3.98 | 8.0 Mops |
| Dispatch | DIMRS Canada | Unavailable | VSELP | 4.2/8.8 | 3.177/6.756 | | | 8.0 Mops |

**Table 4: Speech Coding Systems Performance Comparison [22]**

## Conclusions

Even though the ITU reports MOS scores for many of the same vocoders tested by the DDVPC, the results should not be directly compared because different listening groups and test samples were used. However, both of these groups have extensively analysed vocoder quality and can contribute to an overall assessment for determination of the suitable vocoder to meet a system needs.

Several standards have been approved by various organisations that provide acceptable quality at low data rates. The system engineer must a balance between overall needs by considering the end customer, the operator, and the manufacturer. The customer may be interested in vocoder quality, the operator in coder delay and in accepted standards (for interoperability), and the manufacturer in processing requirements (and their cost impact). Vocoder quality measurements presented here provide one component in the selection of a digital communications system; other dimensions should also be investigated. The total costs of ownership including capacity, physical size, support, power requirements and protection are only a few of the parameters to consider in detail.

## Works Cited


[1] McElroy, Ciarán. "Speech Coding". http://wwwdsp.ucd.ie/speech/tutorial/speech_coding/speech_tut.html (October 1997).

[2] ibid.

[3] ibid.

[4] ibid.

[5] ibid.

[6] Levine, R. Class Notes. Southern Methodist University, EE6302 "Digital Telephony (November 1997).

[7] Swanchara, Steve and Nguyen, Truong. "The Technological and Economical Ramifications of Vocoders in Cellular and PCS". http://fiddle.ee.vt.edu/courses/ee4984/Projects1996/nguyen_swanchara/nguyen_swanchara.html (May 1996, last visited October 1997).

[8] Degener, J. "Digital Speech Compression: Putting the GSM 0.610 RPE-LTP algorithm to work". http://www.ddj.com/1994/1994.12/degener.htm (December 1994, last visited February 1998).

[9] Kuruppillai, Rajan, Mahi Dontamsette and Fil J. Cosentio. *Wireless PCS*. New York: McGraw-Hill, 1997.

[10] Woodard, J and Hanzo, L. "A Channel-adaptive Speech Transceiver", http://www.ecs.soton.ac.uk/research/rj/comms/woodard/rj944.html (last visited March 1998).

[11] Haritaoglu, Esin Darici. "Wideband Speech and Audio Coding", http://wwwumiacs.umd.edu/~desin/Speech1/new.html (last visited December 1997)

[12] Bellamy, John. *Digital Telephony* 2nd Edition. United States of America: John Wiley & Sons, 1991.

[13] Cole, Ronald A, et.al. editors. *Survey of the State of the Art in Human Language Technology*. http://www.cse.ogi.edu/CSLU/HLTsurvey (November 1995, last visited November 1997). Reference from chapter 10.1, written by Isabel Trancoso, and 10.2, written by Bishnu S. Atal and Nikil S. Jayant.

[14] Dynastat Corporation. "Summary of Speech Intelligibility Testing Methods". http://www.realtime.net/dynastat/intellig.html (last visited January 1998)

[15] Dynastat Corporation. "Summary of Speech Quality Testing Methods" http://www.realtime.net/dynastat/quality.html (last visited January 1998)

[16] United States Department of Defense. "A New Federal Standard Algorithm for 2400 bit/s Coded Voice", presented at the IEEE International Conference on Acoustics, Speech & Signal Processing (ICASSP-96), http://www.infosec.rl.af.mil/ddvpc/24results.html (1996, last visited October 1997).

[17] ibid.

[18] Arcon Corporation report in preparation for Rome Laboratories, US Air Force Systems Command. "A Systematic Investigation of the Mean Opinion Score (MOS) and the Diagnostic Acceptability Measure (DAM) for Use in the Selection of Digital Speech Compression Algorithms" http://www.arcon.com/dsl/dldoc1.html (1997, last visited February 1998).

[19] ibid.

[20] ibid.

[21] ITU-R. "Method For Objective Measurements Of Perceived Audio Quality", Preliminary Draft Recommendation ITU-R BS.PEAQ, Document 10-4/19-E, March 1998, ITU-R Task Group 10/4.

[22] ITU-R. "Digitally Coded Speech In The Land Mobile Service", Draft New Recommendation ITU-R M. [8A/XD]' Document 8/65-E, January 1998, ITU-R Working Party 8A.