

[HOME](#)[CORPORATE OVERVIEW](#)[EVALUATION RESULTS](#)[SPEECH SAMPLES](#)[PRODUCTS & SERVICES](#)[PRODUCT LITERATURE](#)[DSP Chips](#)  
[Vocoder Boards](#)  
[Voice Codec Units](#)  
[Application Notes](#)  
[Papers](#)

PRODUCT LITERATURE

## Voice Coding Overview

The last ten years has witnessed explosive growth in the field of digital communication. One of the principal technologies enabling this growth is voice coding, in which an analog speech signal from a microphone is digitally sampled via an A-to-D converter and then efficiently compressed into a digital bit stream for transmission or storage. A corresponding voice decoder receives this bit stream and decompresses it back into a series of digital speech samples suitable for playback through a D-to-A converter and a loudspeaker.

Voice coders can take a number of different forms each of which involves tradeoffs in terms of bit rate (i.e. degree of compression), complexity (i.e. MIPS and memory) and voice quality as well as robustness. This article provides an introduction to voice coding and discusses the features and capabilities of several common voice coding algorithms. In addition this article discusses several important applications of voice coding and introduces methods for selecting the best vocoder for a given application.

Voice coders are normally divided into two broad classes referred to as waveform coders and model based speech coders. In a waveform coder the objective is to reproduce at the decoder the original speech samples on a sample by sample basis. A simple PCM waveform coder accomplishes this by quantizing each speech sample to one of a fixed number of levels. Assuming 8 bits (256 levels) are used per sample and the signal is sampled at 8 kHz the overall data rate is 64 kbps. More involved ADPCM waveform coders apply prediction with differential quantization to reduce the data rate to 24-32 kbps. In any case the process of quantizing the speech samples adds quantization noise which is usually audible as distortion in the decoded signal. The primary advantage of waveform coders is that if the data rate is kept sufficiently high the amount of distortion can be kept reasonably low. Hence waveform coders have historically been prevalent at rates over 16 kbps. Another advantage of traditionally waveform coders is that their complexity is typically modest allowing them to be more readily implemented on early DSP devices.

In contrast to waveform coders, model based speech coders or vocoders use a parametric model to approximate short (10-40 ms) segments of speech. For each segment a set of model parameters are estimated and converted into a bit stream. The decoder converts this bit stream back into model parameters and then uses these parameters to synthesize a speech signal which is perceptually close to the original. In this approach no attempt is made to recreate the original speech samples, instead only the perceptual content as approximated by the model parameters is maintained. The use of a parametric model allows vocoders to operate at lower data

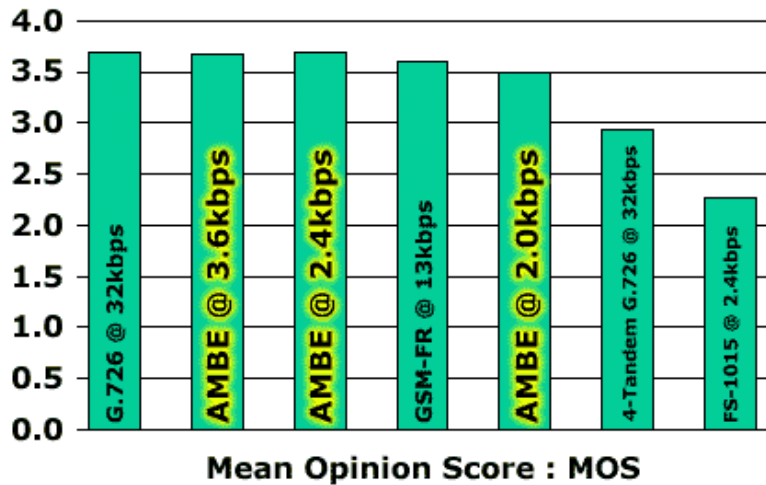
rates (under 8 kbps) than waveform coders, however they require an accurate speech model to obtain good performance. Early vocoders such as the channel vocoder, homomorphic vocoder and LPC vocoder all demonstrated the ability to produce intelligible speech at low to medium data rates. A good example is the 2400 bps LPC-10 vocoder used as a U.S. government standard for secure (i.e. encrypted) telephony.

Over the last decade continued work has improved the performance of voice coders. The challenge with waveform coders has been to try to maintain adequate voice quality while lowering the bit rate. Generally this effort has focused on techniques commonly referred to as CELP (Code Excited Linear Prediction) in which vector quantization is combined with adaptive linear prediction [1]. This approach borrows several concepts from model-based coders in that an all-pole model is used to approximate the speech spectrum and a long-term predictor is used to represent the pitch (i.e. local periodicity) of the speech signal. However in a CELP coder an error signal or residual is computed to compensate for the shortcomings of the linear predictive model. This residual is quantized using vector quantization which typically requires a search for the best vector from a large codebook of candidates. While this approach has made headway, yielding good algorithms at 8 kbps, the performance generally degrades rapidly at lower rates. In addition the vector search employed in CELP coders (and its many variants) has significantly increased the complexity of these algorithms to as high as 20-50 MIPS.

In contrast to waveform coders, the challenge in model-based coders has been to improve the speech model to allow higher voice quality at low bit rates. One approach which has made a significant contribution is the Multi-Band Excitation (MBE) speech model [2]. In this model, which is fundamentally different from the linear-predictive methods found in traditional vocoders as well as CELP, speech is modeled with a fundamental frequency, a set of spectral coefficients and a set of frequency dependent voicing decisions. The inclusion of multi-band voicing information plus new algorithms to analyze and synthesize speech has resulted in new MBE-based vocoders, such as the IMBE<sup>TM</sup> and AMBE<sup>®</sup> vocoders, which can provide very high quality speech at rates between 2-5 kbps. Achievement of high quality speech at such low bit rates is facilitated by the lack of any residual signal in the MBE-based approach. Instead increased emphasis is placed on high fidelity estimation and quantization of the model parameters so that voice quality can be maintained without the need for such an error signal.

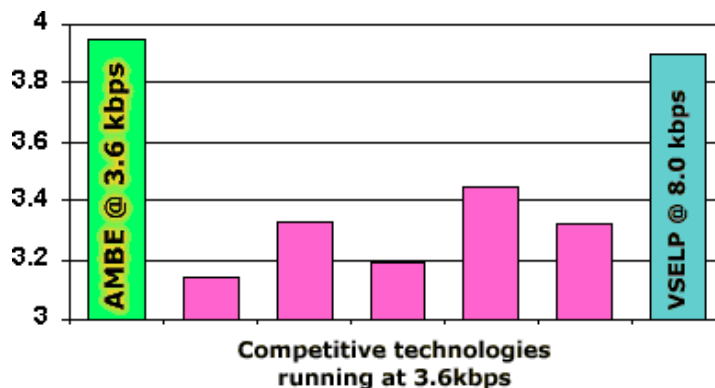
This development of advanced model-based coders has had a significant impact in a number of fields, including wireless communications, voice storage and digital telephony, which require high quality combined with efficient bandwidth utilization (i.e. low bit rate). Figure 1 provides an example of the voice quality, as measured by a standard mean opinion score (MOS) test, which is obtainable by DVSII's state of the art AMBE<sup>®</sup> vocoder. One can see from this figure that the model based AMBE<sup>®</sup> vocoder operating at 2 kbps is able to achieve better quality than the original GSM cellular waveform coder operating at 13 kbps, a factor of 6.5 times higher. In a wireless system this added coding efficiency is typically used to increase the number of users which can be supported across a fixed bandwidth. In addition lowering the data rate in a wireless application generally leads to smaller, less expensive mobile equipment which uses less power and, with additional forward error control (FEC), is more robust

to bit errors found in a typical mobile environment.



**Figure 1: Voice Quality of AMBE® and other Voice Coders**

When selecting a voice coder it is usually necessary to consider a number of application specific factors. The first issue is generally the data rate which is often determined by the available bandwidth or required storage capacity. A second issue is complexity which determines the MIPS and memory requirements of the actual implementation. In most applications lowering complexity usually results in lower hardware costs and power consumption as well as fewer components. A third factor is voice quality which in many cases is the overriding factor determining user acceptance and hence demand. Since voice quality is subjective, it is usually measured via an MOS test in which a group of normal listeners are asked their opinion on a 5 point scale (1= bad, 5=excellent) and the results are averaged together. Typically voice quality can be examined as it varies with speaker, input level, acoustic noise (such as in a car or office) and bit errors, where the exact conditions are representative of the intended application. The consideration of these different factors determine the robustness of the voice coder to the expected operating conditions.



**Figure 2: Inmarsat Mini-M evaluation of Voice Coder Quality for Clean Speech**

For illustration purposes one can consider the selection process conducted by Inmarsat for a new 4800 bps voice coder to be used in their Mini-M mobile satellite telephony system. In this case a gross bit rate of 4800 bps was specified in order to provide efficient global service via notebook sized

mobile terminals. Cost effective implementation favored an algorithm which could be implemented full-duplex in a single DSP using less than 20 MIPs. The voice quality was targeted to match then current cellular quality across various noise conditions. In addition the mobile satellite environment required tolerance to channels with a 1% to 4% bit error rate (BER). In order to evaluate the best voice coder for this application Inmarsat in conjunction with Comsat Laboratories performed an MOS test on six 4800 bps voice coders under a variety of conditions [3], comparing both waveform and model based coders. The results in Figure 2 show that the AMBE® vocoder substantially outperformed all of the other 4800 bps vocoder and had performance approximately equal to the 8 kbps VSELP coder used in IS-54 digital cellular. Further results, presented in Figure 3, showed that the AMBE® vocoder offers good performance across noise conditions actually outperforming the higher rate VSELP coder. Combining these results with the AMBE® vocoder's reasonable 13 MIP complexity led Inmarsat to select the AMBE® vocoder for the Mini-M system. Today this system is in the field providing instant communications from virtually any point on the globe.

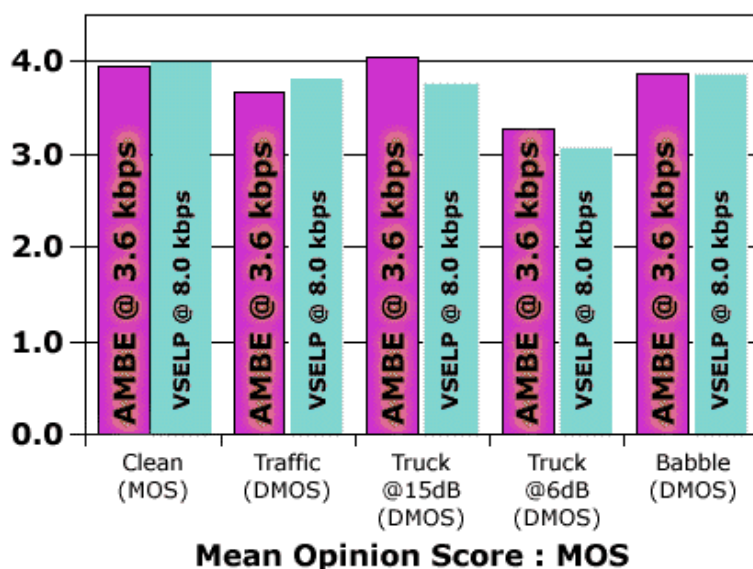


Figure 3: Inmarsat Mini-M evaluation of Voice Coder Noise Robustness

### Summary

In summary a voice coding has essential role in the current expansion of digital communications products and services. Numerous types of voice coders exist with the two principal divisions being waveform coder and model based coders. Waveform coders operating at higher rates offer relatively simple solutions with good performance, while at low rates advanced model based systems such as DVSII's proprietary AMBE® vocoder provide high quality and reasonable complexity. Selecting the best voice coder is very application dependent however a comprehensive evaluation considering complexity, voice quality and robustness is normally the route to the best solution.

[1] B. Atal and M. Schroeder, "Stochastic Coding Of Speech at Very Low Rates", Proceedings of ICC, 1984, pp. 1610-1613.

[2] D. W. Griffin and J. S. Lim, "Multiband Excitation Vocoder", IEEE Transactions on Acoustics, Speech and Signal Processing, Vol 36, No 8,

August 1988, pp. 1223-1235.

[3] S. Dimolitsas, et. al., "*Evaluation of Voice Codec Performance for the Inmarsat Mini-M System*", Tenth International Conference of Digital Satellite Communications, 15-19 May, 1995, pp. 101-105.

Digital Voice Systems, Inc. 234 Littleton Road Westford, MA 01886 USA  
Tel: 978 392 0002 Fax: 978 392 8866 [info@dvsinc.com](mailto:info@dvsinc.com)

Copyright © 2006 Digital Voice Systems, Inc. All Rights Reserved.