ABSTRACT

A data stream from an embedded speech coder has one or more slower data streams included. If part of the high rate data stream is discarded, the slower remaining data can be used to produce a decoded speech signal. Several embedded schemes have been developed for various rate ranges. The system described in this report covers the range from 16 to 64 kbps. At the lowest rate the coder produces a 16 kbps CVSD (continuously variable slope delta modulation) data stream compatible with standard CVSD equipment. At 32, 48 and 64 kbps the coder includes a 2, 4 or 6 bit encoding of the CVSD difference signal at an 8 kHz sample rate. At these higher rates the backbone CVSD decoder output is augmented by the 2, 4 or 6 bit difference to produce an 8 kHz sample output.

INTRODUCTION

In many digital voice communication systems it is valuable to have at hand a digitizer which can operate at several data rates. A choice of a particular rate for reasons of jamming, traffic or design implies an associated quality. In some applications the choice of rate and quality can be dynamically commanded [1]. In still other applications the transmitting encoder may deliver its highest rate encoded speech to a network interface, but somewhere in the path to a destination the data stream is restricted to a lower rate [1]. This kind of operation assumes the higher rate signal contains a low rate signal which can be selected out and used at a receiver. This requires the variable rate encoder to operate as a so called "embedded coder" implying the high rate data stream has this attribute. Although many devices can be realized that will encode speech in different ways and produce a different rate for each, a device producing an embedded stream is more constrained. For example, a delta modulation encoder can produce a data rate equal to the clock frequency. However, if the encoder is producing a 64 kbps data stream and we strip out alternate data bits to drive the receiver-decoder with 32 kbps, the output speech will be inferior to that produced by a 32 kbps encoder driving this decoder. In this sense the 64 kbps encoder is not producing an embedded output. On the other hand a properly quantized PCM data stream can have least significant bits stripped away and the received signal will be the same as a PCM encoder-decoder of the lower rate.

The enhanced CVSD (continuously variable slope delta modulation) encoder-decoder [2] described in this paper is capable of producing an embedded data stream at 64 kbps which contains besides the full rate, useful encoding at 48, 32 and 16 kbps. In addition, at the 16 kbps rate the device operates identically to a CVSD encoder decoder.

BASIC CONFIGURATION

The motivation for this enhanced CVSD coder is a speech digitizer which is compatible at 16 kbps with present CVSD encoders used in various Defense Communication Systems (DCS), but when higher channel capacity is available, produces a speech quality comparable to PCM at 64 kbps. The basic encoder-decoder configuration is seen in Fig. 1. It consists of a CVSD encoder at 16 kHz with filtered input x(t), data stream b(n), and integrated loop signal \( \hat{x}(t) \). This backbone encoder is augmented with the encoding, on an open loop basis, of the difference signal \( \Delta = x(t) - \hat{x}(t) \), at an 8 kHz sample rate. This encoding of the \( \Delta \) is an additive correction to the CVSD approximation to \( x(t) \). At the receiver-decoder the CVSD data stream, b(n), is used to recreate the output signal \( \hat{x}(t) \). If no more data is available, that is, the receiver rate is 16 kbps, the final output is simply the CVSD signal \( \hat{x}(t) \) low pass filtered to produce \( \hat{s}(t) \).

Fig. 1. Basic configuration.
While the basic CVSD encoder is operating at 16 kHz, the error correction, \( \Delta \), need only be sampled at an 8 kHz rate as long as we are operating with a 4 kHz speech band. For a fixed assignment of bits for the \( \Delta \) signal, less noise is generated by allotting all the bits to an 8 kHz sample train rather than half the bits to each 16 kHz sample. Given 2M bits/sample at rate \( f_0 \), if we redistribute these as M bits per 2 \( f_0 \) sample rate, we have increased the noise variance by \( 2^M \) and the bandwidth by a factor of 2. So in the band dictated by \( \frac{f_0}{2} \), the noise variance for the 2 \( f_0 \) sample train is \( 2^{M+1/2} \) times the \( f_0 \) sample train.

Since the basic CVSD delta modulation encoder-decoder must run at 16 kHz to be compatible with standard DCS systems, there are some disadvantages. As shown in Fig. 1a, the reconstructed 16 kHz based analog signal \( \hat{x}(t) \) must be resampled at 8 kHz to provide a 2, 4 or 6 bit encoding of \( \Delta \). At the decoder shown in Fig. 1b, the path from \( \hat{x}(t) \) to the adder must also be sampled at 8 kHz when extra \( \Delta \) samples are available. However when operating at the 16 kbps rate, the full 16 kHz \( \hat{x}(t) \) should be processed by the low-pass filter to minimize output noise. If we look more closely at the CVSD encoder shown in Fig. 2a, the output \( \hat{x}(t) \) signal can be modeled as the 0th order held \( x(t) \) driving an integrator to produce \( \hat{x}(t) \) as in Fig. 2b. The signal \( \hat{x}(t) \) can be considered a 16 kHz sample train low pass filtered by a frequency response

\[
H(f) = \frac{\sin \left( \frac{\pi f}{16000} \right)}{\frac{\pi f}{16000}} \times \frac{1}{f}.
\]

\[\sin \left( \frac{\pi f}{16000} \right) \approx \frac{\pi f}{16000} \text{ for } |f| < \frac{16000}{\pi} \approx 5040 \text{ Hz}.
\]

\[H(f) \approx \frac{1}{f} \text{ for } |f| < \frac{16000}{\pi} \approx 5040 \text{ Hz}.
\]

Fig. 2. CVSD encoder.

Although this is substantial smoothing, it is not adequate rejection outside a 4 kHz band. We know this empirically since the standard CVSD output requires high quality final low-pass filtering of \( \hat{x}(t) \). Given this condition, an 8 kHz downsampling of \( \hat{x}(t) \) generates more noise in the 4 kHz band than a proper prefilter followed by the sampling.

**REQUIRED MODIFIED CONFIGURATION**

To include the 8 kHz downsampling operation properly, the configuration of Fig. 3a is required at the encoder. The dotted filter #1 is required to eliminate 16 kHz sample noise in \( \hat{x}(t) \). Filter #2 is identical to #1 and delays \( \hat{x}(t) \) to compensate for the delay of \( \hat{x}(t) \) from filter #1. The solid LPF in Fig. 3a replaces #1 and #2 in this linear circuit. The decoder of Fig. 3b shows a similar change from Fig. 1b. Again filters in each path can be replaced by a single low-pass filter after the summation. Unlike the case of Fig. 1b, the decoder of Fig. 3b requires no special decision for the 16 kbps case different from the others. That is, in Fig. 1b downsampling only occurred at the rates greater than 16 kbps. In Fig. 3b no downsampling occurs whether or not a \( \Delta \) signal is available at the decoder.

![Fig. 3. Modified configuration.](image)

**CONSIDERATION OF THE \( \Delta \) ENCODING**

The 8 kHz samples of the continuous \( \Delta \) signal must be encoded with some quantizer law. Unfortunately, we cannot easily relate quantization noise introduced by the \( \Delta \) samples to quantization in the reconstructed samples using both the CVSD and \( \Delta \) outputs. For a nonadaptive quantizer curve we can rationalize the choice of a logarithmic characteristic as used for ordinary speech PCM systems even though a long term histogram (1st order probability density) of \( \Delta \) does not reflect exponential behavior.

Both linear and logarithmic encoding tables were used to encode the \( \Delta \) signal. Informal listening at 32 kbps (only 2 bits or 4 distinct levels for the \( \Delta \) indicated a less noisy perceived speech signal when the log encoder was used. At 64 kbps the difference between linear and log encoding was less distinguishable.
If $A = x - \hat{x}$ is small, then either a good match exists between a large $\hat{x}$ and large $x$ or else both $x$ and $\hat{x}$ are small. For a fixed quantizer a conservative approach would allot low error for these cases in spite of the fact that a large $x$ may be able to mask a larger quantization error. For the case of a large $A$ it is reasonable to assume a large $x$ is involved. In this case a larger quantization error is acceptable, based on standard PCM practice.

An adaptive approach to quantizing the $\Delta$ signal uses the CVSD slope waveform $(p(t))$ in Fig. 2a to perform scaling on the $\Delta$ before quantizing or equivalently to shrink and expand the quantization levels with $p(t)$. The slope waveform is used since it has the property of following the speech envelope in some smoothed fashion. Listening to a simulation of such a quantizer in the enhanced CVSD configuration at 32 kbps indicated no clear noise improvement in the output speech signal. In fact such an adaptive scheme requires an increase in encoder-decoder complexity which would be warranted only by a substantial noise reduction. Compared to the fixed log encoder, the adaptive quantizer will actually introduce more quantization error for small amplitude samples occurring during a speech segment with large envelope. The slope signal will track the large envelope and set the quantizer levels correspondingly, distorting any small samples during that time.

EMBEDDED DATA FORMAT

As discussed in the introduction, the enhanced CVSD encoder can be used to generate a data stream at 64, 48, 32, or 16 kbps. What is more interesting is the production of an embedded data stream formatted for use in a packet voice network environment. In such an environment the 64 kbps stream may be received in light traffic, but heavy network traffic may allow only 16 or 32 kbps to reach the destination decoder. For such use a format such as shown in Fig. 4b is necessary. Each 20 ms four separate groups of data words are output to a network from the encoder. The highest priority group consists only of type $W_1$ words which are simply the CVSD bits packed into byte form as shown in Fig. 4a. The next highest priority group consists of type $W_2$ words which are the most significant two bits of the six bit $\Delta$ word packed four to a byte. If only these two highest priority packets are received, a decoder can produce speech at the 32 kbps enhanced CVSD rate. If the next lowest priority packet is received, 48 kbps speech will be output. Finally all four received packets will produce a 64 kbps output speech signal for that 20 ms period.

The receiver needs no special information about data rates. It will simply recreate a speech waveform based on the data in its possession at the time.
REFERENCES


This work was sponsored by the Defense Advanced Research Projects Agency.

The U.S. Government assumes no responsibility for the information presented.