論文 82-19-6-11

Hybrid Companding Delta Modulation with Silence Detection

(묵음 檢出 機能을 使用한 하이브리드 압신 델타 變調器)

趙 東 鎬*, 殷 鍾 官*

(D. H. Cho and C. K. Un)

要 約

본 논문에서는 HCDM(hybrid companding delta modulation)을 사용하여 음성을 부호화할 때, 음성의 간헐성을 이용하여 전송속도를 줄이거나 잡음에 대한 신호비(SQNR)을 증가시키는 연구를 하였다. 음성부분과 묵음(silence)부분을 식별하는 판별기를 이용하여 음성의 묵음부분을 검출하며, 이때 음성부분에 대해서는 HCDM 부호화를 행한다. 음성을 5msec 간격으로 검사하여, 그때 검출되는 묵음부분에 대해서는 그 구간이 묵음이라는 정보만을 전송하며, 수신단에서는 이 정보를 이용하여 묵음부분을 재생한다. 그런데 HCDM 부호기는 2진 신호를 일정한 속도로 또 동기적으로 전송하기 때문에, 버퍼(buffer)를 사용해야 하며 또한 그것을 효율적으로 제어해야 한다. 음성을 부호화할 때, 묵음검출 기능을 이용하는 HCDM 부호기를 사용하면, 재래의 HCDM 보다 잡음에 대한 신호비를 6dB 만큼 증가시킬 수 있거나, 전송속도를 $\frac{1}{3}$가량 줄일 수 있다.

Abstract

In this paper we exploit the use of the intermittent property of speech to reduce the transmission rate or to increase signal-to-quantization noise ratio (SQNR) in coding speech by hybrid companding delta modulation (HCDM). In this scheme we detect silence in speech by a speech/silence discriminator. HCDM coding is done only for speech portion. For silence that is detected in every block of 5 ms, only the information indicating that the block is silence is transmitted. At the receiver silence is created based on this information. Since the HCDM coder transmits binary signal synchronously at a fixed rate, the use of a buffer and its efficient control is essential. By using the HCDM with silence detection in coding speech, we could improve SONR by as much as 6 dB over the conventional HCDM or reduce the transmission rate by one third of the HCDM rate.

I. Introduction

As the number of communication users is increasing rapidly, considerable effort has recently been given to develop bandwidth-efficient coding schemes for transmission of speech. One effective method of reducing redundancy of speech is to use a predictive waveform quantization method such as adaptive delta modulation (ADM) and adaptive

*正会員, 韓国科学技術院 電気 및 電子工学科
(Dept. of Electrical Engineering, KAIST)
接受日字：1982年 4月 27日

differential pulse code modulation (ADPCM). ADM is known to be one of the most effective speech coding methods in the range of 16 to 32 kbits/s. According to a recent study by Un and Lee,[1] hybrid companding delta modulation (HCDM)[2] yields the best performance among many different ADM schemes.

Another effective method of increasing channel capacity is to use a scheme in which silence in conversational speech is detected and not transmitted. One typical example is time assignment speech interpolation (TASI).[3]

If we combine these two methods, that is, predictive coding and speech interpolation, significant bandwidth compression can be achieved in coding speech. This paper deals with such a coding scheme. Specifically, we study here HCDM coder with silence detection. By incorporating a silence/speech detection algorithm in the HCDM coder, one can eliminate transmission of silence, thereby achieving significant bandwidth compression. In our system the output bit stream is transmitted synchronously. This is in contrast with other coders with spech interpolation in which speech is transmitted asynchronously. To have synchronous transmission, a buffer and an effective buffer control algorithm must be used. This aspect will be discussed in detail.

Following this introduction, we describe the algorithm of the HCDM with silence detection (HCDM-SD) in Section II. Computer simulation results of the system and discussion on system performance follow in Section III. Finally we make conclusions in Section IV.

II. Algorithm of HCDM with Silence Detection

1. Overall System

A block diagram of the HCDM-SD is shown in Fig. 1. The system can be divided into three major parts: HCDM coder and decoder, a silence detector, and a buffer and its control unit. The input signal is segmented into blocks of 10 ms duration. By using a silence detector one can decide whether each block is speech

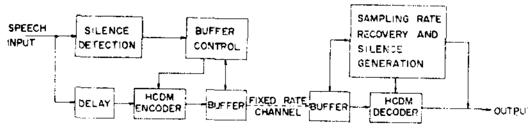


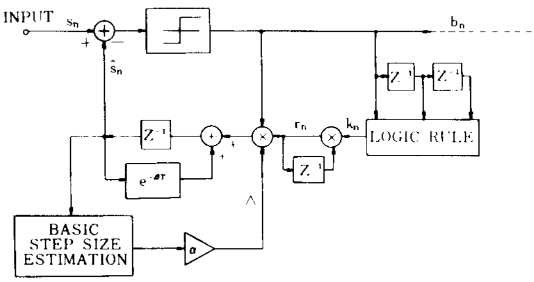
Fig. 1. Block diagram of HCDM-SD system.

or not. When the input signal is decided to be speech, the HCDM coder encodes speech at the rate of f_{sh} and the output bits are stored in the buffer. When silence is detected, code bits are not stored, but only the information indicating that the corresponding block is silence is stored. For a speech segment the output bits of HCDM are organized into packets, each 120 bits with a 2-bit header included; and for a silent segment the HCDM output is just header bits indicating that the segment is silence. The information stored in the buffer is transmitted synchronously at a fixed rate. At the receiver the received signal is stored in a buffer. If the header indicates that the incoming signal is a speech segment, the HCDM decoder decodes the received bits to produce speech. But, if the header indicates that the segment is silence, the decoder is switched off so that artificial silence can be inserted before the next speech segment arrives.

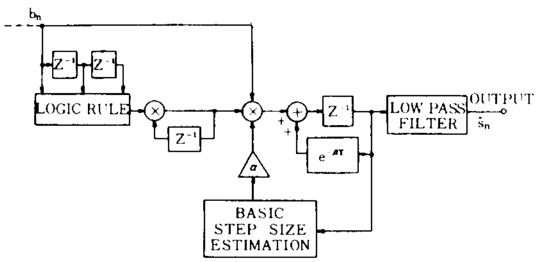
One can note that three critical parts of the HCDM-SD system that affect the system performance are the HCDM coder itself, the silence detector, and the buffer control unit. If silence is detected inaccurately and thus speech portion is cutoff, the received speech would sound unnaturally. Also if overflow occurs in the buffer because of the limitation of the buffer size, speech signal can be lost. Hence, the accuracy of speech detection and the effectiveness of buffer control are essential for the system to work satisfactorily. The HCDM code that is the heart of the system provides efficient coding of input speech. These three critical parts are now described. efficient coding of input speech. These three critical parts are now described.

2. HCDM Coding

The HCDM code is shown in Fig. 2. The



(a) Encoder



(b) Decoder

Fig. 2. Block diagram of HCDM coder

Table 1. HCDM companding logic.

b_n	b_{n-1}	b_{n-2}	Multiplication Factor (k_n)
+	+	+	1.5
–	–	–	1.5
–	–	+	1
+	+	–	1
–	+	+	0.66
+	–	–	0.66
–	+	–	0.66
+	–	+	0.66

feature of the system is that it uses both syllabic and instantaneous companding schemes. The quantizer basic step size is determined by the slope energy of the decoded signal that is estimated every 5 ms. In addition, the step size of HCDM is adjusted at each sampling time according to the logic rule shown in Table 1. Accordingly, given the sampled signal s_n in the ADM coder, sign bits are generated as

$$b_n = \mathrm{sgn}(s_n - \hat{s}_n)$$

with

$$\hat{s}_n = e^{-\beta T}\,\hat{s}_{n-1} + b_{n-1}\,\Delta_{n-1}$$

$$\Delta_n = \Delta\gamma_n$$

$$\gamma_n = k_n\gamma_{n-1}$$

and

$$k_n = f(b_n, b_{n-1}, b_{n-2}),$$

where β is the inverse of a leakage time constant of the prediction loop; γ_n is the instantaneously companded step size at the n_{th} sampling instant; Δ_n is the step size of n_{th} sampling instant; k_n that depends on the present and two previous bits is a multiplication factor determined by the logic rule shown in Table 1; and A is the basic step size that is obtained from the input signal slope energy estimate E computed with the predicted signal over every 5 ms interval. The basic step size of the quantizer is obtained by

$$\Lambda = \alpha E,$$

where α is a scale factor, and E is the average slope energy of the decoded signal.

It is important to choose a proper value of the scale factor α. If the value of α is larger than one, the decoded signal becomes unstable. On the other hand, if α is too small, amplitude of the decoded signal approaches zero regardless of input signal variation. The optimum vlaue of α that gives maximum SQNR may be obtained by obtaining SQNR's as a function of α. When the input signal of HCDM is speech, its optimum value is 0.8. Un et al. obtained this value by computed simulation.[2]

3. Speech Detection

Several speech detectors have been developed since the TASI system first appeared two decades ago.[3-7] Most of these detectors discriminate speech based on level or envelope detection of input signal. Recently, Un and Lee used a different approach in which speech/silence discrimination is made by using linear delta modulation (LDM) bit stream.[7] Our speech detector is a simplified version of Un and Lee's system.

To make speech/silence detection, analog speech signal is coded by an LDM coder, and bit alternations of the LDM output bit stream is counted. The silence/speech decision for a segment of speech is then made by comparing the counted value with a preset threshold.

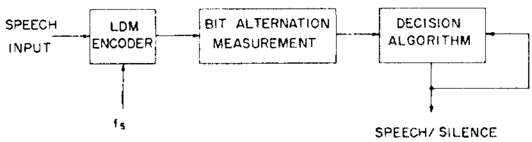


Fig. 3. Block diagram of the silence/speech detector.

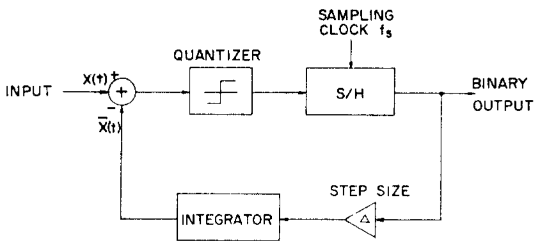


Fig. 4. Block diagram of LDM encoder.

A block diagram of the speech detector used in our system is shown in Fig. 3, and the LDM coder is shown in Fig. 4. In the LDM coder the magnitude of the slope of the estimate $\bar{x}(t)$ cannot exceed the product of the quantizer step size Δ and the sampling frequency f_s, i.e., $\Delta \cdot f_s$. If the slope of the input signal $x(t)$ is greater than $\Delta \cdot f_s$ and thus slope overload occurs, the output bits generated would have the same polarity. On the other hand, if the slope of $x(t)$ is less than $\Delta \cdot f_s$ and thus granular noise occurs, the output bit pattern alternates in bursts by logic 1's and logic 0's. During silence the bit stream alternates in 1's and 0's at successive sampling instants. Accordingly, one can discriminate speech by examining variation of LDM bit patterns. In our system the LDM bit stream is obtained in 5 ms blocks (each block is 120 bits long when the sampling rate is 24 kHz), and the number of bit alternations is counted in each block. The silence/speech decision is made based on the bit alternation rate of n_{th} block, $R(n)$, and the previous decision state. To make a decision, the bit alternation rate is compared with the preset threshold value THV. When $R(n)$ is greater than THV. that portion of input signal is considered as silence. On the other hand, when $R(n)$ is less than THV, that segment is decided to be speech. In the transition of silence to speech and speech to silence, incorrect decisions occur sometimes for a short period of time. To alleviate these problems, a delay time of 10 ms is used in our system. This delay time that has been obtained by computer simulation gives the least decision errors in our system.

4. *Buffer Control*

Since the size of the buffer used in our system is finite, overflow and underflow can occur. Accordingly, an effective buffer control must be used to prevent the undesirable effect.[8] Let f_s be the transmission rate, f_{sd} the sampling frequency to be decided and f_{sh} a sampling frequency that is greater than f_s. Also, let B be the buffer size and b(i) the number of bits in the transmitter buffer at time i. We use the following buffer control algorithm:

$$
\begin{aligned}
f_{sd} &\leq f_s, \text{ if } b(i) = B \\
&= f_{sh}, \text{ if } 0 < b(i) < B, \text{ and } R(n) < THV \\
&= f_s, \text{ if } 0 < b(i) < B, \text{ and } R(n) > THV \\
&\geq f_s, \text{ if } b(i) = 0 .
\end{aligned}
$$

Thus, when the transmitter buffer is full, the sampling rate of the coder must be less than or equal to the actual transmission rate. In the case of a buffer being empty, the coder sampling rate should be greater than or equal to the transmission rate. When the buffer is not full or empty and the bit alternation rate is below the threshold, we use a sampling rate that is greater than the transmission rate. Also, when the buffer is not full or empty and the bit alternation rate is above the threshold, the sampling rate is forced to be the same as the channel transmission rate. One should note that, although the buffer control algorithm described above prevents overflow or underflow, the system performance becomes degarded if the buffer size is too small. Therefore, the buffer must be large enough to guarantee satisfactory performance. In general, as the transmission rate is increased, the buffer size is also increased. The performance gain of HCDM-SD is dependent upon the given delay time and the characteristics of input signal. The performance of HCDM-SD can be improved to some extent as the buffer size or delay time is increased. However, if the buffer size is larger than some threshold value, the performance gain no long increases and the delay time becomes larger. This undesirable large delay time would give unnatural speech quality. Therefore, the buffer size must be compromised according to the delay time, the transmission rate and the characteristics of the input signal. This compromised size may be obtained by computer simulation.

Table 2. Average bit alternation rate of LDM signal.

Sampling Frequency	Bit Alternation Rate for 5 ms Windowing	Average Bit Alternation Rates		
		Speech		Silence
		Voiced	Unvoiced	
8 kHz	40	5	23	31
16 kHz	80	5	20	66
24 kHz	120	5	27	106

III. Computer Simulation Results and Discussion

Computer simulation has been done using various male and female speech of different characteristics. The speech used in simulation was bandlimited to 3.4 kHz. In general, bit alternations of LDM signal depend on the LDM sampling frequency and input speech characteristics. Hence, it is necessary to optimize the LDM sampling frequency for correct silence/speech decision. Accordingly, we first examined variation of bit alternation rate by changing LDM sampling rate. Table 2 shows the bit alternation rate R(n). It can be noticed from Table 2 that for the silent portion, the bit alternation rate increases as the sampling rate becomes higher, while it remains fairly constant for the speech portion. Also, it becomes more difficult to distinguish silence from speech as the sampling frequency becomes lower. Therefore, a proper sampling frequency must be chosen for accurate silence/speech discrimination. According to our simulation results, the sampling frequency must be greater than 16 kHz for correct decision.

From Table 2, it is seen that bit alternation rate is high for silent portion, but is low for the speech segment. Therefore, it is easy to set a proper threshould values by which silence/speech decision is made. The threshold levels for bit alternation rate must be optimized to minimize the decision errors. Optimum threshold value is set at the midway between the average bit alternation rate of speech and the same for silence. The optimum threshold values THV are 50 at the sampling frequency of 16 kHz and 70 at 24 kHz. With this optimum value, the silence/speech decision for speech can be made correctly, providing that the input speech is noise free.

Fig. 5 shows a flow chart of our decision algorithm. The bit alternation rate for the LDM output is measured and compared with the threshold value, and the output AA is produced at each windowing interval. If the bit alternation rate is greater than THV, AA is level '1'. Otherwise, AA is level '0'. The

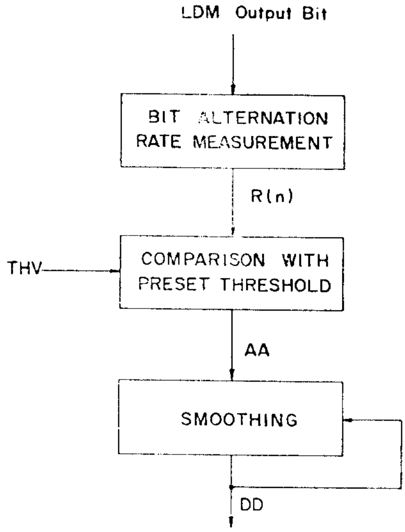


Fig. 5. Flowchart of the decision algorithm.

silence/speech decision is made based on the state of AA and its previous state. The decision state DD is level '1' or level '0'. Level '1' means silence and level '0' implies speech. In the transition of silence to speech and speech to silence, the state of R(n) alternates between low and high states for a short time interval, typically 5 to 20 ms. In the continuous speech portion, silence portion can be detected infrequently. Also, in the long silence duration, speech portion may be detected because of noise. To alleviate these phenomenon, a smoothing scheme is used. In the speech-to-silence transition state, a delay time of 20 ms gives an error-free decision. Also, in the silence to speech transition state, a delay time of 5 ms can correct decision errors due to short noise.

To examine the performance of HCDM with silence detection, signal-to-quantization noise ratio (SQNR) of the system has been obtained as a function of the input signal level over the range of 80 dB. Fig. 6 shows the result along with the performance of the conventional HCDM. It is seen that the performance of HCDM-SD is about 6 dB better than that of HCDM. The reason for this performance improvement is as follows. When the portion of input signal is considered as silence, only the side information that indicates silence is transmitted. By doing so, bandwidth compression of as much as 33% can be achieved for a given SQNR; or SQNR can be improved by increasing the sampling rate. In our HCDM-SD system, the transmission rate is fixed at 16 kbits/s, but speech portion is sampled at 24 kHz unless the buffer is overflow. Fig. 7 shows the performance of HCDM-SD and HCDM when channel is noisy. Again, the performance of HCDM-SD is better than HCDM. In this case, we assumed that no synchronization errors occur.

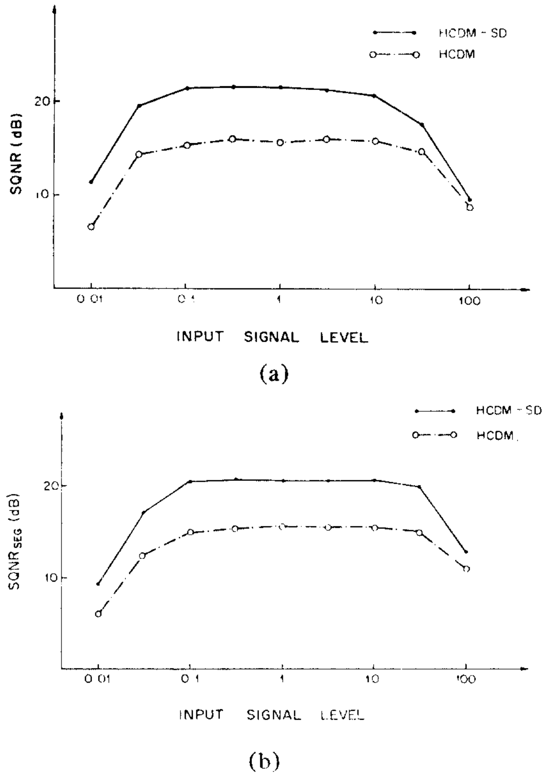


Fig. 6. (a) SQNR's of HCDM-SD and HCDM vs input signal level.
(b) Segmented SQNR's of HCDM-SD and HCDM vs input signal level.

As mentioned, in the HCDM-SD system a time buffer control scheme is used because the channel rate is different from the source coding rate. In our simulation, a 3 kbit buffer was used. Hence a delay time of 3/16 sec results. This delay would give perceptually a negligible effect.

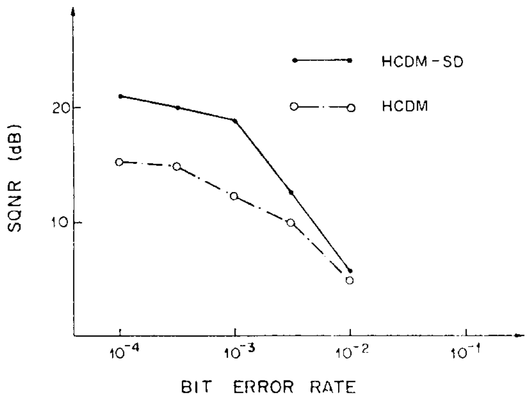


Fig. 7. SQNR's of HCDM-SD and HCDM vs channel error rate.

When the buffer size is made very large or unlimited, one can improve the performance further by as much as 9 dB over the conventional HCDM. However, the time delay resulting from the use of a large buffer would give unnatural speech quality. Hence, a compromised buffer size must be decided based upon the sampling frequency used.

Finally, considering implementation of the HCDM-SD algorithm using standard integrated circuits, we need hardwares such as a silence detector, a buffer, and a buffer control logic in addition to the HCDM codec itself. Hence, the hardware of HCDM-SD system is about twice as complex as that of the conventional HCDM. However, the LSI technology is being rapidly progressed in recent years. Therefore, the cost and hardware complexity should be no problem since the system is LSI-implementable.

IV. Conclusions

In this paper we have studied speech coding by HCDM with silence detection. In this scheme we detect short silence in speech by a silence/speech discriminator. Waveform coding is done only for speech portion. According to our simulation results of the HCDM-SD system, it performs reliable and yields a bandwidth compression of about 33%, or a SQNR improvement of about 6 dB. This improvement results from the elimination of short pauses in speech. If one has the TASI gain by eliminating long pauses in coding speech, further bandwidth reduction and performance improvement can be achieved.

It should be noted that the silence detection scheme can be used in any speech coder. Therefore, it is possible to get the same bandwidth compression effect or performance improvement by utilizing this scheme in waveform coders such as PCM and ADPCM. The proposed scheme can also be used for economical storage of speech in a computer.

References

1. C.K. Un and H.S. Lee, "A study of the comparative performance of adaptive delta modulation systems", *IEEE Trans. Commun.*, vol. COM-28, pp. 96-101, Jan. 1980.
2. C.K. Un, H.S. Lee and J.S. Song, "Hybrid companding delta modulation", *IEEE Trans. Commun.*, vol. COM-29, pp. 1337-1344, Sept. 1981.
3. K. Bullington and J.M. Fraser, "Engineering aspects of TASI," *BSTJ.*, vol. 38, pp. 353-364, Mar. 1959.
4. P.G. Drago, A.M. Molinari, and F.C. Vagliani, "Digital dynamic speech detectors", *IEEE Trans. Commun.*, vol. COM-26, pp. 140-145, Jan. 1978.
5. E. Fariello, "A novel digital speech detector for improving effective satellite capacity", *IEEE Trans. Commun.*, vol. COM-20, pp. 55-60, Feb. 1972.
5. J.A. Jankowski, Jr., "A new digital voice activated switch", *COMSAT Techn. Rev.*, vol. 6, no. 1, pp. 159-178, Spring 1976.
7. C.K. Un and H.H. Lee, "Voiced/Unvoiced/silence discrimination of speech by delta modulation", *IEEE Trans. Acoust., Speech, Signal Processing,* vol. ASSP-28, pp. 398-407, Aug. 1980.
8. J.K. Cavero and S.K. Lee, "A simple buffer control for variable rate communications", *IEEE Trans. Commun.*, vol. COM-24, pp. 1045-1048, Sept. 1976.