THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 53, No. 5, May–June 1974
Printed in U.S.A.

On the Correlation Between Bit Sequences in Consecutive Delta Modulations of a Speech Signal

By N. S. JAYANT

(Manuscript received November 16, 1973)

We consider a communication link in which a band-limited speech signal is delta-modulated, detected, and filtered by a low-pass filter, and the analog output is delta-modulated again with an identical encoder. We are concerned with the correlation C between equal-length bit sequences, designated {b} and {B}, that result from the two stages of delta modulation. We study C as a function of the sequence length W; the starting sample T in {b}; the time shift L between {b} and {B}; the signal-sampling frequency F; and a parameter $P(\geqq 1)$ which specifies the speed of step-size adaptations in the delta modulators. ($P = 1$ provides nonadaptive, or linear, delta modulation.)

Computer simulations have confirmed that for small time shifts L and for statistically adequate window lengths W, C is a strong positive number (0.4, for example). Moreover, the C function tends to exhibit a maximum $C_{\max}$ at a small nonzero value of L (between 1 and 5, say) reflecting a delay introduced by the low-pass filter preceding the second delta modulator; and when W is on the order of 100 or more, the dependence of $C_{\max}$ on the starting sample T is surprisingly weak. Also, in the range of F and P values included in our simulation, $C_{\max}$ increased with F and decreased with P. Finally, the positive C values for small L are retained even when the delta modulators are out of synchronization in amplitude level and step size, as long as the delta modulators incorporate leaky integrators and finite, nonzero values for maximum and minimum step size.

With a given T, the C(L) function can exhibit significant nonzero values even for large L. However, these values are both positive and negative; and if correlations are averaged over several values of T, the average C(L) function tends to be essentially zero for sufficiently large L ($L \geqq 100$, say), while still preserving the strong positive peaks at a predictable small value of L. This observation is the basis of an interesting application

where the value of C is used to determine whether or not two digital codes, appearing at different points in a speech communication system, carry identical speech information.

I. THE PROBLEM

Consider a speech signal subjected to two successive stages of delta modulation, with an intermediate stage of low-pass filtering, as shown in Fig. 1. A previous paper[1] has studied how signal quality degrades as a function of the number of delta modulations. The present paper is concerned with the amount of correlation that exists between the bit sequences $\{b\}$ and $\{B\}$ from the two (identical) delta modulators. Specifically, we employ computer simulations to study the correlation

$$C = \frac{1}{W} \sum_{i=T}^{T+W} b_i B_{i+L}.$$

It is assumed that $\{b\}$ and $\{B\}$ are zero-mean sequences with equiprobable ± 1 entries. Apart from being a function of the window duration W and time shift L, the correlation C will also depend on the signal-sampling frequency F and a parameter P specifying the step-size logic used in the delta modulators. The delta-modulator simulations are described in Section II and the properties of C are described in succeeding sections.

The studies reported in this paper were prompted by an interesting potential application where the value of the correlation C would be used to determine whether or not two digital codes (appearing at different points in a speech communication network) carry the same speech information. More specifically, we were considering a telephonic system that incorporated digital and analog signal terminals capable of being interconnected via a common switching network. The problem was to determine whether digital terminals communicating with each other (in other words, handling the same speech information) could be detected by digitally correlating the signals of each digital terminal with the signals at other digital terminals in the system.[2,3] The digital coding under consideration was delta modulation, and the results of this paper indeed suggest that the detection of communicating terminals should be possible on the basis of appropriate bit correlations.

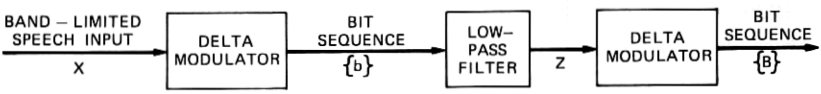

Fig. 1—Block diagram of the simulated speech communication system.

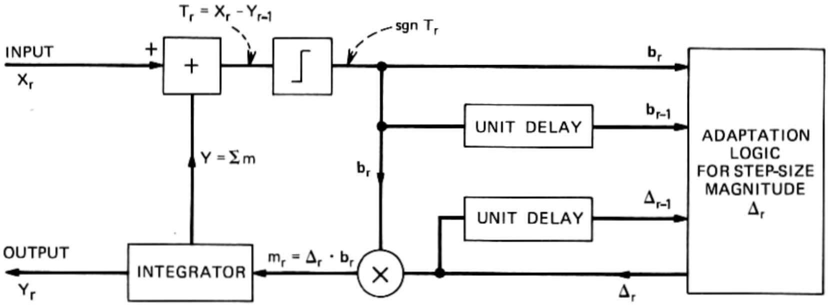

Fig. 2—Schematic diagram of an adaptive delta modulator.

II. SIMULATION DETAILS

The delta modulator utilized in our simulations is schematized in Fig. 2 and is the same instantaneously adaptive delta modulator (ADM) discussed in Ref. 4. Basically, it is described by the equations

$$b_r = \text{sgn}\,(X_r - Y_{r-1}),$$
$$Y_r = Y_{r-1} + \Delta_r \cdot b_r,$$

and

$$\Delta_r = \Delta_{r-1} \cdot P^{b_r \cdot b_{r-1}},$$

where X_r is the amplitude of the input sample r, and Y_{r-1} is the amplitude of the latest staircase approximation to it. The parameter P ($\geqq 1$) automatically increases step size when Y is not tracking X fast enough ($b_r = b_{r-1}$), and decreases it when Y is hunting around X ($b_r = -\, b_{r-1}$). Nonadaptive or linear delta modulation (LDM) corresponds to the special case of $P = 1$.

The speech signal is a 1.5-second male utterance of "Have you seen Bill?" that is band-limited to 3.3 kHz. The sampling rate, unless otherwise noted, is 60 kHz. A plot of the speech waveform appears in Fig. 3, where a number at the right of a line represents the last 60-kHz sample in that line. The original signal samples are quantized to a 12-bit accuracy, and have integer amplitudes in the range -2^{11} to $+2^{11}$. Finally, the low-pass filter is a programmed recursive filter with an 18-dB/octave roll-off. This seems to represent adequate filtering for toll-quality speech reproduction using ADM at 60 kHz.

III. DEPENDENCE OF CORRELATION ON TIME SHIFT

Figure 4 shows the dependence of C on the time-shift L for two different values of starting sample T. It is interesting to observe that both the functions show a maximum at $L = L_{\max} = 4$. Even more

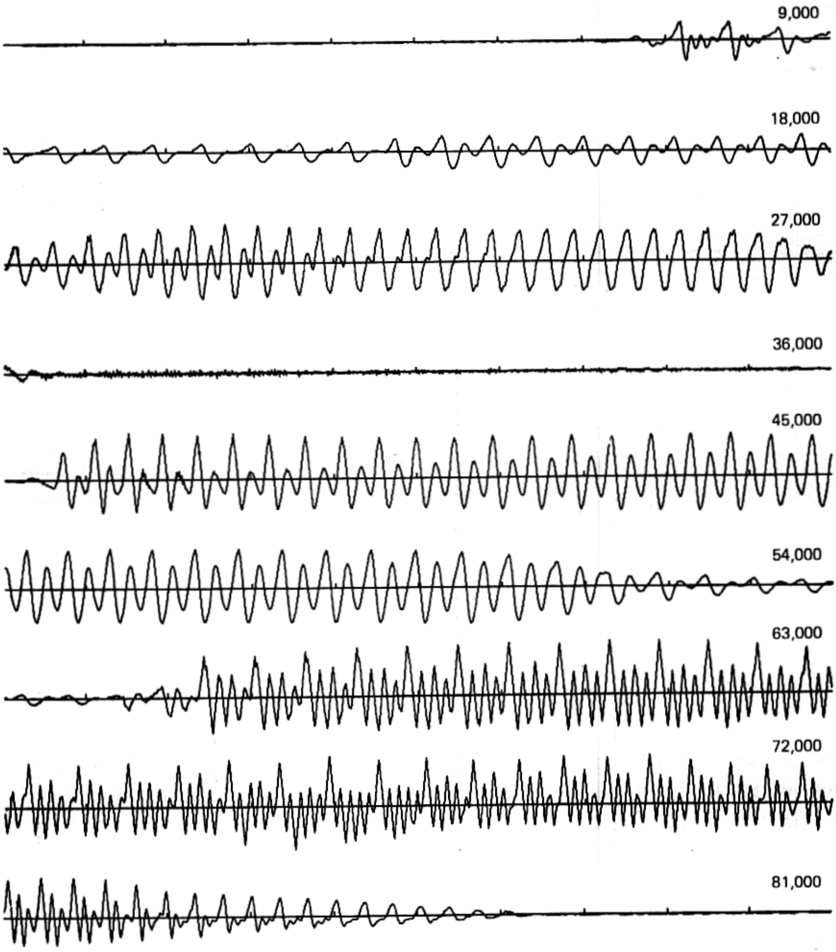

Fig. 3—The speech waveform of "Have you seen Bill?"

interesting is the fact that respective values of $CXZ(L)$, the correlation between the speech input X and the low-pass filter output Z, are also maximized (as determined in a separate simulation) at $L = 4$. It would seem that the nonzero value of L_{max} in Fig. 4 is to be attributed to the delay introduced by the low-pass filter. Actual values of C_{max} and L_{max} depend on the short-term speech spectrum and the nature of low-pass filtering, as determined by the parameters T and F (see Tables I and II). It is a general result, however, that the $C(L)$ function always shows a unique, strongly positive maximum value at a small

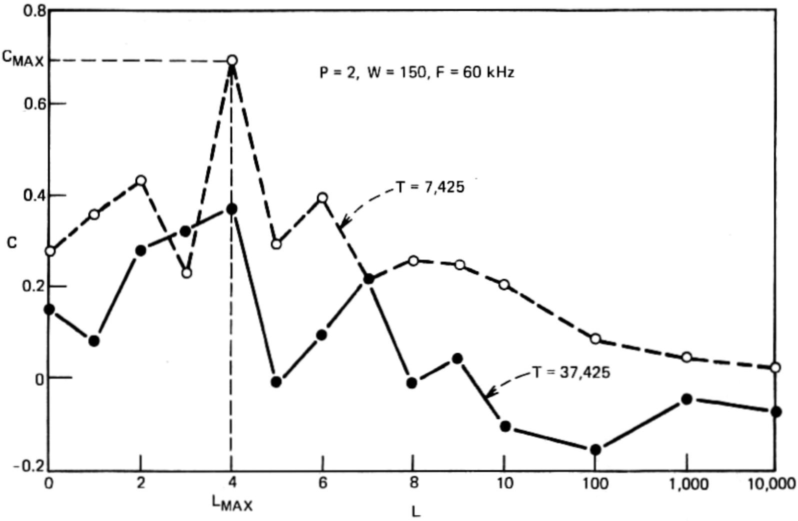

Fig. 4—Dependence of correlation C on time shift L.

value of L. Secondary peaks at large values of L tend to be less unique, and they tend to be randomly positive and negative depending on the part of the speech utterance being considered, as determined by T.

IV. DEPENDENCE OF MAXIMUM CORRELATION ON SAMPLING FREQUENCY

Table I indicates a tendency for C_{max} to decrease with decreasing sampling frequency. This may be ascribed to the fact that, at a lower sampling rate, delta modulation provides a cruder approximation to the input signal. The bits, therefore, carry more signal-independent noise information, and they have corresponding random properties that cause a decorrelation between $\{b\}$ and $\{B\}$.

Table I — Dependence of maximum correlation C_{max} on sampling frequency F ($P = 2$, $W = 1000$)

F / T	60 kHz		40 kHz	
	L_{max}	C_{max}	L_{max}	C_{max}
17425	4	0.46	3	0.32
37425	4	0.37	1	0.28
57425	4	0.48	2	0.33

Table II — Dependence of correlation C on starting sample T and time shift L ($W = 150$, $P = 2$, $F = 60$ kHz; numbers in parentheses are values of L_{max})

T \ L	0	L_{max}	10	100	1000	10000
7425 (4)	0.27	0.69	0.20	0.08	0.04	0.01
17425 (4)	0.35	0.48	0.24	−0.09	−0.04	−0.11
27425 (5)	0.23	0.47	0.11	−0.03	−0.03	0.03
37425 (4)	0.15	0.37	−0.11	−0.16	−0.08	−0.08
47425 (2)	0.29	0.33	0.31	−0.13	−0.11	−0.08
57425 (4)	0.37	0.43	0.11	−0.04	0.21	0.13
67425 (1)	0.39	0.44	0.37	0.27	0.22	−0.03
Average of C values (over T)	0.29	0.46	0.18	−0.01	0.03	−0.02

V. DEPENDENCE OF MAXIMUM CORRELATION ON STEP-SIZE MULTIPLIER P

Table III demonstrates how C_{max} tends to decrease with increasing P. Larger values of P increase the high-frequency excursions of the staircase function Y. These are filtered out by the low-pass filter. This leads to lesser correlation between the filter output Z and the bit sequence $\{b\}$ and, thence, to a decorrelation of $\{B\}$ and $\{b\}$.

VI. DEPENDENCE OF MAXIMUM CORRELATION ON WINDOW LENGTH

Our results so far have tacitly assumed window length values that represent bit sequences whose durations are of the order of a few milliseconds. Figure 5 shows C explicitly as a function of W. It is seen that very stable indications result with W in the order of 1000, although values close to a respective asymptote are sometimes reached for W values in the order of 100. In fact, a window length of $W = 10$ is seen to be sufficient, for all values of T in Fig. 5, to bring out the strong positive nature of C_{max}. The convergence of the three curves in Fig. 5

Table III — Dependence of maximum correlation C_{max} on step-size multiplier P ($F = 60$ kHz, $W = 1000$)

T	37425		37000	
P	L_{max}	C_{max}	L_{max}	C_{max}
1.0	4	0.91	4	0.89
1.5	4	0.66	4	0.62
2.0	4	0.34	4	0.44

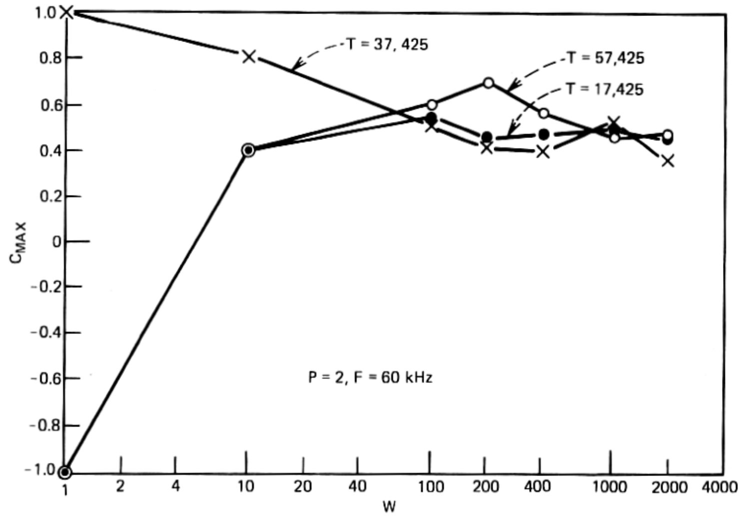

Fig. 5—Dependence of correlation C on window length W.

is not at all surprising. Note that, by definition, C should indeed be independent of T for $W \to \infty$. The results of Fig. 5 were based on a search for $C_{\max}$ in the range $0 \leqq L \leqq 10$. Except for $W = 1$, unique maxima were noted at nonzero values of L. For $W = 1$, the value of $C_{\max}$ was surprisingly constant in the range $0 \leqq L \leqq 10$, the constant value being $+1$ for one value of T, and -1 for the other two.

VII. DEPENDENCE OF MAXIMUM CORRELATION ON WINDOW LOCATION

As seen in the last section, $C_{\max}$ is a significant function of the starting sample for finite W. Table II shows the values of $C_{\max}$ for seven equally spaced values of T. The average value of $C_{\max}$ is 0.46 and the standard deviation is only 0.10. Note also that C values for large L are smaller in general, and the effect is more noticeable when correlations are averaged over T. This is because the positive $C_{\max}$ values always add up, while C values for large L, being randomly positive or negative, tend to average out to values close to zero.

At least one interesting application of the preceding observations has been suggested.[2,3] Suppose the second delta modulator has several potential speech inputs including the input Z resulting from X. The function C would then assume the strong positive values of Table II only when the input to the second delta modulator is indeed the DM version of the speech X; and it would show values of $C \to 0$ if the input

Table IV — The effect of unsynchronized delta modulators ($T = 37425$, $P = 2$, $F = 60$ kHz, $W = 1000$. Values in parentheses are for $W = 150$)

Case	Initial Conditions Y_1	Y_2	$\vec{\Delta}_1$	$\vec{\Delta}_2$	Integrators	Limits Step Size	L_{max}	C_{max}
I	0	0	1	1	Perfect	$(0, \infty)$	4 (4)	0.34 (0.37)
II	0	0	1	1	Leaky	(25, 250)	2 (3)	0.48 (0.83)
III	0	−50	1	−10	Leaky	(25, 250)	2 (3)	0.47 (0.76)
IV	0	−50	1	−10	Perfect	$(0, \infty)$	5 (1)	0.11 (0.16)

was an entirely different speech signal* (possibly due to a different speaker). This effect will be more pronounced if the averaging process indicated in Table II is carried out. We are suggesting, in other words, a means of determining whether or not two digital DM codes, appearing at different points in a speech communication network, carry the same segment of speech information. The basic recipe is a DM bit correlator with a window of 0.1 to 1 ms, and a window location T that seems to be quite uncritical, especially when time diversity (averaging over T) is possible.

VIII. EFFECT OF UNSYNCHRONIZED DELTA MODULATORS

In practice, the two delta modulators in Fig. 1 can be unsynchronized in amplitude Y and step size Δ when either or both of them are in some kind of a transient state of operation. It is an interesting result of our study that the strong positive values of C_{max} are retained even during such asynchronous periods, provided the delta modulators operate with a leaky integrator, and with finite and nonzero limits on step size. Leaky integration decreases the effect of amplitude history and, hence, the effect of amplitude asynchrony. Finite and nonzero limits on step size provide potential meeting points for the two step-size sequences, although they may begin with a different starting value.

In Table IV, Y_1 and Y_2 represent initial amplitudes for the two delta modulators, while $\vec{\Delta}_1$ and $\vec{\Delta}_2$ are the initial (signed) step sizes. The step-size limits, 250 (maximum) and 25 (minimum), include a significant range of step sizes that are called for in the adaptive delta modulation of speech (with $F = 60$ kHz, and with signal amplitudes in the range -2^{11} to $+2^{11}$).[4] Finally, the leaky integrators of Table IV leaked 5 percent of signal amplitude in a sampling period.

* This situation is hypothesized to be equivalent to the case of large L.

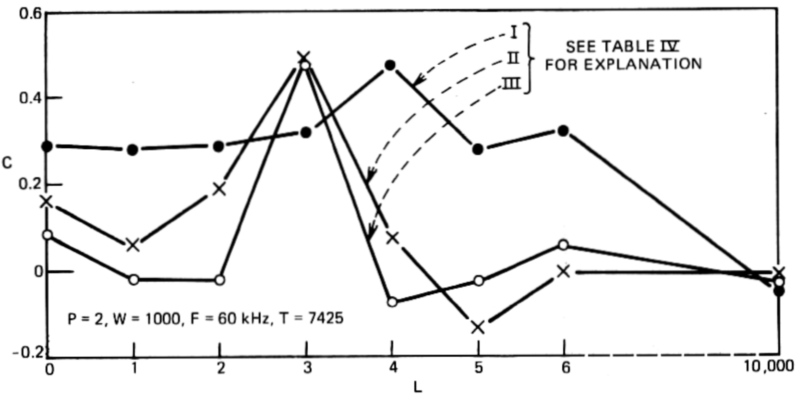

Fig. 6—Dependence of correlation C on time shift $L \leqq$—example of unsynchronized delta modulators.

Note that Table IV shows that leaky integration and finite, nonzero step-size limits are imperative in the asynchronous case (rows III and IV, Table IV) to preserve a strong positive C_{max}; they are also desirable to boost the value of C_{max} in the synchronous case (rows I and II, Table IV). (The boost is quite significant for $W = 150$). A separate simulation showed that finite (nonzero) step-size limits and leaky integrators were effective only when employed in unison; and in one study of C as a function of L, they also sharpened the peak at L_{max} (see Fig. 6).

Finally, Table V is a counterpart of Table II for the case of unsynchronized encoders. The step-size limits are 25 and 250, the leak is 1 percent in a sample duration, and $P = 1.5$. (The last two numbers are probably more representative than the corresponding values in

Table V — Dependence of correlation C on starting time T and time shift L with unsynchronized delta modulators ($W = 10$, $P = 1.5$, $F = 60$ kHz)

T \ L	0	L_{max}	10	100	1000	10000
7425	−0.4	0.0	0.0	0.0	0.0	0.0
27425	−0.4	0.4	−0.4	−0.4	−0.2	0.2
47425	0.4	0.4	0.4	−0.4	−0.2	−0.2
67425	0.6	0.6	0.6	0.6	0.6	0.0
Average of C values (over T)	0.05	0.35	0.15	−0.05	0.05	0.00

Table IV.) Finally, we have reduced the window duration to $W = 10$. This results in obviously crude $C(L)$ functions for a given beginning sample T. But, as in Table II, when $C(L)$ values are averaged over T, the resulting C function shows a clear tendency to decay to near-zero values for $L \geqq 100$. The values of C_{max} in Table V represent largest values as seen in a finite search $(0 \leqq L \leqq 5)$. None of these was a unique maximum, which is possibly due to the insufficient duration (0.16 ms) of the window, $W = 10$.

REFERENCES

1. N. S. Jayant and K. Shipley, "Multiple Delta Modulation of a Speech Signal," Proc. IEEE (Letters), September 1971, p. 1382.
2. J. L. Flanagan and N. S. Jayant, "Digital Detection of Intra-Office Calling," unpublished work.
3. J. L. Flanagan and N. S. Jayant, "Digital Signal Detection in Telephonic Communication Systems," U. S. Patent Application, December 12, 1973.
4. N. S. Jayant, "Adaptive Delta Modulation with a One-Bit Memory," B.S.T.J., *49*, No. 3 (March 1970), pp. 321–342.