[27] J. Capon, "On the asymptotic effieiency of locally optimum detectors," *IRE Trans. Info. Thy.*, vol. IT-7, pp. 67-71, Apr. 1961.
[28] B. W. Stuck, "On discrete time processing of signals corrupted by additive stable noise," (Abstract only), *Proceedings of the Eighth Annual Princeton Conference on Information Sciences and Systems,* Mar. 1974, p. 405.
[29] J. H. Miller and J. B. Thomas, "Numerical results on the convergence of relative efficiencies," *IEEE Trans. Aerospace and Electronic Systems,* vol. AES-11, pp. 204-209, Mar. 1975.
[30] S. O. Rice, "Mathematical analysis of random noise," *Bell Sys. Tech. J.*, vol. 23, pp. 282-382, July 1944.
[31] J. B. Thomas, "Nonparametric detection," *Proc. IEEE,* vol. 58, pp. 623-631, May 1970.

Detecting the Presence of Speech Using ADPCM Coding

R. W. SCHAFER, SENIOR MEMBER, IEEE, K. JACKSON, J. J. DUBNOWSKI, AND L. R. RABINER, FELLOW, IEEE

Abstract—**When speech is coded using a differential pulse-code modulation system with an adaptive quantizer, the digital code words exhibit considerable variation among all quantization levels during both voiced and unvoiced speech intervals. However, because of limits on the range of step sizes, during silent intervals the code words vary only slightly among the smallest quantization steps. Based on this principle, a simple algorithm for locating the beginning and end of a speech utterance has been developed. This algorithm has been tested in computer simulations and has been constructed with standard integrated circuit technology.**

I. INTRODUCTION

In many types of sophisticated speech-processing systems there is a need to detect the presence of speech at the input to the system. For example, in speech-recognition or speaker-verification systems, it is necessary to locate automatically the beginning and end of a speech utterance as a preliminary to detailed analysis and subsequent comparison to a reference pattern. Similarly, in digital transmission systems that utilize gaps in conversations in order to reduce the average bit rate [1], [2], it is necessary to automatically distinguish between speech and silence or background noise.

Another example, providing the motivation for the work described in this paper, is in the preparation of a vocabulary for a computer voice response system [3]. In this case, isolated words and phrases are represented in digital form by an adaptive differential pulse-code modulation (ADPCM) system for efficient storage in digital memory [4]. In order to produce the most natural concatenation of these basic elements into a longer message, it is necessary to determine precisely the beginning and end of each word or phrase so that there are no unnecessary gaps in the concatenated signal. In order to avoid tedious manual editing, an automatic system for detecting the beginning and end of speech is required. In considering this problem for ADPCM-coded speech, it was found that the nature of ADPCM coding leads to a very simple method[1] for determining the beginning and end of speech directly from the ADPCM representation [5].

We shall first discuss the fundamentals of ADPCM coding, then state the simple algorithm for finding the end points, and finally discuss a particular hardware implementation of the algorithm. Although our results are primarily motivated by the voice-response application, the basic system could be used whenever it is necessary to detect the presence of speech.

Paper approved by the Associate Editor for Data Communication Systems of the IEEE Communications Society for publication after presentation at the International Conference on Communications, San Francisco, CA, June 1975. Manuscript received September 2, 1975; revised December 19, 1975.

R. W. Schafer and K. Jackson are with the School of Electrical Engineering, Georgia Institute of Technology, Atlanta, GA 30345.

J. J. Dubnowski and L. R. Rabiner are with Bell Laboratories, Murray Hill, NJ 07974.

II. ADPCM SPEECH CODING

The basic ADPCM system, shown in Fig. 1, is a conventional differential PCM system with a fixed first-order predicator and an adaptive quantizer with a one-word memory [4]. Rather than the samples of the input speech signal $x(n)$, the difference

$$\delta(n) = x(n) - y(n) \tag{1}$$

between the input and an estimate $y(n)$ of the input is quantized. If the estimate is good, the variance of the difference will be smaller than the variance of the input $x(n)$. Thus, for a given number of quantization levels, it is possible to quantize $\delta(n)$ with less error than $x(n)$. If each quantization level is represented by a digital code word, the quantized difference signal can be represented by a digital code word $c(n)$. An approximation $\hat{x}(n)$ to the input speech sample is constructed by adding the quantized difference signal to the estimate $y(n)$; i.e.,

$$\hat{x}(n) = y(n) + \hat{\delta}(n). \tag{2}$$

Thus from (1) and (2), it can be seen that $\hat{x}(n)$ differs from $x(n)$ by the error due to quantizing $\delta(n)$. The estimate $y(n)$ is obtained from the reconstructed signal by a simple first-order linear prediction; i.e.,

$$y(n) = a\hat{x}(n-1).$$

The quantization error can be further reduced by adjusting the range of the quantizer so as to match the peak amplitude of the signal. This means that for low-level signals such as fricatives, the step size (difference between successive quantization levels) should be small compared to that required to represent high-level voiced speech sounds.

The quantizer used in our system is of the "mid-riser" type as depicted in Fig. 1. A 4-bit (or 16-level) quantizer is used; however, for simplicity Fig. 1 shows a 3-bit (or 8-level) quantizer. The quantizer levels are coded as follows: The eight positive levels are represented by the 4-bit binary numbers 0000 through 0111, with 0000 representing the smallest positive level and 0111 representing the largest. Likewise, the 8 negative levels are represented by the numbers 1000 through 1111, with 1000 representing the negative level closest to zero, and 1111 representing the most negative level. (This labeling of quantizer levels was used in the system described in [6], and is different from the labeling of [4].) Thus the most significant bit can be thought of as the sign bit and the remaining 3 bits can be thought of as the magnitude. Specifically $\hat{\delta}(n)$ is

[1] The algorithm to be discussed in this paper was designed for detecting speech already coded in ADPCM format and primarily for application in a high signal-to-noise ratio environment. It is not suggested that signals which are digitized in other formats (e.g., PCM, ADM) be converted to ADPCM format to utilize this speech-detection algorithm.

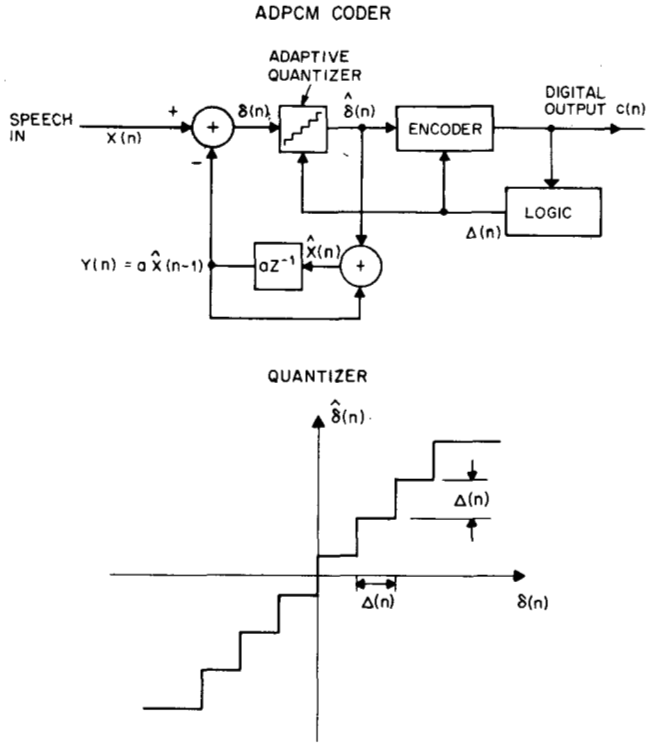


Fig. 1. Differential PCM system with an adaptive quantizer.

related to the code words $c(n)$ by

$$\hat{\delta}(n) = [\tfrac{1}{2}\Delta(n) + \Delta(n) \cdot | c(n) |] \operatorname{sgn} [c(n)] \tag{3}$$

where $\Delta(n)$ is the quantizer step size.

In Fig. 1, the block labeled LOGIC monitors the coded output and adapts the step size on the basis of the previous quantizer output. For example if the magnitude of the previous code word is large, then the step size is increased, whereas if the magnitude of the previous code word is small, the step size is decreased. This is accomplished through the relation

$$\Delta(n) = M\Delta(n-1), \tag{4}$$

where M is a function of the previous code word $c(n-1)$. If $| c(n-1) |$ is small then $M < 1$ so that $\Delta(n)$ decreases, and if $| c(n-1) |$ is large, $M > 1$ so that $\Delta(n)$ increases. The details of the adaptation strategy are given in [4].

The step-size adaptation is equivalent to compressing the amplitude of the speech signal so that the quantizer treats low-level signals much the same as high-level signals. The adaptation strategy is designed so that all quantization levels are used about an equal percentage of time. However, when the signal amplitude approaches zero, the step-size will also tend to approach zero. For this reason and for reasons concerning practical implementation, the step-size variations according to (4) are constrained by

$$\Delta_{min} \leqslant \Delta(n) \leqslant \Delta_{max}. \tag{5}$$

When the signal level becomes very small for a time period of a few milliseconds, the step-size falls to its minimum value and remains there, and the difference signal will then fall within the very lowest quantization levels. Thus, when no speech is present at the input, the code words will vary only slightly from sample to sample, and the magnitude of the code words will be small. This property of ADPCM coding is the basis for determining the presence of speech.

III. THE SPEECH-DETECTION ALGORITHM

Fig. 2(a) shows a plot of the sequence of code words for the beginning of the word three. The largest positive excursion corresponds to code word 0111 and the largest negative excursion corresponds to 1111. The sampling rate is 6 kHz and there are 256 samples/line so that each line represents approximately 40 ms of speech. Note that for the entire first line and the first part of the second line, the code words remain within the first few positive and negative values, but then in the second line, the code word sequence rather abruptly begins to vary among all the possible code word values. The first line and part of the second correspond to silence, while the remainder of the second line corresponds to the low-level fricative |th| at the beginning of the word |three|. This can be seen in Fig. 2(b), which shows the speech waveform corresponding to the code word sequence in Fig. 2(a). Thus, we associate a persistence of small-magnitude code words with silence and the frequent appearance of large magnitude code words is associated with speech (either voiced or unvoiced). This basic characteristic of the ADPCM code word sequence is placed clearly in evidence by a measurement of what is termed the "code word energy," which is defined as the sum of code word magnitudes over a 100-sample, or 16-ms, window. That is,

$$E(n) = \sum_{k=n-99}^{n} | c(k) |. \tag{6}$$

Fig. 2(c) shows the quantity $E(n + 50)$ for the code word sequence of Fig. 2(a). (The shift of 50 samples aligns the energy measurement with the center of the 100-sample interval.) The quantity $E(n)$ ranges between 0 (all code words

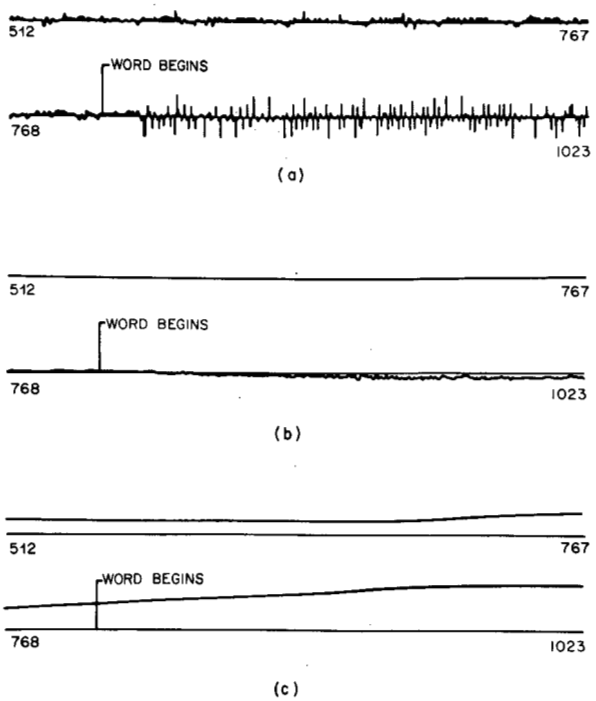


Fig. 2. Illustration of properties of ADPCM coding. (a) Code word sequence for the beginning of the word /three/. (b) Corresponding speech waveform. (c) Code word energy. (Each line represents 256 samples at 6-kHz rate.)

in the 100-sample interval either 0000 or 1000) and 700 (all code words in the 100-sample interval either 0111 or 1111). For an adaptive quantizer that is designed so that all quantizer levels are used about the same percentage of time $E(n)$ is approximately 350 during speech intervals and it is significantly lower during silence.

This suggests a rather simple algorithm for determining the beginning and end of a speech utterance. The code word energy is computed at each sample, and compared to a preset threshold which is set midway between zero and the average code word energy for speech. When the code word energy exceeds this threshold for 300 consecutive samples (50 ms), then the sample at which $E(n + 50)$ first exceeded the threshold is chosen as the beginning of the utterance. The energy computation and comparison to the threshold is continued at each sample time and when the code word energy falls below the threshold and remains below for 1000 consecutive samples (160 ms), then the sample at which $E(n + 50)$ first fell below threshold is chosen as the end of the speech utterance. Requiring the threshold to be exceeded for 50 ms prevents short clicks and breath noise from being falsely labeled as the beginning of speech. Likewise, requiring the code word energy to fall below the threshold for 160 ms ensures that a stop consonant within a word or phrase will not be mistaken for the end of an utterance.

The beginning of the word |three| as determined by this algorithm is marked on the code word sequence in Fig. 2(a), on the speech waveform in Fig. 2(b), and on the code word energy in Fig. 2(c). It is obvious that the code word energy displays the silence/speech distinction in a very simple way. More examples of this type are given in [5].

IV. HARDWARE IMPLEMENTATION

The details of implementing the ADPCM coder described in Section II in digital hardware are given elsewhere [7]. Using this hardware system to generate the ADPCM code words, the algorithm described in Section III was implemented in hardware using standard integrated circuits. The hardware implementation of the speech-detection algorithm is conveniently described in terms of two basic parts: 1) computation of the code word energy and 2) logical operations for detecting the presence of speech at the input to the ADPCM coder.

Computation of Code Word Energy

The computation of the code word energy as defined in (6) is depicted in Fig. 3. This portion of the system consists of a 100-word 4-bit shift register to store the required 100 code words, two counters for implementing the summation, and logic to control the summation process. The most recent 100 code words are stored in the 4-bit recirculating shift register. The control logic of the energy computation is informed through the DTA RDY control signal each time the ADPCM coder produces a new code word $c(n)$. (This occurs at a sampling rate of 6 kHz.) This initiates a synchronous computation sequence which generates the code word energy $E(n)$. First the energy accumulator is set to zero. Then, the least significant three bits (the magnitude) of the new code word are loaded into the shift register through the data selector. At the same time, the fourth bit of the new word is set to mark the most recent word. Then the data selector is set to recirculate the shift register from output to input. As each three-bit code word magnitude becomes available at the output of the shift register, it is added to the accumulated sum. After 100 shifts occur, the control logic of the energy computation signals the speech detection logic unit that a new value of code word energy is available. The fourth bit of the shift register is used as a marker or flag bit to indicate the completion of 100 shifts. This bit is set when a new word enters the shift register and is cleared when that word is recirculated. This scheme avoids the use of an eight-bit counter and associated logic to monitor the position of the shift register.

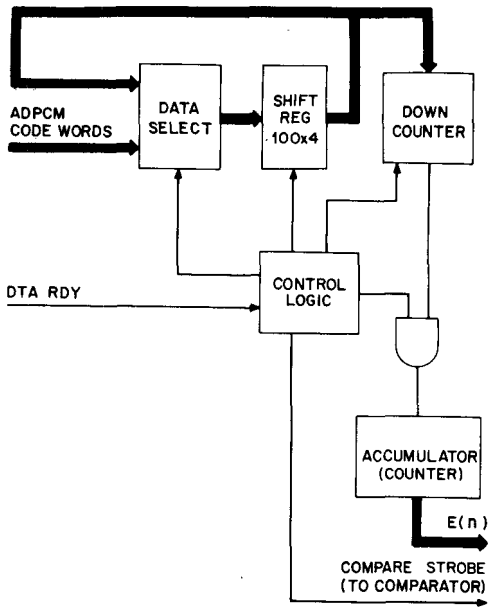


Fig. 3. Block diagram of code-word energy computation.

The accumulation of the code word energy is accomplished by loading each code word into a down-counter when it becomes available at the output of the shift register. Then, before clocking the shift register again, the down-counter and the accumulator (which is nothing more than an up-counter) are simultaneously clocked until the down-counter reaches zero. The up-counter is 10 bits long to accommodate the unlikely occurrence of the maximum value of 700. In this way, $E(n)$ is computed without an arithmetic unit.

Since the energy computation unit is self synchronizing after being clocked by the ADPCM coder, the main requirement is that the 100-word accumulation be accomplished in the sampling period of the ADPCM coder. In the present implementation, the shift register speed permits the ADPCM clock rate to be as high as 10 kHz; however, if sampling rates higher than 6 kHz are used, the 100-sample averaging window will correspond to proportionately less time.

Speech-Dectection Logic

The speech-detection logic unit is depicted in Fig. 4. It consists of a comparator, flip-flops for memory, a 10-bit counter, and simple combinational control logic. At each sample time the speech-detection unit receives a new value of the code word energy, which is compared to a preset threshold number T. If $E(n) > T$, the output of the comparator assumes the logical TRUE state. This control signal is stored in a one-bit latch whose output is an input to the combinational logic circuit. When the control signal DTA RDY becomes TRUE again, this logic outputs a pulse on the START-OF-UTTERANCE line if the SPEECH PRESENT signal is FALSE and if $E(n)$ has been above threshold for 288 consecutive samples, as determined by the 10-bit counter. (The number 288 is the closest convenient number to 300.) This pulse also sets the SPEECH PRESENT level to TRUE. Likewise, when SPEECH PRESENT is TRUE and $E(n)$ falls below threshold and remains below for 1024 consecutive samples (closest convenient number to 1000), the combinational logic outputs a pulse on the END-OF-UTTERANCE line, which in turn resets the SPEECH PRESENT flip-flop to FALSE.

The operation of the speech-detection system is illustrated by Fig. 5. In the top part of the figure, the upper trace is the SPEECH PRESENT control level and the lower trace is the speech signal that was input to the system. Notice that the

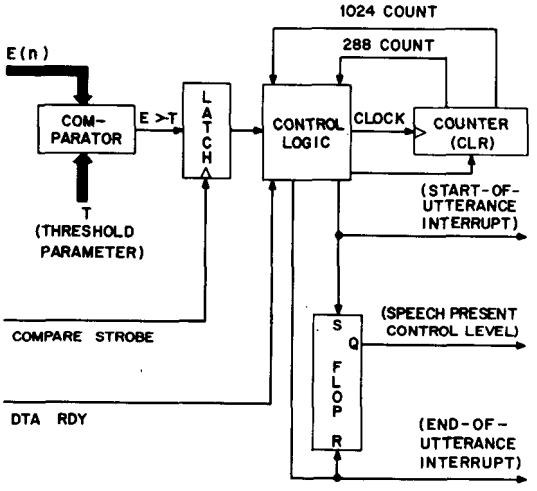


Fig. 4. Block diagram of logic for determining the presence of speech from the code word energy.

control level becomes TRUE about 50 ms after the beginning of the speech utterance. Likewise the lower half of the figure shows the same two signals at the end of an utterance. In this case the control level becomes FALSE about 160 ms after the end of the speech utterance.

V. APPLICATION TO DIGITAL RECORDING

Fig. 6 shows how the speech detection system can be used to facilitate recording of ADPCM coded speech in digital computer memory. Code words are received from a hardware ADPCM coder simultaneously by both the computer and the speech-detection system. An elastic 288-word storage is maintained in the computer to save potentially valid speech data. The START-OF-UTTERANCE interrupt can be used to cause the computer to save the most recent 288 words and to continue saving ADPCM samples until the END-OF-UTTERANCE interrupt is received. At this time, the last 1024 words received prior to the END-OF-UTTERANCE interrupt can be discarded.

This system provides simple control of the process of recording the vocabulary for a computer voice response sys-

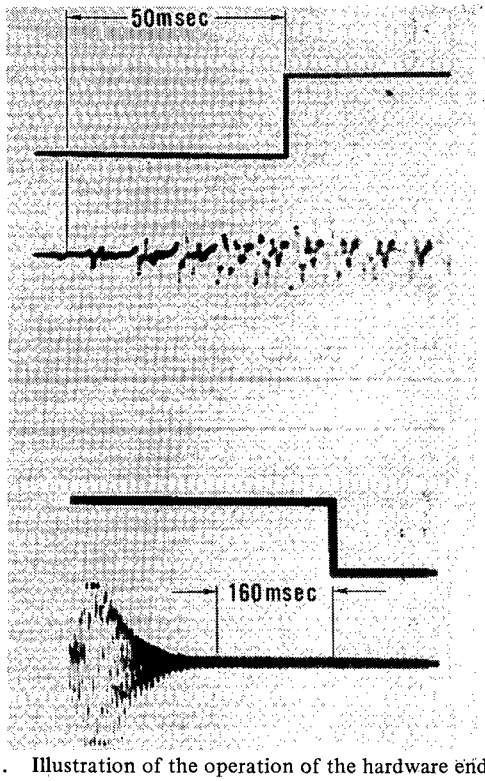


Fig. 5. Illustration of the operation of the hardware end-point detector.

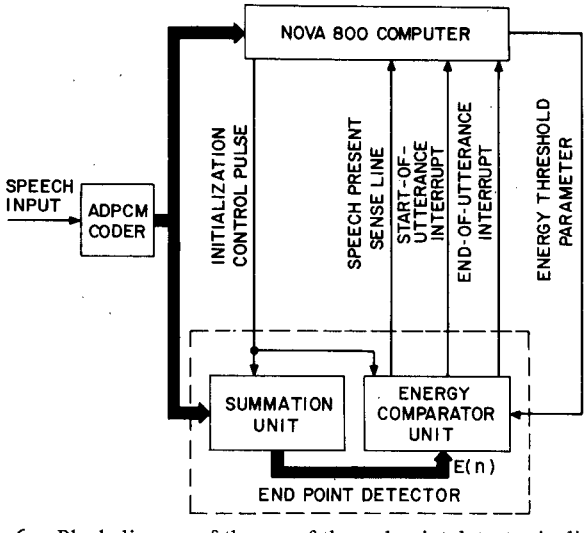


Fig. 6. Block diagram of the use of the end-point detector in digital recording of speech.

tem. The silences between words and phrases are thus automatically eliminated before storage in digital memory. Clearly, the configuration of Fig. 6 could also be used to control the recording of speech utterances for on-line speech-recognition or speaker-verification systems which could be implemented in the main computer.

REFERENCES

[1] H. Mieda and M. G. Schachtman, "TASI quality–Effects of speech detectors and interpolation," *Bell Syst. Tech. J.*, pp. 1455-1473, July 1962.

[2] J. A. Sciulli and S. J. Campanella, "A speech predictive encoding communication system for multichannel telephony," *IEEE Trans. Commun.*, vol. COM-21, pp. 827-835, July 1973.

[3] L. H. Rosenthal, L. R. Rabiner, R. W. Schafer, P. Cummiskey, and J. L. Flanagan, "A multiline computer voice response system utilizing ADPCM coded speech," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-22, pp. 339-352, Oct. 1974.

[4] P. Cummiskey, N. S. Jayant, and J. L. Flanagan, "Adaptive quantization in differential PCM coding of speech," *Bell Syst. Tech. J.*, vol. 52, pp. 1105-1118, Sept. 1973.

[5] L. H. Rosenthal, R. W. Schafer, and L. R. Rabiner, "An algorithm for locating the begining and end of an utterance using ADPCM coded speech," *Bell Syst. Tech. J.*, vol. 53, pp. 1127-1135, July-Aug. 1974.

[6] S. L. Bates and R. W. Schafer, "Digital hardware for PCM to ADPCM conversion" (Abstract), *Acoust. Soc. Amer.* (suppl.), vol. 56, Fall 1974.

[7] S. L. Bates, "A hardware realization of a PCM-ADPCM code converter," M.S. thesis, Dep. Elec. Eng., Massachusetts Inst. Technol., Cambridge, Jan. 1976.

Signal Design for Low-Error Probability in Fading Dispersive Channels

ERNESTO CONTE, MAURIZIO LONGO, AND EDUARDO MOSCA

Abstract–The signal design problem for FSK communication via fading dispersive channels is considered. The channel is modeled as a linear filter whose time-varying impulse response is a sample function from a zero-mean Gaussian random field of arbitrary WSSUS type. The additive noise component in the received waveforms is supposed to be a zero-mean white Gaussian random process, and maximum likelihood demodulation is assumed. The signal design procedure here adopted consists of minimizing a known upper bound on the error probability, whereas the previous similar design method by Daly intended maximizing an upper bound on the detection probability for radar-astronomy targets. Though with slightly different optimal numerical values, here, as in Daly's problem, the signal design depends on a single parameter which is a simple functional of the channel time-frequency covariance function and of the signal envelope ambiguity function. A detailed example shows how the results of this concise paper can be used to optimize signal parameters and to predict the performance loss due to nonoptimal signal envelopes.

I. INTRODUCTION

In this concise paper the signal design problem in fading dispersive communication channels is considered. Previous works in the area particularly relevant to the present investigation are those by Kennedy [1] and Daly [2]. The first consists of a detailed study concerning fading dispersive communication channels and their performance when high rate FSK signaling is used together with channel encoding. The second presents a systematic signal design method for detecting typical radar-astronomy targets exhibiting fading dispersive behavior or, equivalently, for "on-off" communication through fading dispersive channels. An extension of Daly's signal design method to binary transmitted-reference communication was also provided by one of the authors [3].

Paper approved by the Associate Editor for European Contributions of the IEEE Communications Society for publication without oral presentation. Manuscript received February 18, 1975; revised October 15, 1975. This work was supported in part by the C.N.R.

E. Conte and M. Longo are with the Istituto Elettrotecnico, Naples, Italy.

E. Mosca is with the Istituto di Elettronica, Florence, Italy.

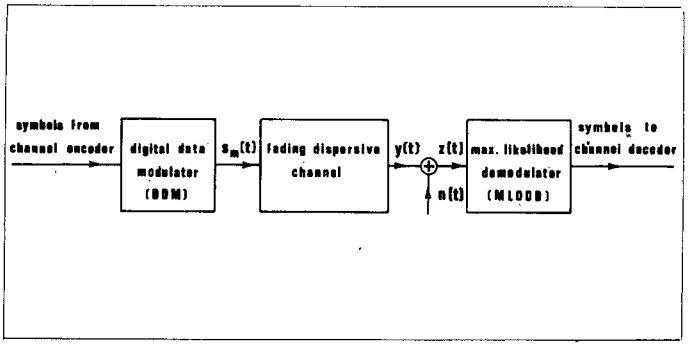


Fig. 1. The section of the communication system under consideration.

The aim of this concise paper is to show that a modified Daly's signal design method can be used for synthesizing the envelope of FSK signals in fading dispersive channels. This signal design method is referred as "modified" Daly's method in that whereas [2] concerns signal selection so as to maximize the Kullback-Leibler information number and hence an upper bound on the detection probability for a fixed false-alarm receiver, in the present study optimality is conceived with regard to minimizing an upper bound on the error probability per transmitted symbol.

II. FORMULATION OF THE PROBLEM

The section of the communication system that will now be considered is shown in Fig. 1. The digital data modulator (DDM), the physical channel, the noise source, and the digital data demodulator (DDD) make up the so-called discrete channel. The input to the discrete channel is a sequence of symbols from an M-ary alphabet delivered by the discrete channel encoder. The DDM maps in a one-to-one way each M-ary symbol at its input into one of the following FSK waveforms

$$s_m(t) \triangleq s(t) \exp(j2\pi mFt), \qquad m = 0,1,\cdots,M-1 \tag{1}$$

where $s(t)$ denotes an arbitrary complex envelope with energy E_s

$$E_s \triangleq \int_{-\infty}^{\infty} |s(t)|^2 \, dt.$$

The physical channel is modeled by the input-output description

$$y(t) = \int_0^{\infty} h(t,\tau)x(t-\tau)\, d\tau$$

where $x(t)$ is the input to the channel, $h(t,\tau)$ is the time-varying channel impulse response, which, in turn, is assumed to be a Gaussian random process in both variables with zero mean

$$E[h(t,\tau)] = 0, \qquad \forall t,\tau$$

and covariance

$$K_h(t,t',\tau,\tau') \triangleq E[h(t,\tau)h^*(t',\tau')] = \sigma_h(t-t',\tau)\delta(\tau-\tau').$$

The last equality is equivalent to assuming a wide sense stationary uncorrelated scatterer (WSSUS) channel which satisfactorily models the great majority of physical channels in the applications [5]. It is convenient to introduce another function related to $\sigma_h(t,\tau)$, namely the channel time-frequency covariance function

$$\sigma_H(t,f) \triangleq \int_{-\infty}^{\infty} \sigma_h(t,\tau)e^{-j2\pi f\tau}\, d\tau,$$

i.e., the Fourier transform of $\sigma_h(t,\tau)$ considered as a function of τ with fixed t. Because of the stochastic nature of the WSSUS channel the output signal $y(t)$, conditional to the input signal $x(t)$, is a Gaussian stochastic process with zero mean, $E[y(t)] = 0, \forall t$, and covariance

$$K_y(t,t') \triangleq E[y(t)y^*(t')] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \sigma_H(t-t',f-f')X(f)X^*(f') \cdot e^{j2\pi(ft-f't')}\, df\, df'$$

where $X(f)$ is the Fourier transform of $x(t)$. The signal $z(t)$ at the input of the DDD is given by

$$z(t) = y(t) + n(t)$$

where $y(t)$ is the channel output described above, and $n(t)$ is an additive white noise term with power spectral density equal to N_0 W/Hz, consisting of a Gaussian stochastic process with mean zero, $E[n(t)] = 0, \forall t$, and covariance

$$E[n(t)n^*(t')] = N_0\delta(t-t').$$

The DDD is assumed to be a miximum likelihood DDD (MLDDD) which has to choose one among M mutually exclusive, equiprobable hypotheses

$$H_m: z(t) = y_m(t) + n(t), \qquad m = 0,1,\cdots,M-1 \tag{2}$$

where $y_m(t)$ is the useful process at the channel output conditional to the transmission of $s_m(t)$.

It turns out that the use of a maximum likelihood demodulator together with the FSK signals (1) yields a symmetric discrete channel. Therefore, the performance of the communication system under study is describable by a unique parameter, namely the error probability, viz., the probability that the MLDDD chooses H_μ given H_m

$$P_e = \Pr\{H_\mu \mid H_m\}, \qquad \mu \neq m. \tag{3}$$

The problem to be solved is the following: given the set of FSK signals (1), the noisy fading dispersive channel and the

MLDDD of Fig. 1, find the complex envelope $s(t)$ so as to minimize P_e. Unfortunately, this problem does not appear mathematically tractable, and therefore our strategy will consist of choosing $s(t)$ so as to minimize an upper bound on P_e.

It is to be noticed that the hypothesis testing problem (2) concerns properly the "one-shot" case. By this, it is meant that a single signal is transmitted and a single decision is made by the MLDDD. On the other hand, in practical instances, signals $s_m(t)$ are sent sequentially over the physical channel which, because of its finite memory, may introduce intersymbol interference. Throughout this concise paper it will be assumed, as in [1], that intersymbol interference is eliminated by the use of a suitable technique, e.g., by frequency stepping successive signals $s_m(t)$ in a deterministic way. A further advantage introduced by frequency stepping consists of the suppression of the statistical dependence among adjacent signals. Due to channel fading this dependence would otherwise be present even in the absence of intersymbol interference. When the above assumptions are in force, the discrete channel of Fig. 1 becomes a discrete memoryless channel.

III. ERROR PROBABILITY OF MAXIMUM LIKELIHOOD DEMODULATION

The analysis in [1] and [4] of the communication system of Fig. 1 yields a tight upper bound on the error probability per symbol (3) corresponding to the M-ary FSK signaling (1), to an arbitrary WSSUS fading dispersive channel with Gaussian white additive output noise and to maximum likelihood demodulation, viz.,

$$P_e \leqslant \min_{0<\rho\leqslant 1} \{(M-1)^\rho \exp[-\Gamma(\rho)]\} \tag{4}$$

$$\Gamma(\rho) \triangleq \sum_{p=1}^{\infty} (1+\rho) \ln\left(1+\frac{\rho}{1+\rho}\frac{\lambda_p}{N_0}\right) - \ln\left(1+\frac{\lambda_p}{N_0}\right). \tag{5}$$

In (5) the λ_p's are the (nonnegative) eigenvalues of the covariance kernel $K_y(t,t')$ of the useful process $y(t)$ at the channel output conditional to the transmission of $s_0(t) = s(t)$, viz.,

$$K_y(t,t') \triangleq E[(y(t)y^*(t') \mid s_0(t)] = \sum \lambda_p \phi_p(t)\phi_p^*(t')$$

where $\{\phi_p(t)\}$ is a sequence of orthonormal functions

$$\int_{-\infty}^{\infty} \phi_p(t)\phi_r^*(t)\,dt = \delta_{pr}.$$

It is to be noticed that (4) and (5) are valid provided that the orthogonal modulator waveforms (1) retain their orthogonality after transmission through the channel; such a condition holds if the following inequality is fulfilled:[1]

$$F > W_s + B$$

where F is the minimal frequency separation of the FSK signals $s_m(t)$ as in (1), W_s is the frequency bandwidth of the envelope $s(t)$, and B is the maximum frequency spread produced by the channel.

Let $s(t)$ be the signal at the WSSUS channel input of given energy E_s; then the average energy E_y of the corresponding process $y(t)$ at the output depends only on E_s and not on the specific waveform $s(t)$. In fact,

$$E_y \triangleq E\left[\int |y(t)|^2\,dt \mid s(t)\right]$$

$$= \int_0^\infty d\tau \int_{-\infty}^{\infty} \sigma_h(0,\tau)\,|s(t-\tau)|^2\,dt$$

$$= E_s \int_0^\infty \sigma_h(0,\tau)\,d\tau = E_s \sigma_H(0,0).$$

Therefore, if it is assumed, as in (1), that E_s is fixed, E_y is a constant for all admissible $s(t)$. Furthermore,

$$E_y = \int_{-\infty}^{\infty} K_y(t,t)\,dt = \sum_{p=1}^{\infty} \lambda_p.$$

Thus, it is convenient to introduce the normalized eigenvalues

$$\gamma_p \triangleq \lambda_p / E_y$$

for which $\sum_{p=1}^{\infty} \gamma_p = 1$. Denoting by α a signal-to-noise ratio (SNR) at the MLDDD front end

$$\alpha \triangleq E_y / N_0 \tag{6}$$

(4) can be rewritten as follows

$$P_e \leqslant \min_{0<\rho\leqslant 1} (M-1)^\rho \exp[-\alpha E(\rho)]$$

where

$$E(\rho) \triangleq \frac{1}{\alpha} \sum_{p=1}^{\infty} \left[(1+\rho)\ln\left(1+\frac{\alpha\rho}{1+\rho}\gamma_p\right) - \rho \ln(1+\alpha\gamma_p)\right].$$

The remaining part of the paper will exclusively be concerned about the binary FSK signaling, viz., $M = 2$. Nevertheless, an approach similar to the one that will be developed hereafter can be used in connection with the general case $M \geqslant 2$. Setting $M = 2$, one has

$$P_e \leqslant \exp(-\alpha E_0)$$

where

$$E_0 = \max_{\{\gamma_p\}} \left\{ \max_{0<\rho\leqslant 1} E(\rho) \,\middle|\, \sum_{p=1}^{\infty} \gamma_p = 1 \right\}.$$

It is easy to show [1] that $\max_{0<\rho\leqslant 1} E(\rho) = E(1)$, and hence

$$E_0 = \max_{\{\gamma_p\}} \left\{ E(1) \,\middle|\, \sum_{p=1}^{\infty} \gamma_p = 1 \right\}$$

$$E(1) = \sum_{p=1}^{\infty} \gamma_p f_1(\alpha\gamma_p)$$

[1] See [1, pp. 72–74].

$$f_1(x) \triangleq \frac{2}{x}\ln\left(1+\frac{x}{2}\right)-\frac{1}{x}\ln(1+x).$$

The behavior of $f_1(x)$ is plotted in [1, p. 140] where it is shown that $f_1(x)$ takes on its maximum value at $x = 3.07$ and $f_1(3.07) = 0.15$. A direct consequence of this fact and of the form of $f_1(x)$ is the following theorem introduced in [1] and proved in detail for the present case in [4].

Theorem 1 (Optimality of equal strength diversity links): For a given energy of the admissible FSK binary ($M = 2$) signals (1), and hence for a fixed SNR α (6), the minimum value of the upper bound on the error probability P_e

$$P_e \leqslant \exp\left[-\alpha\Sigma_p \gamma_p f_1(\alpha\gamma_p)\right] \tag{7}$$

corresponds to a set of $\alpha/3.07$ nonzero eigenvalues of equal strength, more precisely,

$$\gamma_p = \left(\frac{3.07}{\alpha} \wedge 1\right), \qquad p = 1,2,\cdots,\left(\frac{\alpha}{3.07} \vee 1\right) \tag{8}$$

where $(a \wedge b) \triangleq \min\{a,b\}$ and $(a \vee b) \triangleq \max\{a,b\}$

Remarks

Remark 1: If $\alpha/3.07$ is not an integer number, owing to the fact that the maximum of $f_1(x)$ is quite broad, one can select the number of normalized eigenvalues equal to N_1, the largest integer less than $\alpha/3.07$ or equal to N_2, the smallest integer greater than $\alpha/3.07$; the choice between these two values is of little interest at least when $\alpha/3.07$ is $\gg 1$. For moderate values of $\alpha/3.07$ one can use the following decision rule:

$$f_1(\alpha/N_1) \underset{N_2}{\overset{N_1}{\gtrless}} f_1(\alpha/N_2).$$

Remark 2: Taking into account (8), (7) yields

$$P_e \leqslant \begin{cases} \exp(-0.15\alpha), & \alpha \geqslant 3.07 \\ \exp\left[-\alpha f_1(\alpha)\right] & \alpha \leqslant 3.07. \end{cases}$$

Remark 3: Notice that the λ_p's represent the degrees of freedom of the useful output process $y(t)$. Moreover, $\Sigma_{p=1}^{\infty} \lambda_p/N_0 = \alpha$. Hence $\alpha\gamma_p = \lambda_p/N_0$ can be referred to as the SNR per degree of freedom. Therefore Theorem 1 asserts that the optimum output process $y(t)$ is characterized by $(\alpha/3.07 \vee 1)$ degrees of freedom with an SNR per degree of freedom equal to $(3.07/\alpha \wedge 1)$. This result is in complete agreement with Pierce's conclusion [6] for the similar classical problem of allocating the available SNR over a number of diversity links with binary orthogonal signals over a fading channel. The similarity between the present and the classical result leads one to reinterpret the nonzero eigenvalues in the above theorem as representing implicitly equal strength diversity links, and talk about *implicit diversity*.

IV. OPTIMIZATION OF THE SIGNAL ENVELOPE

In order to find an optimum envelope $s(t)$ it is convenient to begin by introducing the autocorrelation function $R_y(t,\tau)$ of the useful process at the channel output conditional to the transmission of $s_0(t) = s(t)$

$$R_y(t,\tau) \triangleq E\left[y\left(t+\frac{\tau}{2}\right)y^*\left(t-\frac{\tau}{2}\right) \mid s_0(t)\right]$$

and the corresponding energy density function

$$S_y(t,f) \triangleq \int_{-\infty}^{\infty} R_y(t,\tau)e^{-j2\pi\tau f}\,d\tau.$$

It is assumed that the duration of $s(t)$ and the channel memory are finite. Consequently the process $y(t)$ can be considered to be identically zero outside the interval $[-T/2, T/2]$ with probability 1. Under these circumstances, it can be shown [4] that the energy density function $S_y(t,f)$ corresponding to the optimum set of eigenvalues in Theorem 1 equals the rectangular shape $P(t,f)$

$$S_y(t,f)_{\text{opt}} = P(t,f) \triangleq \begin{cases} C, & -T/2 \leqslant t \leqslant T/2 \\ & -W/2 \leqslant f \leqslant W/2 \\ 0, & \text{elsewhere} \end{cases} \tag{9}$$

where

$$C \triangleq \left(\frac{3.07}{\alpha} \wedge 1\right) E_y$$

$$W \triangleq \left(\frac{\alpha}{3.07} \vee 1\right) 1/T.$$

The problem of optimizing $s(t)$ has now become that of finding a complex envelope $s(t)$ yielding an energy density function $S_y(t,f)$ equal to $P(t,f)$ as defined by (9). However, since the rectangular shape of $P(t,f)$ is not realizable, a different strategy must be followed. An attempt can be made to find an $s(t)$ so as to minimize the normalized mean-square error between the optimal $P(t,f)$ and the actual $S_y(t,f)$ corresponding to a given $s(t)$

$$D(s) \triangleq \frac{\int_{-\infty}^{\infty} |P(t,f) - S_y(t,f)|^2\,dt\,df}{\int_{-\infty}^{\infty} P^2(t,f)\,dt\,df}$$

Unfortunately, even this optimization problem does not appear to lead to any intuitively appealing conclusion. Therefore, as in Daly's work [2], a lower bound on $D(s)$ may be established as follows and $s(t)$ can be chosen so as to minimize such a lower bound. Expanding the square and substituting for $P(t,f)$ its expression given in (9) we have

$$D(s) = 1 - \frac{2}{E_y}\int_{-T/2}^{T/2}\int_{-W/2}^{W/2} S_y(t,f)\,dt\,df + \frac{TW}{E_y^2}\iint S_y(t,f)\,dt\,df.$$

Using the Schwarz inequality we get the lower bound

$$D(s) \geqslant (1 - TWD_0)^{1/2} \tag{10}$$

where

$$D_0 \triangleq \frac{1}{E_y^2}\iint S_y(t,f)\,dt\,df = \iint |\sigma_H{}^0(\tau,\nu)|^2\,|\psi_S{}^0(\tau,\nu)|^2\,d\tau\,d\nu$$

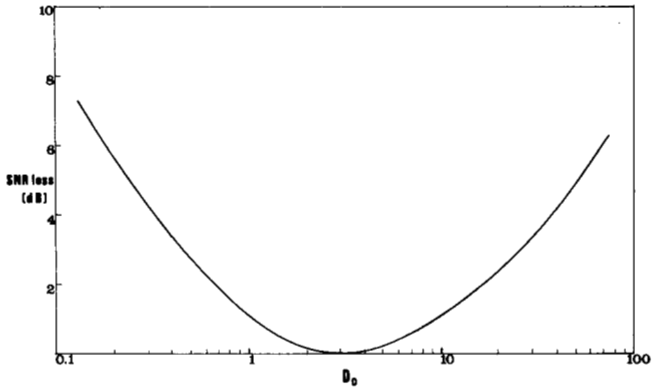


Fig. 2. Plot of $f_1(x)$.

$$\sigma_H{}^0(\tau,\nu) \triangleq \frac{\sigma_H(\tau,\nu)}{\sigma_H(0,0)}; \qquad \psi_S{}^0(\tau,\nu) \triangleq \frac{\psi_S(\tau,\nu)}{\psi_S(0,0)}$$

and $\psi_S(\tau,\nu)$ is the ambiguity function of the signal envelope $s(t)$

$$\psi_S(\tau,\nu) \triangleq \int_{-\infty}^{\infty} s\left(t+\frac{\tau}{2}\right) s^*\left(t-\frac{\tau}{2}\right) e^{-j2\pi\nu t}\, dt.$$

We sum up as follows.

Signal design procedure: Given the SNR α and the WSSUS channel time-frequency covariance function $\sigma_H(\tau,\nu)$, find an $s(t)$ minimizing the lower bound (10) on $D(s)$, i.e., an $s(t)$ such that

$$D_0 \triangleq \frac{1}{TW} = \left(\frac{3.07}{\alpha} \wedge 1\right)$$

where D_0 is given by (11).

Remark 4: Notice that for a given channel the signal design procedure depends only on the ambiguity function of $s(t)$. Even if it is usually very difficult to solve (11) exactly for a given D_0, nevertheless this equation gives a good insight into the basic relationship between some signal properties and their effects on the performances of the communication system. More precisely it can be stated, as in [4], that for low SNR's ($\alpha < 3.07$) the signal design procedure amounts to making $\psi_S(\tau,\nu)$ as much as possible similar to $\sigma_H(\tau,\nu)$; for moderate and high SNR's ($\alpha > 3.07$) $\psi_S(\tau,\nu)$ must be made different enough from $\sigma_H(\tau,\nu)$ so as to attain the optimum value $(3.07/\alpha)$ for D_0. In any case the signal design procedure is useful in the parametric optimization of the signal envelope as in the following example.

Remark 5: It can be shown [2], [4] that if the process $y(t)$ has N predominant eigenvalues λ_p of approximately equal strength λ, D_0 in (11) equals $1/N$ for

$$D_0 = \frac{\sum_{p=1}^{\infty} \lambda_p{}^2}{\left(\sum_{p=1}^{\infty} \lambda_p\right)^2} \cong \frac{N\lambda^2}{N^2\lambda^2} = \frac{1}{N}. \tag{12}$$

Therefore, since $TW = (\alpha/3.07 \vee 1)$ equals the optimum number of degrees of freedom of $y(t)$ (Remark 2), (12) indicates that one must choose $s(t)$ so as to make the actual number of predominant eigenvalues of $y(t)$ as close as possible to the optimum number of degrees of freedom. This is a reassuring conclusion indicating that the minimization of the lower bound (10) on $D(s)$ can effectively lead to a satisfactory design procedure.

As long as the hypothesis of equal strength predominant eigenvalues is satisfied at least approximately, (8) yields an estimate of P_e which is conjectured to be in reasonable agreement with the actual P_e [2]

$$P_e \propto \exp\left[-\alpha f_1(\alpha D_0)\right]. \tag{13}$$

In Fig. 2 the SNR loss is plotted as a function of D_0. This loss is given by $10(\log_{10} f_1(3.07) - \log_{10} f_1(\alpha_0))$ and represents the increase in SNR that, according to (13), is needed to compensate P_e for a nonoptimum value of αD_0. Hereafter, this loss will be referred to as the conjectured SNR loss.

Example: Let the modulus of the normalized time-frequency covariance function of the WSSUS channel be

$$|\sigma_H{}^0(\tau,\nu)| = \left[1 - \frac{|\tau|}{\theta}\right]_+ |\operatorname{sinc} \mu\nu| \tag{14}$$

where $[x]_+ \triangleq (x \vee 0)$ and $\operatorname{sinc} x \triangleq \sin \pi x/(\pi x)$. Roughly speaking, this corresponds to a channel with memory of μ seconds and with shortest fading circleof 2θ s. Let the modulus of the complex envelope $s(t)$ be a linearly frequency modulated rectangular pulse whose normalized ambiguity function is [7]

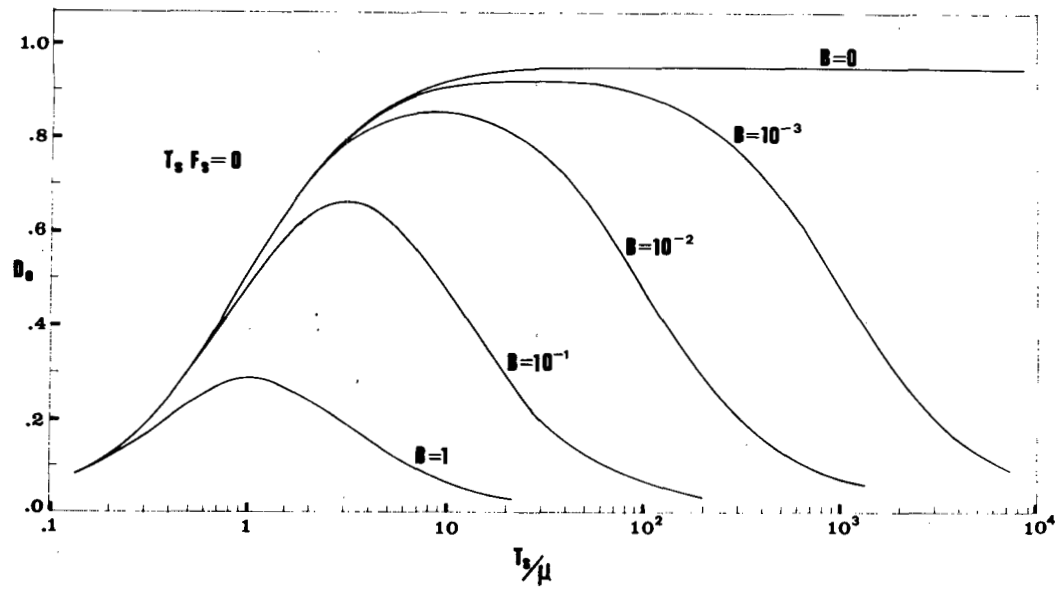
$$|\psi_S{}^0(\tau,\nu)| = \left[1 - \frac{|\tau|}{T_s}\right]_+ \cdot \left|\operatorname{sinc}\left[\left(\nu - \frac{F_s}{T_s}\tau\right)(T_s - |\tau|)\right]\right|. \tag{15}$$

In the above expression T_s represents the signal duration and F_s the total frequency deviation. D_0 was calculated introducing (14) and (15) in (11) as a function of T_s/μ and for several values of T_sF_s and $\mu/2\theta$. The corresponding plots are shown in Fig. 3(a)-(h). In order to show how these plots can be used, let the channel be such that

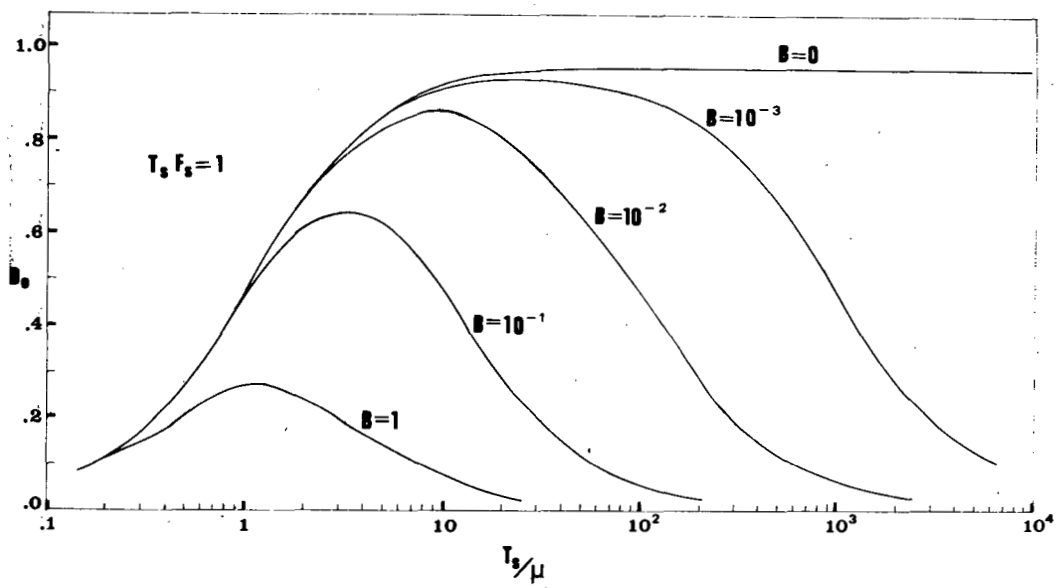
$$\mu/2\theta = 0.1$$

and assume that because of practical considerations the signal duration is chosen to be ten times the channel memory,

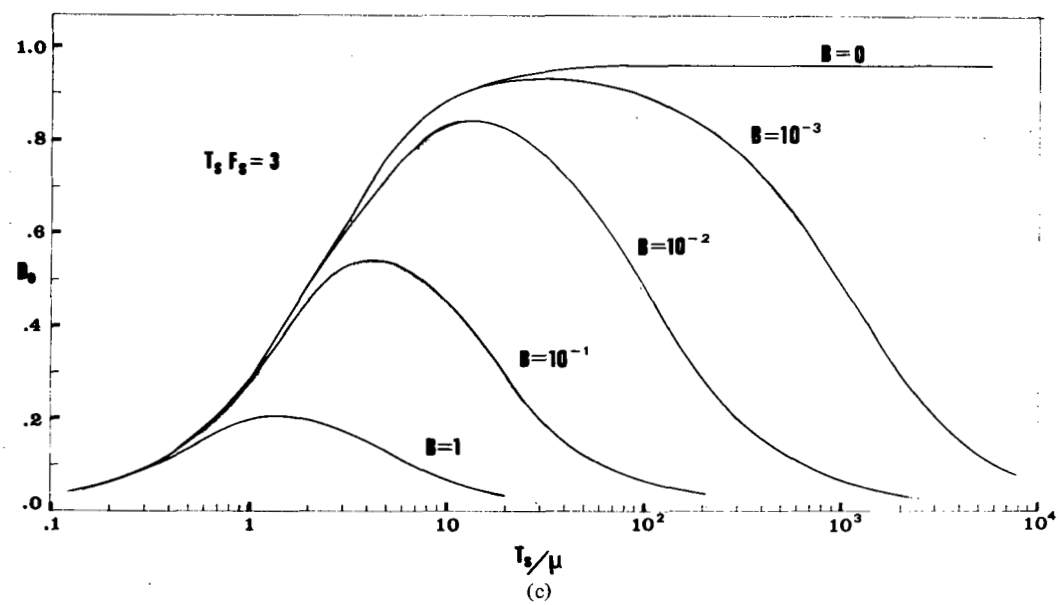
[2] Equation (13) only intends to indicate the dependence of P_e on the exponent and does not imply that the exponential must be multiplied by a unity factor to yield the numerical value of P_e. For more refined expressions of P_e the reader is referred to [1].



(a)

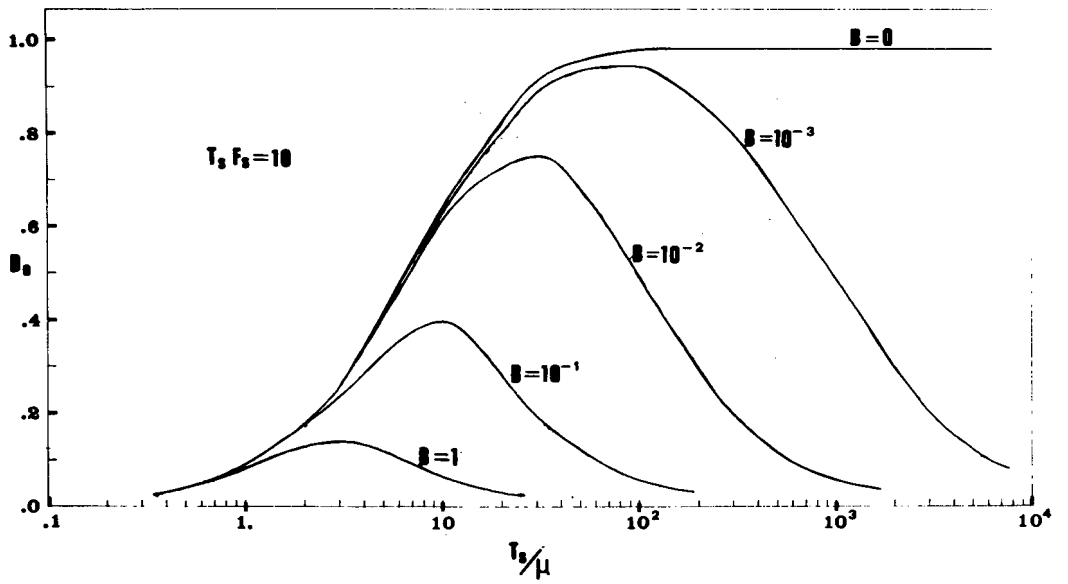


(b)

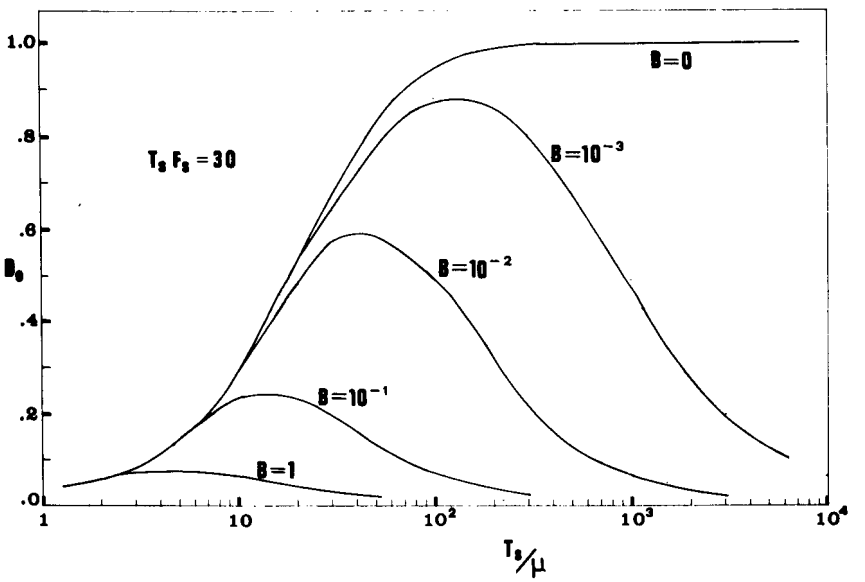


(c)

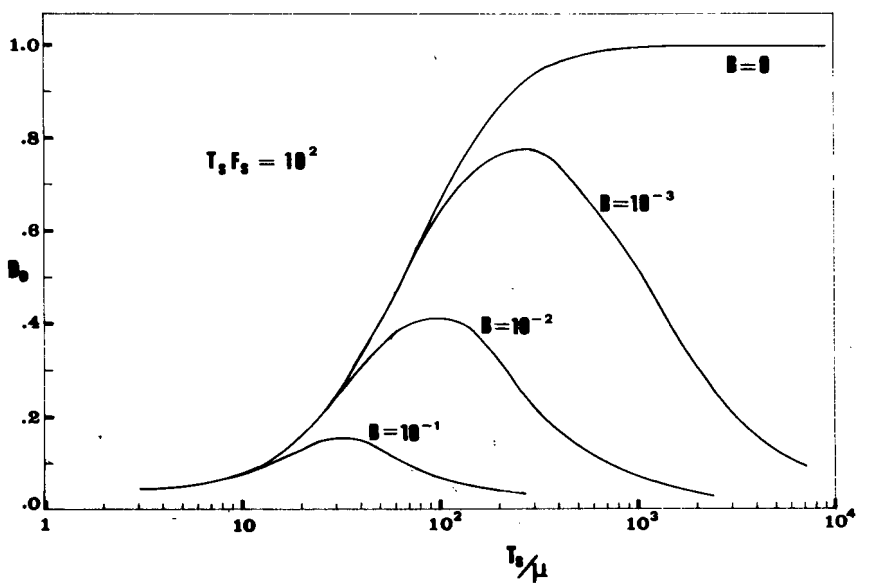
Fig. 3. Plots of D_0 versus T_s/μ for several values of T_sF_s and $B = \mu/2\theta$.



(d)

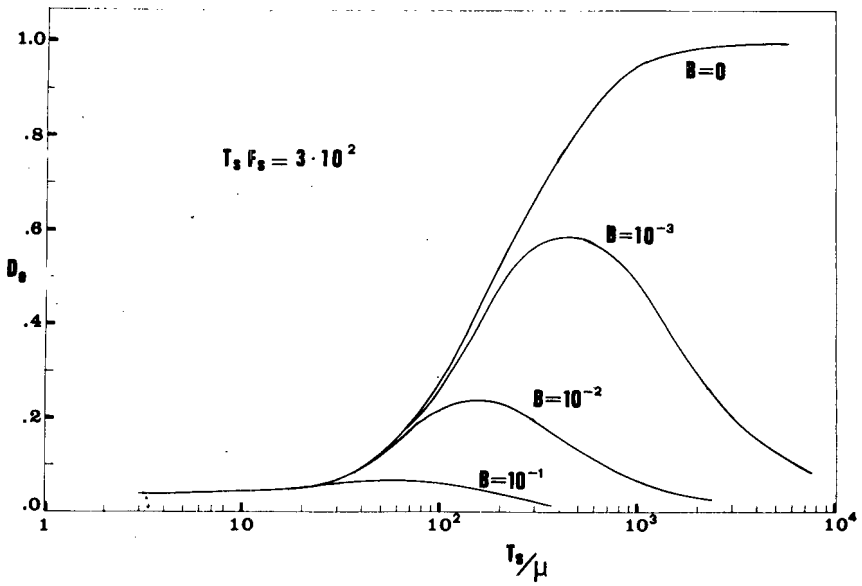


(e)

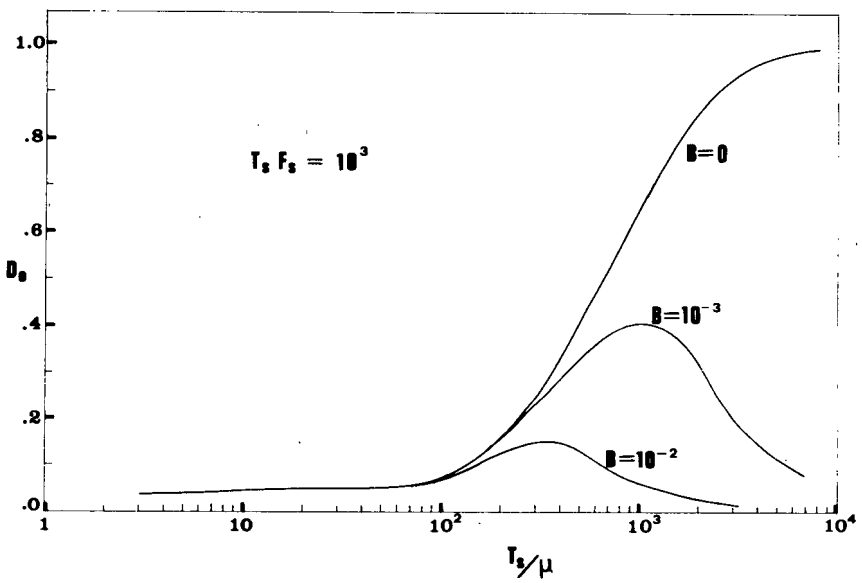


(f)

Fig. 3. (Continued.)



(g)



(h)

Fig. 3. (Continued.)

$T_s/\mu = 10$.

Moreover suppose that $\alpha = 10$, i.e.,

$3.07/\alpha \cong 0.3$.

Looking at the plots one finds that the optimum value 0.3 of D_0 is attained setting T_sF_s between 10 and 30, approximately

$T_sF_s \cong 20$.

If one sets $F_s = 0$, from Fig. 3(a) the corresponding value of D_0 can be found to equal 4.07. Hence, from Fig. 3 for $\alpha D_0 = 47$ the conjectured SNR loss is found to equal about 5 dB.

V. CONCLUSIONS

It is found that the modified Daly's signal design method leads to identify a single parameter, viz., D_0 in (11) relevant to the minimization of an error probability upper bound. For simple situations these results can be shown to agree with previous indications [8] of which the present work constitutes a definite refinement.

In spite of simple shape of the channel time-frequency covariance function the corresponding plots in Fig. 3(a)-(g) can provide good insight into the relative weight on the error probability played by the important signal parameters.

Though our study considers the optimization of FSK signal envelope in an incoherent mode of reception, the qualitative agreement among the results of this concise paper and previous works, such as [8]-[10], indicates that a suitably modified Daly's signal design method can be extended even to different signaling schemes.

After completion of the present paper another work in the area [10] was published whose main emphasis is on deriving the probability of error for ON-OFF communications over a general (not necessarily WSSUS) time-varying dispersive channel. Whereas this greater generality allows one to compute error probabilities for a wide class of possible channels, limited insight is provided into the signal design problem. The WSSUS assumption, for instance, removes the stumbling block which in the general case prevents one from minimizing the probability of error under a transmitted energy constraint.

REFERENCES

[1] R. S. Kennedy *Fading Dispersive Communication Channels.* New York: Wiley, 1969.
[2] R. F. Daly "Signal Design for Efficient Detection in Dispersive Channels" *IEEE Trans. Inform. Theory,* 206-213, Mar. 1970.
[3] E. Mosca "Signal Selection in Transmitted Reference Communication Systems" 1972 Int. Conf. on Communications, Philadelphia, Penn., June 1972.
[4] E. Conte, M. Longo, and E. Mosca "A Procedure for Signal Selection for Digital Communications via Fading Dispersive Channels" submitted for publication to *Alta Frequenza.*
[5] P. A. Bello "Characterization of Randomly Time-Variant Linear Channels" *IEEE Trans. Communications Systems,* 360-393, Dec. 1963.
[6] J. N. Pierce "Theoretical Limitations on Frequency and Time Diversity for Fading Binary Transmissions," *IRE Trans. Comm. Systems,* 186-189, June 1961.
[7] H. L. Van Trees *Detection, Estimation and Modulation Theory,* vol. III, page 292, New York: Wiley, 1971.
[8] R. S. Kennedy and I. L. Lebow "Signal Design for Dispersive channels" *IEEE Spectrum,* 231-237, March 1967.
[9] P. Monsen "Digital Transmission Performance on Fading Dispersive Diversity Channels" *IEEE Trans. Communication Systems,* 33-39, Jan. 1973.
[10] S. K. Chow and A. N. Venetsanopoulos: "Optimal On-Off Signaling over Linear Time-Varying Stochastic Channels" *IEEE Trans. on Information Theory,* 602-609, September 1974.

Correspondence

Proof of a Conjecture on the Interarrival-Time Distribution in an $M/M/1$ Queue with Feedback

P. J. BURKE

Abstract—**In a Jackson-type queuing network with feedback, the equilibrium state distribution of each queue is that of an $M/M/s$ system. In support of a previous conjecture that nevertheless the input processes in such a network are not Poisson, the marginal interarrival-time distribution for an equilibrium $M/M/1$ queuing system with feedback, counting both fed-back and exogenous customers as arrivals, is calculated. Since this distribution is a mixture of two exponentials, the total input to such a system is not Poisson.**

We consider a queuing system in which a customer (call) rejoins the queue with fixed probability p at the conclusion of his service. All distributions involved are equilibrium distributions. The calls originally arrive at the queue in a Poisson stream and are served by a single exponential server. The "total" input process (input) consists of the superposition of the stream of originally arriving (exogenous) calls and the stream of calls rejoining the queue. Since these are not independent, the statistical character of the input is not evident.

The state of the system is defined as the number of calls in the system waiting or being served. It is known that the state distribution of the queuing system described above is the same as that of an ordinary $M/M/1$ system with an enhanced traffic parameter. In fact, in a network of exponential-service multiserver queues with feedback, where the fed-back calls go to any queue in the network with fixed probabilities depending only on the source and target queues, and where the exogenous inputs are independent Poisson processes, the state distribution of each queue is that of an ordinary $M/M/s$ queue without feedback but with traffic parameters depending on the feedback probabilities. (This is a result of Jackson [1].) The point is that, in spite of the presence of feedback, the queues behave, in certain respects, as if the total input processes were Poisson. In this connection the present author wrote [2], ". . . The combined input to a queue, new arrivals and returning customers, is apparently not Poisson." This conjecture was supported only by a heuristic argument, which was not altogether convincing to some readers. Hence, it was thought worthwhile to provide a rigorous proof that the input is not Poisson in the simplest case of a Jackson-type network, namely, a single $M/M/1$ queue with feedback. The proof takes the form of a calculation of the marginal distribution of an interarrival interval of the input of such a queue, and it provides as a byproduct the explicit form of this distribution.

Before proceeding with the calculation it is well to make the point explicitly that fed-back calls, and hence also the total input stream, see the same state distribution as the exogenous calls. That is, we assert the proposition that the state distribution at arbitrary arrival instants, fed-back or exogenous, is the same as that at the exogenous arrival instants alone. To establish this proposition we recall first that in any equilibrium queuing system with individual arrival and service, the departing (not fed-back) calls leave the same state distribution as seen by exogenous calls [3]. Next we observe that the fed-back calls leave the same state distribution as the departing calls since calls which are fed back (or which leave) are selected with fixed probabilities (thus independently of the state). Since fed-back calls see the same state that they leave, the proposition follows.

In addition to p, the feedback probability, let

λ original arrival rate
μ service rate
π_i equilibrium probability of state i at an arbitrary instant
q $= 1 - p$.

It is known [1] that, if $\lambda/\mu q < 1$,

$$\pi_i = [1 - (\lambda/\mu q)](\lambda/\mu q)^i, \tag{1}$$

Paper approved by the Associate Editor for Computer Communication of the IEEE Communications Society for publication without oral presentation. Manuscript received November 20, 1975; revised January 6, 1976.

The author is with Bell Laboratories, Holmdel, NJ 07733.