Voice-band Modems: A Device to Transmit Data Over Telephone Networks

1. Basic Principles of Data Transmission

V Umapathi Reddy

V U Reddy is with the Electrical Communication Engineering Department, Indian Institute of Science. His research areas are adaptive signal processing, multirate filtering and wavelets, and multicarrier communication.

Over the last 40 years, there has been continuous evolution in the design of voice-band modems – starting at a data rate of 300 bits per second in late 1950s, a rate of 33,600 bits per second has been achieved in 1995. Realising such high data rates over the voice band of 3400 Hz is a remarkable feat made possible by combining sophisticated techniques from three disciplines, communication theory, signal processing and information theory. In this article, we begin with a brief introduction of voice-band modems and then introduce the basic principles of data transmission.

Introduction

Voice-band modems were introduced in the late 1950s to transmit data through the public switched telephone network (PSTN). *Figure* 1 shows how two computer terminals, located far apart, can talk to each other with the help of modems. The word modem is coined from modulator-demodulator. Since PSTN does not trans-

Box 1. The Twisted-Wire-Pair

The twisted-wire-pair telephone line, which is laid between the central office and the subscriber's premises can support a much larger bandwidth than 3400 Hz used for telephone communication. The bandwidth restriction of 3400 Hz over public switched telephone network is imposed by multiplexing equipment on the network side of the central office. The so-called digital subscriber line technology aims at utilizing the larger bandwidth of the twisted-wire pair to transmit the data at rates much beyond those possible with voice-band modems. But, this is beyond the scope of this article.

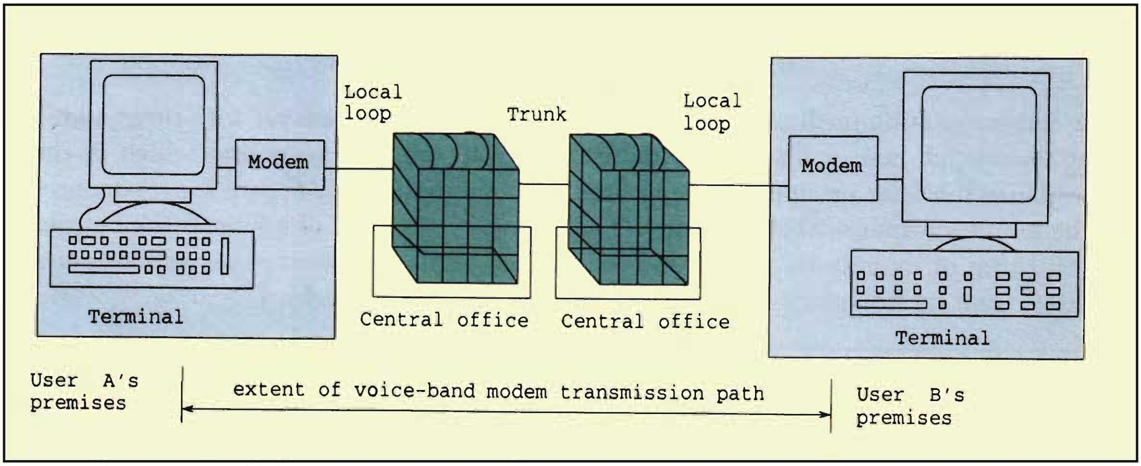


Figure 1. Voice-band modem reference model. (From Starr and others, Understanding Digital Subscriber Line Technology, Prentice-Hall,1999.)

port frequencies below approximately 200 Hz, the data needs to be modulated. Modulation, broadly speaking, is a process which maps the data into a frequency band which the communication medium supports, and demodulation is the inverse process of modulation.

The first voice-band modem, used for interconnecting computers was the Bell 202, which provided a data rate of 1200 bps (bits per second). This was a four-wire modem, where two two-wire pairs are used in simplex mode (i.e., transmission takes place one way permanently), one for each direction. The recent V.34 modem provides a data rate of 33,600 bps over classical two-wire dial-up telephone lines. This is a two-wire full-duplex modem (i.e., sends data continuously in both directions on the same wire-pair). By transmitting 33,600 bps in 3.6 kHz voice band, V.34 modems send nearly 10 bps/Hz, which is a remarkable feat in the sense that it approaches the theoretical limit. To achieve this remarkable rate, advanced developments from three disciplines, communication theory, information theory and signal processing, have been combined.

Modulation, broadly speaking, is a process which maps the data into a frequency band which the communication medium supports, and demodulation is the inverse process of modulation.

A Brief Description of Data Transmission Principles

Voice-band modem uses passband data transmission, where a baseband (band around zero frequency) infor-

Box 2. Communication Channel

A communication medium or channel is usually modelled as a filter with noise added at its output, and a filter is characterized by its impulse response, which is the response the filter produces when excited with an impulse. In *Figure* 2, $h(t)$ denotes the impulse response of the channel. The frequency response of a filter is the Fourier transform of its impulse response. For example, if $g(t)$ is the impulse response of a filter, then its frequency response, denoted by $G(f)$, is given by $\int_{-\infty}^{\infty} g(t)e^{-j2\pi ft}dt$.

mation signal is converted, by modulation, into a passband (band away from zero frequency) signal which can transit a passband channel, and a demodulation is performed in the receiver to recover the baseband information signal. *Here, the channel refers to physical communication medium, which is the telephone line in the case of voice-band modems.* A passband transmission system can, however, be reduced to an equivalent baseband system and the performance of the former can be studied from that of the latter. Here, we use a baseband transmission model and for ease of exposition, digital PAM (pulse amplitude modulation) in presenting the basic principles of data transmission.

Figure 2 gives a general baseband signaling model where the raw binary data $\{a_n\}$ (a_n takes values from the alphabet $\{1, 0\}$) is encoded into PAM symbol sequence $\{d_n\}$.

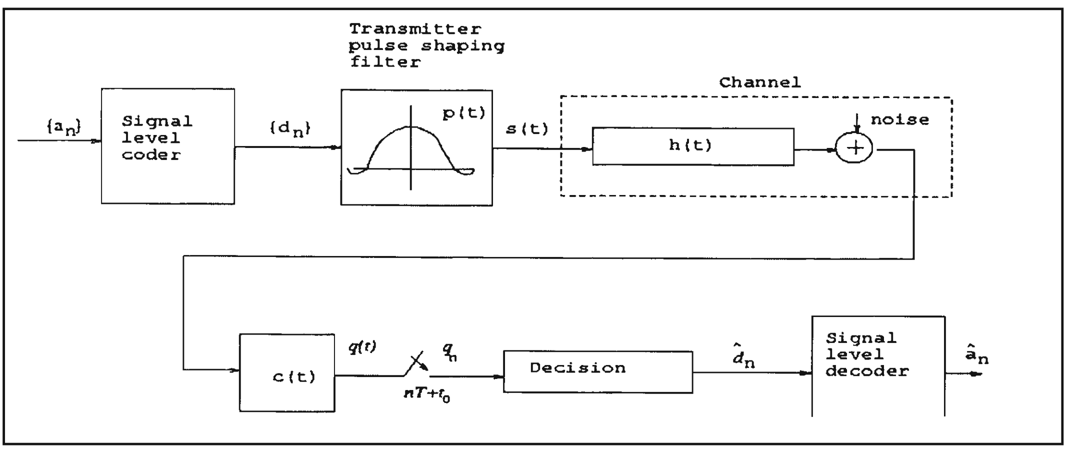


Figure 2. A general baseband signaling model.

An example of such encoding is shown in *Figure* 3 for an L-ary PAM with L=8. Here, each block of three successive binary digits of raw data is mapped to one of the eight amplitude levels. *Note that the mapping of 3 binary digits to 8 possible signal amplitudes is such that the adjacent signal amplitudes differ by one binary digit. This mapping is called Gray encoding. It is preferred because the most likely errors caused by noise involve erroneous selection of an adjacent amplitude to the transmitted signal amplitude which leads to only a single bit error.* $p(t)$ is the pulse which the transmitter filter produces when excited with an impulse (selection of a suitable pulse shape will be explained later). The transmitted signal is then given by

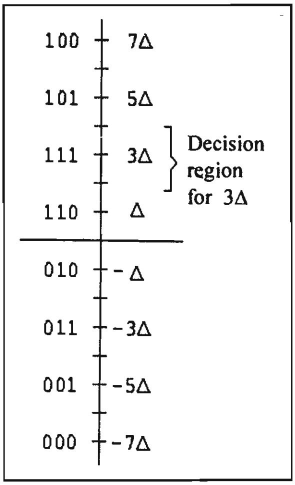


Figure 3. Pulse amplitude levels chosen from 8-level alphabet.

$$s(t) = \sum_{k=0}^{\infty} d_k\, p(t-kT), \qquad (1)$$

where T denotes the symbol interval, i.e., the time interval between the successive symbols d_k, and the symbol rate is $1/T$ The number of bits carried by each symbol, in L-ary PAM, is $\log_2 L$. The bit rate is then given by $\frac{1}{T}\log_2 L$. The channel noise is assumed to be white, i.e., the noise power is uniform over all frequencies, and Gaussian. The decision circuit will map q_n to one of the amplitudes depending upon its value. For example, if q_n falls in the decision region for the amplitude level 3Δ (see *Figure* 3), then it is mapped to 3Δ. The signal level decoder decodes the estimated symbol $\hat{d}_n$ into the corresponding data bits. One of the factors, which determines the fraction of the transmitted symbols that are received correctly on average, is the spacing between the symbols corresponding to the closest amplitude levels (which is 2Δ in *Figure* 3). But note that, larger the Δ, higher is the average transmitted power.

Selection of $p(t)$ and $c(t)$: Equation (1) shows that the symbols d_k are the amplitudes of the transmitted pulse $p(t)$. The bandwidth of the pulse $p(t)$ should be

Any distortion introduced by the channel, within the channel bandwidth, can be corrected by suitably choosing $c(t)$. But the distortion, caused because the channel bandwidth is not adequate to support the pulse $p(t)$, cannot be compensated by any $c(t)$.

compatible with that of the channel. Otherwise, the received signal at the input of the receiver filter $c(t)$ will be severely distorted. Any distortion introduced by the channel, within the channel bandwidth, can be corrected by suitably choosing $c(t)$. But the distortion, caused because the channel bandwidth is not adequate to support the pulse $p(t)$, cannot be compensated by any $c(t)$.

We now motivate the design of $p(t)$ and $c(t)$. Consider *Figure* 2. Let $x(t) = p(t) \star h(t) \star c(t)$ where $\star$ denotes convolution operation. *Output of a filter is the convolution of its impulse response with the input to the filter. For example, if $f(t)$ is the input to the receiver filter $c(t)$, then its output at any time t is given by $\int_{-\infty}^{\infty} f(\tau)c(t-\tau)d\tau$.*

In terms of $x(t)$, we can write

$$q(t) = \sum_{k=0}^{\infty} d_k x(t-kT) + v(t), \qquad (2)$$

where $v(t)$ is the response of the receiver filter to noise. Sampling $q(t)$ at time instants $nT + t_o$, $n = 0,1,\ldots,$ we get

$$q(nT+t_o) \equiv q_n = \sum_{k=0}^{\infty} d_k\, x(nT+t_o-kT)+v(nT+t_o), \quad (3)$$

which can be expressed as

$$q_n = d_n x_0 + \sum_{\substack{k=0 \\ k\neq n}}^{\infty} d_k x_{n-k} + v_n, \qquad (4)$$

where $x_n \equiv x(nT + t_o)$, and similarly v_n. If $x(t)$ is such that $x_k = 0$ for $k \neq n$, then q_n is a scalar multiple of the transmitted symbol d_n (when noise is absent). If $x(t)$ does not satisfy this condition, the second term in the right-hand side of (4) will not be zero, which will then interfere with the symbol d_n. *Since this interference arises from the symbols around the symbol under consi-*

deration, it is called the intersymbol interference (ISI).

One possible $x(t)$ which satisfies the condition for zero ISI is

$$x(t) = \frac{\sin \pi(t - t_o)/T}{\pi(t - t_o)/T}. \quad (5)$$

Figure 4 shows $x(t)$ and $x(t - T)$. Note that the samples of $x(t)$ at $kT + t_o$, $k \neq 0$, are zero. *We should point out here that we ensure zero ISI only when the sampling instants are synchronised with symbol timings at the receiver.* The corresponding frequency response (i.e., the Fourier transform of $x(t)$) with $T = 1/2W$ is given by

$$X(f) = \begin{cases} Te^{-j2\pi f t_o} & |f| \leq W \\ 0 & \text{otherwise.} \end{cases} \quad (6)$$

The reason for sampling at $t = nT + t_o$ other than at $t = nT$ is that the delay and distortion in the channel filter $h(t)$ might make some other sampling instants preferable.

The difficulty with the choice of $x(t)$, given in (5), is that $x(t)$ extends from $t = -\infty$ to $t = \infty$ (see *Figure* 4),

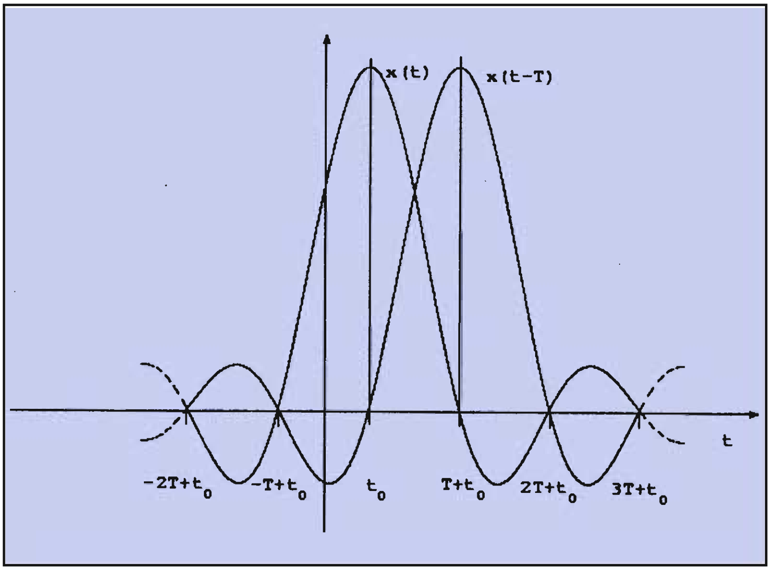


Figure 4. Condition for avoidance of ISI – the pulses equal zero at all sampling times except their own.

and no amount of delay can make it causal, i.e., $x(t - T_D) \neq 0$ for $t < 0$ for any finite value of T_D. Such signal cannot be generated by any real-life filter. The same conclusion will be arrived from the frequency response $X(f)$, given in (6), which is of brick wall shape. Such shape is unrealisable with any real-life filter.

To design a realisable $x(t)$, a certain amount of excess bandwidth is generally used. A particular pulse, for $1/T < 2W$, that has desirable properties and has been widely used in practice, is the raised cosine spectrum

$$X_{rc}(f) = \begin{cases} T & 0 \leq |f| < \frac{1-\beta}{2T} \\ \frac{T}{2}\left(1 - \sin\left(\frac{|2\pi fT| - \pi}{2\beta}\right)\right) & \frac{1-\beta}{2T} \leq |f| \leq \frac{1+\beta}{2T} \\ 0 & |f| > \frac{1+\beta}{2T} \end{cases} \tag{7}$$

and the corresponding time-domain pulse is given by

$$x_{rc}(t) = \frac{\sin(\pi t/T)}{\pi t/T} \frac{\cos(\pi\beta t/T)}{1 - 4\beta^2 t^2/T^2}, \tag{8}$$

where β is the rolloff factor (also called the excess bandwidth factor). *Figure* 5 illustrates the raised cosine pulse and its spectrum for several values of β. Note that $x_{rc}(t = nT) = 0$ for $n \neq 0$ (t_o is assumed to be zero here) as required for zero ISI.

The bandwidth occupied by the signal beyond the frequency $1/2T$ is called the excess bandwidth. For example, $\beta = 0.5$ corresponds to 50% excess bandwidth.

We may point out here that the raised cosine pulse given in (8), strictly speaking, is not of finite duration even for $\beta > 0$, i.e., $x_{rc}(t - T_D) \neq 0$ for $t < 0$ for any finite value of T_D. However, since the tails of the pulse decay as $1/t^3$ for $\beta > 0$, the pulse may be considered to be of finite duration, for all practical purposes, when excess bandwidth is used.

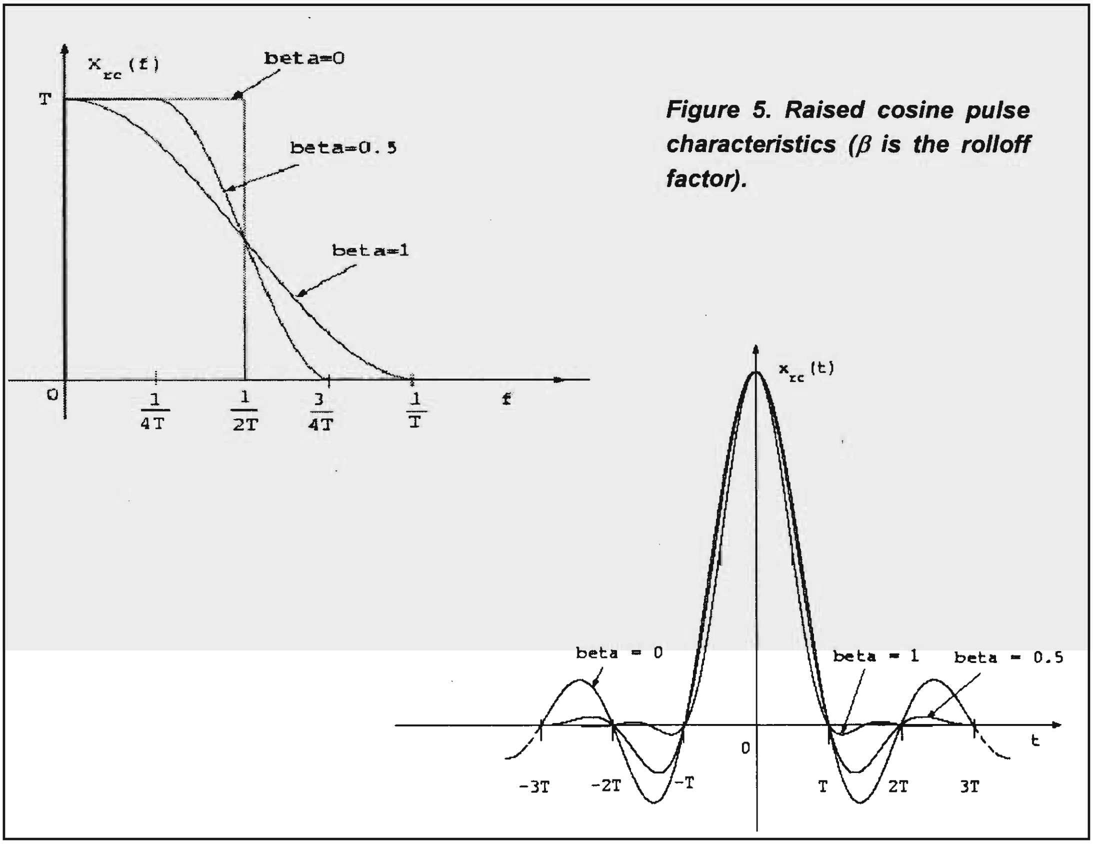


Figure 5. Raised cosine pulse characteristics (β is the rolloff factor).

We now consider the design of $p(t)$ and $c(t)$ under two different scenarios: (i) non-ideal, but known channel and (ii) non-ideal unknown channel. In each case, the bandwidth of the channel is restricted to W Hz.

Case 1: The channel is non-ideal, but known. To design $p(t)$ and $c(t)$ for zero ISI, we need to satisfy

$$P(f)H(f)C(f) = X_{rc}(f)e^{-j2\pi f t_d}, \quad |f| \leq W. \quad (9)$$

We then choose $C(f)$ such that the output signal-to-noise ratio (SNR) at the sampling instant is maximized. This gives

$$|C(f)| = K_1 \frac{|X_{rc}(f)|^{1/2}}{|H(f)|^{1/2}}, \quad |f| \leq W, \quad (10)$$

and the corresponding $P(f)$ as

$$|P(f)| = K_2 \frac{|X_{rc}(f)|^{1/2}}{|C(f)|^{1/2}}, \quad |f| \le W, \tag{11}$$

where K_1 and K_2 are arbitrary scale factors. The delay t_d is selected to satisfy

$$\Theta_p(f) + \Theta_h(f) + \Theta_c(f) = 2\pi f t_d, \tag{12}$$

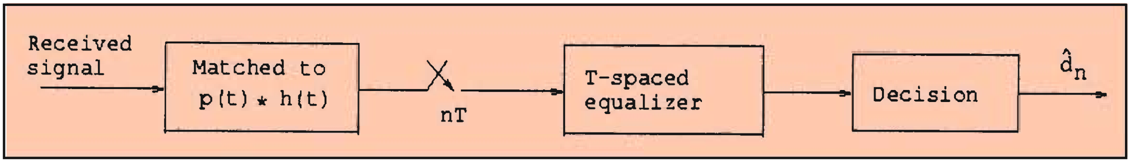
where $\Theta_p(f)$, $\Theta_h(f)$ and $\Theta_c(f)$ are the phase responses of the transmitter pulse shaping filter, channel and receiver filter, respectively. The phase response of a filter is defined as follows.

The frequency response of a filter, in general, is complex-valued, and one possible representation of this is the polar form. That is, if $G(f)$ is the frequency response of a filter, then $G(f) = |G(f)|e^{-j\Theta_g(f)}$, where $\Theta_g(f)$ is the phase response of the filter.

The problem with the above design is that when the channel suffers from a large attenuation in a band of frequencies within the bandwidth of W Hz (which is likely in practice), the receiver filter will have a large gain in that band since $|C(f)|$ is inversely proportional to $|H(f)|$ (see (10)), which will enhance the input noise thereby resulting in SNR loss at the output.

A preferred approach in this case is to choose a raised cosine shape for $p(t)$ and select $c(t)$ as a matched filter matched to $p(t) \star h(t)$, followed by a symbol-spaced (T-spaced) equalizer (see *Figure* 6). The popular form of equalizer, currently employed in voice-band modems, is a linear equalizer and its coefficients are chosen based on certain minimum mean square error criterion. The

Figure 6. A receiver structure when the channel is non-ideal, but known.



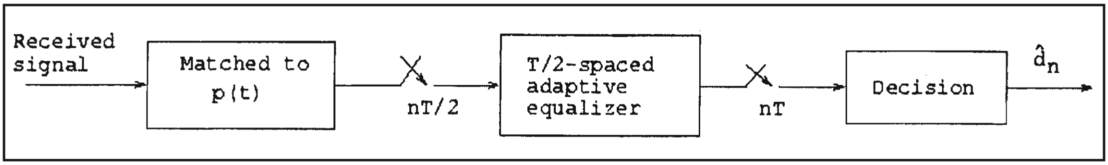


Figure 7. A receiver structure when the channel is non-ideal and unknown (a desired structure in practice).

role of the equalizer is to minimize the ISI component (second term on the right hand side of (4)) which arises because the overall response, $p(t) \star h(t) \star c(t)$, is not guaranteed to be a raised cosine pulse. However, when the noise is absent, the T-spaced equalizer with infinite length will eliminate the ISI completely.

Case 2: The channel is non-ideal and unknown. A preferred approach in this case is to choose $p(t)$ as a raised cosine pulse and the receiver filter $c(t)$ as a matched filter matched to $p(t)$, followed by a $T/2$-spaced adaptive equalizer (see *Figure* 7). The front-end filter can now be viewed as a noise-limiting filter. The equalizer is made adaptive since we do not have the knowledge of the channel which is required in designing the fixed equalizer. Also, we use $T/2$-spaced equalizer because the performance of the T-spaced equalizer will be very sensitive to the sampling instants when the front-end filter is not matched to $p(t) \star h(t)$. On the other hand, this sensitivity is much less with $T/2$-spaced equalizer. The equalizer coefficients are chosen using an adaptive algorithm and the most popular adaptive algorithm is the so-called least mean square (LMS) algorithm. We may point out here that since we do not have the exact knowledge of the channel in practice, and it may not be stationary, i.e., the channel response may vary slowly with time, a receiver with $T/2$-spaced adaptive equalizer is used in practical systems.

Suggested Reading

[1] J G Proakis, *Digital Communications* (third edition), McGraw-Hill, 1995.

[2] R D Gitlin, J F Hayes and S B Weinstein, *Digital Communication Principles*, Plenum Press, 1992.

Address for Correspondence
V Umapathi Reddy
Electrical Communication Engineering
Indian Institute of Science,
Bangalore 560012, India.
Email: vur@ece.iisc.ernet.in