

AN ADAPTIVE-RATE DIGITAL COMMUNICATION SYSTEM FOR SPEECH

J. E. Kleider and W. M. Campbell

Motorola SSTG, Scottsdale, AZ, USA

ABSTRACT

Current digital voice communication systems allow only modest levels of protection of the coded speech and often do not follow the dynamic changes that occur in the transmission channel. We present a method that provides optimal voice quality and intelligibility for any given transmission channel condition. The approach is performed via adaptive-rate voice (ARV) coding using an adaptive-rate modem, channel coding, and a multimode sinusoidal transform coder. In general, the receiver utilizes channel state information to not only optimally demodulate and decode the currently corrupted symbols from the channel, but also to inform the transmitter, via a feedback channel, of the optimal strategy for voice/channel coding and modulation format. We compare several source-channel coding schemes at multiple transmission symbol rates and compare the performance to fixed aggregate-rate channel-controlled variable rate voice coding systems.[†]

1. INTRODUCTION

In most present-day digital voice communication systems, the available channel bandwidth is divided between a fixed-rate voice coder, a fixed-rate channel coder (forward error control, FEC), and system control functions (protocols). The bit allocations for the channel coding are designed based on some "nominal" channel conditions. Often, however, the transmission channel experiences degradations, such as fading, far worse than the nominal conditions, and the speech quality degrades severely. Past work [1, 2] has suggested channel-controlled variable rate (CCVR) coding as one possible solution to extending the speech quality for low received signal-to-noise ratios. In such a system, a bit rate allocation tradeoff is performed between the voice coding and the FEC, which is based on the state of the channel. The performance increase of CCVR, however, can be limited by the fixed aggregate-rate of the system. Performance can be achieved by more closely matching the modulation strategy to the available channel capacity.

We show that further performance gains are achieved by utilizing adaptive-rate modulation and demodulation (ARM). Generally, if the modem symbol rate, R_s , is cut in half to $R_s/2$, the system requires 3 dB less received signal power, P_r , to achieve the same received signal-to-noise ratio, E_s/N_o (or E_b/N_o), at a given bit error rate (BER). Our approach uses 3 symbol rate changes, R_s to $R_s/4$, to provide an improvement in the required received signal power, P_r , to achieve the same received E_b/N_o . The algorithm is applied to both bandwidth efficient (PSK) and power efficient (PPM) modems.

The method maximizes the synthesized speech quality using the sinusoidal transform voice coder [3] at multiple bit rates by adapting the system in an optimal fashion to the degrading communication channel. The channel model used is based on complex additive white gaussian noise (AWGN) and Rayleigh fading. The ARM produces an improvement in BER versus relative received signal power, which translates into significantly lower average spectral distortion of the vocoder. When used in conjunction with channel coding and a multimode vocoder, a large increase in speech quality at the receiver is produced.

The adaptive-rate system discussed herein is shown in Figure 1. Both objective and subjective speech quality measures are used for system evaluation. Spectral distortion of the multimode STC vocoder (MMSTC) is used for the objective measure, while informal listening tests are used to provide subjective feedback of the adaptive-rate synthesized speech quality. Performance enhancements provided by this technique are shown by comparison to a fixed aggregate-rate channel-controlled variable rate speech and channel coded system. The comparison system aggregate rate is 9.6 kbps, similar to the system proposed in [1]. However, our system uses multiple maximal-distance-property convolutional codes for error protecting the vocoder parameters.

Experimental results indicate that the adaptive-rate technique discussed here produces a significant increase in speech quality in the AWGN channel, and roughly equivalent speech quality in the Rayleigh fading channel as opposed to the fixed aggregate-rate configuration. For the AWGN channel, a reduction of 3 dB is observed in the average received signal power, with no increase in spectral distortion. For the Rayleigh channel, equivalent system performance is demonstrated, while system complexity is reduced significantly. Informal listening tests support these results.

Section 2 describes further each block of the adaptive system of Figure 1, with the experimental results presented in Section 3. Conclusions are given in Section 4.

2. SYSTEM DESCRIPTION

A detailed description of each block in Figure 1 is given in this section. Note that this adaptive-rate system can be designed with any voice coder, FEC design, and at any rate desired.

A. Modulation Format

We have applied this technique using two different adaptive-rate modems, M -ary PSK and M -ary PPM. Results are presented for 16-ary PPM. The use of PPM is motivated by its effectiveness in reducing multipath effects and for spread spectrum applications. For the PPM modem, N data bits are symbol encoded using M distinct symbols, where each symbol is orthogonal to any other. The PPM symbol is determined by a pulse that is delayed

[†] This work was supported by Rome Lab under contract from the U.S. Government, Contract No. F19628-94-C-0069.

Figure 1. Block diagram of adaptive-rate system.

according to the sequence of N data bits. Since each delay corresponds to a distinct sequence of N data bits, the encoding table contains $M = 2^N$ unique delays. [4] provides a description of the M -ary PPM signal set. The probability of a symbol error, P_M , for M -ary PPM (orthogonal signal set) can be found from [5] to be

$$P_M = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left[1 - \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^y e^{-x^2/2} dx \right)^{M-1} \right] \exp \left[-\frac{1}{2} \left(y - \sqrt{\frac{2E_s}{N_0}} \right)^2 \right] dy. \quad (1)$$

Next we provide an analytical expression for the performance of PPM in a Rayleigh fading channel. Since PPM is an orthogonal signaling scheme, we can use the theoretical bit error rate performance for orthogonal coherent FSK as an approximation for PPM. First we assume a “quasi-static” system. The Rayleigh fading performance [6] of orthogonal coherent FSK is

$$P_b \cong 1/(2\Gamma) \quad \{\text{for high } \Gamma (\geq 10 \text{ dB})\}. \quad (2)$$

where Γ is defined as

$$\Gamma = (E_b/N_o) \mathbf{E}[R^2]. \quad (3)$$

$\mathbf{E}[R^2]$ is the average value of R^2 , and R is a Rayleigh-distributed random variable (representing the multiplicative fade, assuming phase is approximately constant).

B. Channel Model

We utilize a fully interleaved Rayleigh fading channel with complex additive white Gaussian noise. Specifically, we assume that the j^{th} received symbol, $y(j)$, is related to the j^{th} transmitted symbol, $x(j)$, according to

$$y(j) = R(j)x(j) + n(j). \quad (4)$$

Here, $x(j)$ is one of 16 possible symbols in the 16-ary PPM signal set and $n(j)$ is a zero-mean complex Gaussian random variable with variance $N_o/2$. $R(j)$ is the multiplicative fading coefficient for the fully interleaved Rayleigh fading channel and is Rayleigh distributed with $\mathbf{E}[R^2] = 1$. To ensure the Rayleigh channel is “fully interleaved,” we assume that $R(j)$ and $R(i)$ are independent for $i \neq j$. The fade occurs such that the Rayleigh coefficient is approximately constant over the symbol duration.

C. Speech Coder

A multimode coder is utilized in this work, primarily because the performance at each bit rate is not compromised when compared to an embedded approach [1]. The vocoder architecture is based on the STC algorithm at 4 different bit rates. There is a more graceful degradation in speech quality using the MMSTC. The four vocoder bit rates used are 9.6 kbps, 4.8 kbps, 2.4 kbps, and 1.2 kbps. The strategy used for switching between each vocoder is critical so that there are no annoying artifacts when a rate switch is requested.

Spectral distortion was chosen for the objective measure of speech quality. Informal listening tests were performed to gauge the subjective speech quality and were found to correlate well with the spectral distortion results. The objective quality was found by measuring the spectral distortion of each strategy versus the average received signal power divided by the noise spectral density, P_r/N_o . Average spectral distortion, denoted here as SD , can be expressed as

$$SD = 10 \log_{10}(e) \sqrt{2} \|a_{cep} - \hat{a}_{cep}\|_2 \quad (5)$$

where a_{cep} is the cepstrum of the linear prediction vectors representing the spectrum. \hat{a}_{cep} is the estimated cepstrum at the receiver due to any remaining bit errors entering the vocoder.

D. Channel Coding and Interleaver

Multiple standard convolutional codes are used in this work as possible candidate codes for the variable channel coding. The channel coding is applied with equal weighting to the information data bits. Rates of 1/2, 1/4, and 1/8, with $K = 7$ as defined in [5], are the multiple channel code rates utilized for the error protection. Viterbi decoding is used with a correlation receiver to decode the received data. Perfect interleaving is assumed, so as to effectively randomize the channel in addition to perfect synchronization at the receiver.

E. Channel Status and System State Estimators

Near-optimal demodulation at the receiver is performed by sampling the output of the system state estimator. The function of the system state estimator is to decode the current vocoding, channel coding, and modulation strategy used by the transmitter. The function of the channel status estimator is to estimate the received signal envelope in (4). It also monitors the bit-error and symbol-error statistics, and the synthesized speech distortion. These statistics are used by the channel status estimator to monitor the long- and short-term conditions of the channel.

An estimate of $R(j)$ can be performed by periodically inserting pilot symbols into the transmitted data stream [7]. Accurate estimates can be made by comparing the received pilot symbols to a known transmitter insertion sequence. Once $R(j)$ is found, the system transmission strategy is determined, the received signal is demodulated, and the bit rate allocation strategy is fed to the Viterbi and speech decoders, and to the de-interleaver, if necessary.

From [1], the fading envelope can be expressed as $R(x)$ and is shown to be the product of the long- and short-term fading coefficients, $R_l(x)$ and $R_s(x)$, respectively. An estimate of the long-

term fading coefficient [8] is shown to be $R_f(x)$ times the average of $R_s(x)$, where x is the distance between the transmitter and receiver. $2W$ is the averaging window size, and if λ is used to denote the RF carrier wavelength, W is commonly in the range of $20\lambda - 40\lambda$. The averaging process assumes that $R_f(x)$ is approximately constant over the width of the window. The channel status estimator uses these fading statistics to determine the new strategy for the next transmission, and outputs this information in the feedback channel to the transmitter.

E. Switching Algorithm

The transmitter utilizes the information accessed from the feedback channel to determine the new vocoding, channel coding/interleaving, and modulation strategy. The receiver confirms the new strategy received from the transmitter. The switching strategy for the modem is straightforward, and is performed by the appropriate changes in sampling frequency and receiver correlation time (bandwidth). The protocol used for switching is also straightforward, and is based on the channel statistics and a perceptually acceptable transition location between the vocoder speech frames.

3. EXPERIMENTAL RESULTS

BER performance analysis of any digital communication system uses the energy per bit divided by the noise spectral density (E_b/N_o) as the standard performance measure. E_b/N_o can be used to compare one digital modulation format to another, but does not reflect the performance increase achieved by an adaptive-symbol-rate modem. In the adaptive-rate system we illustrate the performance increase by using the fact that the received average signal power, P_r , is equal to the energy per bit times the bit rate, R_b , or equivalently

$$P_r = E_b R_b. \quad (6)$$

Dividing both sides by N_o gives

$$P_r/N_o = (E_b/N_o)R_b \quad (7)$$

where P_r/N_o can be used as the adaptive-rate performance measure and shows the BER improvement gained by changing the bit rate R_b (or symbol rate R_s). Note that when channel coding is used

$$P_r/N_o = (E_{cb}/N_o)R_{cb} \quad (8)$$

where E_{cb} is the energy per coded bit and R_{cb} is the coded bit rate. R_{cb} is the sum of the bit rates of the vocoder and channel coder:

$$R_{cb} = R_v + R_c. \quad (9)$$

R_v and R_c are the voice and channel coder bit rates, respectively.

To illustrate the advantages over more conventional approaches (i.e., the fixed-symbol-rate system), comparisons are made between two systems, listed in Tables 1 and 2, for both AWGN and Rayleigh channels. The new approach, reference Table 2, is shown to perform better. It is the most desirable when examining the appropriate trade-offs, such as system complexity and performance.

Table 1. Fixed Aggregate-Rate Channel-Controlled System.

Strategy	R_v (bits/sec)	R_c (bits/sec)	R_{cb} (bits/sec)
1	9600	0	9600
2	4800	4800	9600
3	2400	7200	9600
4	1200	8400	9600

Table 2. Adaptive-rate Speech Communication System.

Strategy	R_v (bits/sec)	R_c (bits/sec)	R_{cb} (bits/sec)
5	9600	0	9600
6	4800	4800	9600
7	2400	2400	4800
8	1200	1200	2400
9	1200	3600	4800

Figures 2, 3, 4, and 5 are plots of the average spectral distortion. For all plots, coherent PPM modulation with exact phase recovery is assumed.

Note that strategies 1 and 5 are equivalent, as are strategies 2 and 6. The results show that the adaptive-rate system (Table 2) requires approximately 3 dB less received power to achieve equivalent spectral distortion performance in the AWGN channel, when compared to the system in Table 1. While the received signal performance is approximately equivalent in the Rayleigh channel, it is important to note that only a single-rate channel code (rate = 1/2) is required for the adaptive-rate system. Note also that strategies 4 and 9 exhibit roughly equivalent performance to strategy 8 in Rayleigh fading, but are inferior in AWGN. Since it is straightforward to change the modem rate in terms of complexity, it may be more desirable to choose an adaptive-symbol rate approach as opposed to the computational complexity involved in channel code rate reductions (rate 1/2 to 1/8) as is done with the fixed aggregate rate system in Table 1.

The switching strategy can be determined by minimizing the system's average spectral distortion at a given received power. This is accomplished by choosing the optimal trade-off between distortion due to channel bit errors with the distortion due to a lower speech encoding rate. Table 3 shows the performance of the adaptive-rate system and the strategies that produce the best overall performance in AWGN and Rayleigh channels, resulting in lower overall system complexity. An improvement of 3 dB is found for AWGN, while similar performance is demonstrated for Rayleigh slow fading. We expect the new adaptive-rate system to also be desirable for Rician fading where a proportionally higher level of the received signal is from the direct line-of-site path.

Table 3. Adaptive-rate system performance versus strategy.

Strategy	P_r/N_o (dB-Hz) [AWGN]		P_r/N_o (dB-Hz) [Rayl.]	
	from	to	from	to
8	$-\infty$	39.1	$-\infty$	50
7	39.1	41.9	50	51.8
6	41.9	45.3	51.8	75.8
5	45.3	$+\infty$	75.8	$+\infty$

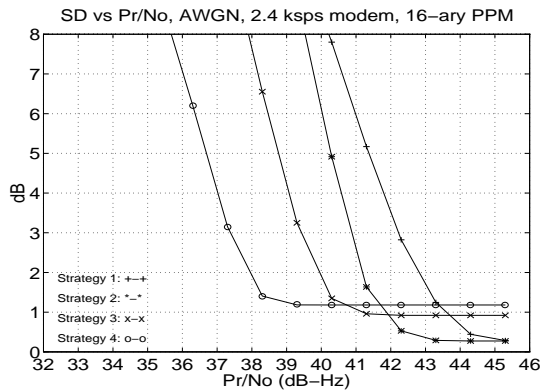


Figure 2. Fixed aggregate-rate system performance in AWGN.

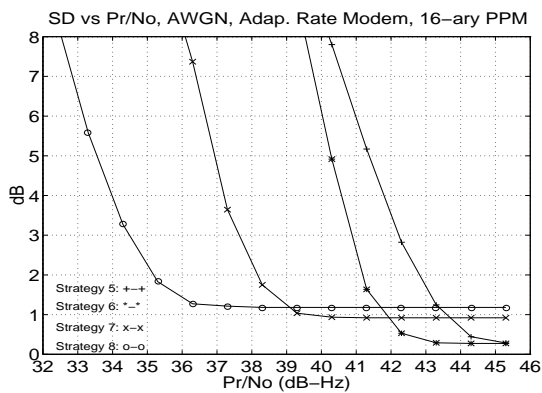


Figure 3. Adaptive-rate system performance in AWGN.

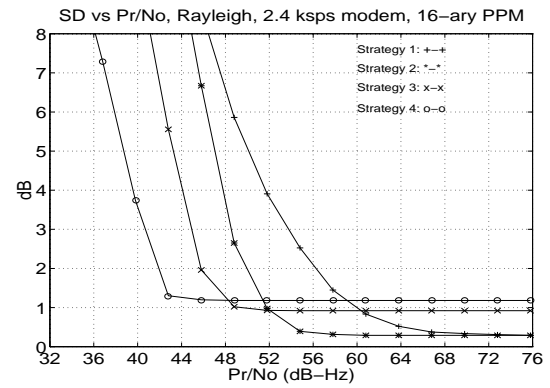


Figure 4. Fixed aggregate-rate performance in Rayleigh fading.

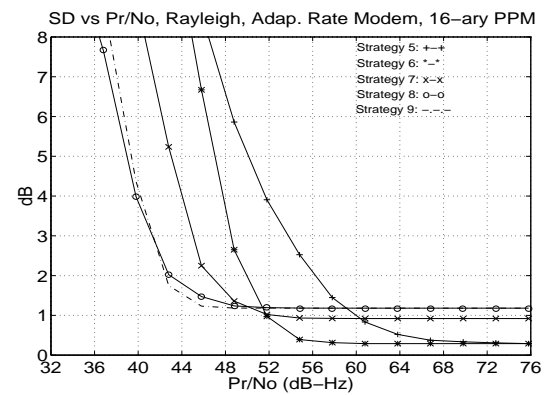


Figure 5. Adaptive-rate system performance in Rayleigh fading.

4. CONCLUSION

An adaptive-rate voice coding method was presented, which chooses the optimal rate and method of vocoding, channel coding, and modulation in response to increased channel bit errors. The proposed system combines the principles of channel capacity and rate-distortion theory to provide the highest speech quality possible, for any given channel condition. A lower rate of source encoding in the vocoder causes an increase in the speech distortion, and is a function of the rate-distortion bound. The speech distortion is minimized by jointly minimizing the distortion in the vocoder due to lower encoding rates and to channel bit errors. Through such a technique a significant gain in speech quality at the receiver can be achieved.

We would like to thank R. McAulay and MIT/Lincoln Labs for use of the STC vocoders and Terry Champion for his support during the project.

5. REFERENCES

[1] E. Yuen, P. Ho, and V. Cuperman, "Variable Rate Speech and Channel Coding for Mobile Communication," Proc. 43rd IEEE/VTS Vehicular Tech. Conf., 1994, pp. 1709 - 1713.

[2] D. J. Goodman and C.-E. Sundberg, "Combined Source and Channel Coding for Variable-Bit-Rate Speech Transmission," Bell Sys. Tech. Journal, Vol. 62, No. 7, Sept. 1983, pp. 2017-2037.

[3] R. J. McAulay and T. F. Quatieri, "Low Rate Speech Coding Based on a Sinusoidal Model," Chapter 1.6, *Advances in Speech Signal Processing*, S. Furui and M. M. Sonhdi (Eds.), Marcel Dekker, N.Y., 1992

[4] J. F. Providakes and Y. Barsoum, "Modelling and Simulation of a Pulse Position Modulation (PPM) Communications System Over Urban Multipath Channels," IEEE MIL-COM91, pp. 10.2.1 - 10.2.7

[5] J. G. Proakis, *DIGITAL COMMUNICATIONS*, 3rd Ed. McGraw-Hill, 1995.

[6] K. Feher, *Wireless Digital Communications Modulation & Spread Spectrum Applications*, Prentice-Hall, N.J., 1995.

[7] J. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels," *IEEE Transactions on Vehicular Technology*, vol. 40, pp. 686-693, Nov. 1991.

[8] W. Lee, "Estimate of local average power of a mobile radio signal," *IEEE Transactions on Vehicular Technology*, vol. 34, pp. 22-27, Feb. 1985.