Capacity Enhancement of Cellular CDMA by Traffic-Based Control of Speech Bit Rate

Tetsuji Kawashima, *Member, IEEE*, Vishal Sharma, *Student Member, IEEE*, and Allen Gersho, *Fellow, IEEE*

***Abstract*—Variable rate speech coding is now recognized as an important system component for high-capacity cellular networks because it exploits speech statistics to reduce the average bit rate, which results in reduced interference and increased capacity. Once a variable rate capability is available, an additional capacity enhancement can be achieved by introducing *network control* of the user bit rate in response to changing traffic levels. In this paper, we introduce the concept of network control of rate and propose a particular network-control method for code-division multiple access (CDMA) systems. Based on an $M/M/\infty//M$ queueing model applied to a cell under heavy traffic conditions and a new performance measure called *averaged speech quality*, we obtain simulation results to demonstrate how network control of rate can achieve improved speech quality or increased capacity for a given quality objective.**

I. Introduction

A VARIABLE rate speech coder generates a transmitted data signal whose rate varies over the duration of a call. Variable rate coders [1] can be divided into two main categories:

1) *source-controlled* variable rate coders: where the coding algorithm responds to the time-varying local character of the speech signal to determine the data rate, with the data rate varying from frame to frame or over a few frames;
2) *network-controlled* variable rate coders: where the coder responds to an external control signal to switch the data rate to one of a predetermined set of possible rates. The external control signal is assumed to be remotely generated, in response to requests for signaling information or, as proposed in this paper, in response to traffic levels in the network.

Source-controlled variable rate coding is already used in actual systems and has been adopted as the TIA industry standard IS-95 for code-division multiple access (CDMA) cellular networks [6]. Network-controlled variable rate coding based on traffic loading appears to be a new concept, however, and will also prove to be useful in making more efficient use of the limited radio spectrum. Although it can be applied to both time-division multiple access (TDMA) and frequency-division multiple access (FDMA) systems, in this paper we propose its application to a CDMA system because CDMA offers a natural and easy way to benefit from variable rate coding in cellular networks. Reducing the coding rate of a user correspondingly reduces interference to other users. Furthermore, each user transmits a wideband signal that covers the entire spectral band, which is shared by a large number of users. Thus, there is no family of frequency channels (as in FDMA systems) and no assignment of time slots to different channels (as in TDMA systems), and the rather complex overhead required for network control in other systems is eliminated. In the IS-95 standard, each frame may have one of four rates and the receiver automatically identifies the rate without requiring side information.

In the past decade, major advances have been made to low bit-rate speech coding for telecommunications [4], but only recently has attention been addressed specifically to variable rate speech coding for CDMA (see, for example, [1]–[3]). Most algorithms of current interest are based on the family of techniques known as *code-excited linear prediction* (CELP). In particular, the QCELP variable rate algorithm [2] was adopted by the TIA as a service option for the IS-95 standard. For a tutorial on CELP and other speech coding algorithms, see [5].

While many methods of network-controlled variable rate speech coding are possible, perhaps the simplest method is one in which the network directs all transmitters to use a lower rate during periods of heavy traffic. Under normal or light traffic conditions, a higher rate coder performs better (in terms of the speech quality experienced by the users) than a lower rate coder because the performance in this case is limited not by the interference but by the coding rate. On the other hand, under heavy traffic conditions, coding at a higher rate can lead to more severely degraded quality due to channel bit errors, because the performance in this case is limited not by the coding rate but by the total interference in the network. Therefore, as illustrated in Fig. 1 for the case of two alternative bit rates, a favorable trade-off between the coding rate and the perceived speech quality may result. The same principle can be extended to network control of variable rate coders with more than two bit-rates by incrementally reducing the rate at suitably chosen successively increasing levels of traffic.

In this paper we demonstrate that this hypothesis is indeed valid based on an idealized but reasonably realistic model of the problem. We utilize an objective measure of perceived speech quality to quantitatively derive a relation between speech quality and the offered traffic load. This calculation is done is two steps. The first step involves calculating the

Manuscript received January 30, 1995; revised August 22, 1995 and September 22, 1995. This work was supported in part by Sony Corp. and Motorola, Inc.
T. Kawashima is with Sony Corporation 4-16-1, Okata, Atsugi-shi, 243 Japan.
V. Sharma and A. Gersho are with the Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106-9560 USA.
Publisher Item Identifier S 0018-9545(96)02899-X.

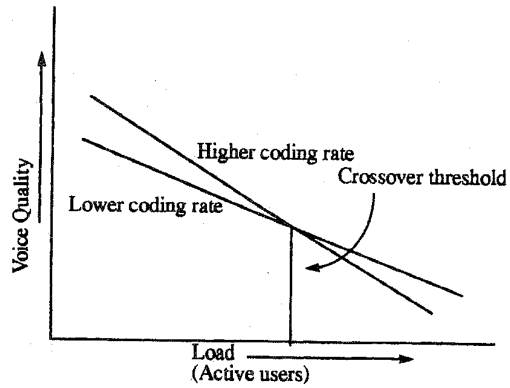


Fig. 1. Representative curves of voice quality versus number of active users for the case of two alternative bit rates.

traffic load (or the number of users) as a function of the error rate at the receiver. The second step involves calculating the error rate at the receiver as a function of the speech quality. These results are combined to obtain a relationship between the speech quality and the number of users in the network.

II. Relation Between the Traffic Load and the Error Rate

Most research that has been done to obtain a relation between the traffic load and the error rate has until now concentrated on calculating the bit error rate (BER) at the receiver versus the number of users per cell. This is because these studies concentrated chiefly on the communications rather than the source coding aspects of such systems (see, for example, [7] and [8]). In actual systems, however, speech quality usually depends on the frame erasure rate (FER) rather than the BER. Bit errors are correctable by employing forward error correction (FEC) codes, so a frame of speech is in error only when the error correcting capability of the FEC code is exceeded. Thus, at the output of the decoder it is the FER and not the BER that is the parameter of interest. The CDMA standard of the TIA, for instance, employs convolutional codes (different codes for the uplink and the downlink) for forward error correction. If the number of errors in a frame exceeds the error correcting capability of the code, the frame is treated as an erasure. Although in an actual system, the receiver uses a block code (with a very weak error correcting capability of correcting one error) to correct such extra errors (see Fig. 2), we will ignore its effect for the purposes of our analysis.

In this paper, we use an additive white Gaussian noise (AWGN) approximation model of the cellular channel for the downlink, that is, the connection from the base-station to the mobile (see for example, [7] and [9]). As in the paper by Milstein *et al.* [7], we assume that the mobile is situated at the intersection of four square-shaped cells, a possible worst case scenario (see Fig. 3). Although the uplink is more severe than the downlink and should also be considered, in this paper we will only use a simple model of the downlink to confirm the effectiveness of the network-controlled variable rate coding scheme. The model is based on a frequency-duplex system, where every mobile in the system experiences interference from all base stations, but experiences no interference from the transmitters of other mobiles. The signal from each base station is composed of K direct-sequence waveforms (asynchronous with one another), where K is the total number of users in a cell. As in [7], we also assume that the composite signal from each base station independently undergoes flat fading with a Rayleigh distribution, that is, all signals that arrive at a mobile from a given base station propagate over the same path, and so are assumed to fade in unison.

As pointed out previously, Milstein *et al.* [7] have calculated the BER as a function of the number of users per cell. Pursley and Taipale [9], however, have calculated an upper bound on packet error rate (PER), which is the quantity of interest for us, in terms of the first error-event probability of the convolutional code used. Therefore, we modified the analysis in [7] to obtain a relation between the FER and the number of users per cell. A packet error is defined as the generation of errors in a sequence of data bits (i.e., the packet) that are beyond the correcting capability of the FEC code. Since the system initializes the convolutional decoder at the beginning of every frame, the PER and FER are the same in our case, provided that we ignore the effect of the additional block code. Since this additional code can correct at most one error, we will ignore its effect in our analysis.

The upper bound on the FER is calculated based on the first error-event probability of the convolutional code used, which is itself calculated only as an upper bound [11]. The fact that these probabilities are upper bounds makes our results pessimistic. Since we compare a system with network-controlled variable rate coding to one without it, under the same pessimistic conditions, however, the qualitative nature of our results still remains valid.

The upper bound on the first-event error probability P_u of an AWGN channel is given by

$$P_u \leq T(D)|_{D=\exp(-E_b R_c/N_0)} \tag{1}$$

where $T(D)$ is the transfer (or generating) function of the convolutional code used [11], E_b is the transmitted energy per bit, N_0 is the one-sided power spectral density of the noise, and R_c is the code rate of the convolutional code used. The BER of a cellular channel approximated as an AWGN channel has been calculated in [7] to be

$$\begin{aligned} &P(e|\alpha_1, \alpha_2, \alpha_3, \alpha_4) \\ &\quad = \phi\left(-\left[\frac{K}{3G}\left(1 + \frac{\alpha_1^2 + \alpha_2^2 + \alpha_3^2}{\alpha_1^2}\right)\right]^{-1/2}\right) \end{aligned} \tag{2}$$

where $\phi(x)$ is given by

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2}\, dy \tag{3}$$

and where K is the number of users in the cell, G is the processing gain of the CDMA system, and $\alpha_1, \alpha_2, \alpha_3$, and α_4 are the "fading" variables, which are independent and identically distributed Rayleigh random variables, representing the fading of the amplitudes of the signals received from the four base stations (see Fig. 3).

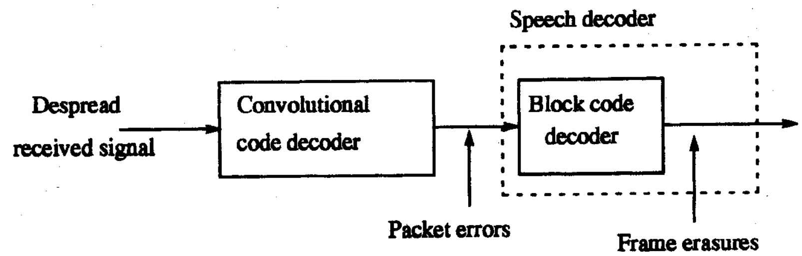


Fig. 2. Receiver block diagram. The errors that are not corrected by the convolutional decoder give rise to packet errors. If the packet errors cannot be corrected by the block code decoder, a frame erasure results. For our analysis, we ignore the effect of the single-error correcting block-code decoder, so packet errors and frame erasures are synonymous for our purposes.

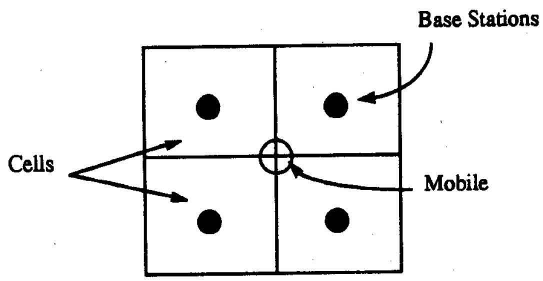


Fig. 3. A mobile at the intersection of four cells—a worst case placement.

When BPSK signaling is used, the BER for an AWGN channel is also given by

$$P_e = \phi(-\sqrt{2E_b/N_0}) \tag{4}$$

where ϕ is as defined in (3).

Both (2) and (4) represent the BER in an AWGN channel using BPSK signaling, and they should therefore be the same. Thus, equating the arguments of the ϕ function in the two equations, we obtain

$$E_b/N_0 = \frac{1}{2}\left[\frac{K}{3G}\left(1 + \frac{\alpha_1^2+\alpha_2^2+\alpha_3^2}{\alpha_1^2}\right)\right]^{-1} \tag{5}$$

where E_b/N_0 is the effective bit energy to interference density ratio, which can be used in (1) to obtain a bound on the first error event probability P_u, conditioned on the random variables $\alpha_1, \alpha_2, \alpha_3$, and α_4. The unconditional first error-event probability can then be calculated by averaging over the distribution of ratio of the "fading" variables as

$$P_u \le \int_o^\infty T(D)|_{D=\exp(-E_b R_c/N_0)} f_Z(z)\, dz \tag{6}$$

where $z = \alpha_1^2 + \alpha_2^2 + \alpha_3^2/\alpha_1^2$ and $f_Z(z) = 3z^2/(1+z)^4$ as shown in [7].

To obtain meaningful bounds on the first error-event probability, it was necessary to truncate the probability distribution of the random variable z at a value that would yield values of $T(D)$ less than unity. Therefore, we used a truncated and renormalized distribution function in (6). For this purpose, we numerically solved the equation $T(D) = 1$ to obtain an upper threshold value, say D_o, for D. Since $D = \exp(-E_b R_c/N_0)$ we obtain

$$\left(\frac{E_b}{N_0}\right)_{\max} R_c = \ln(1/D_0). \tag{7}$$

Combining (7) and (5) and remembering that the ratio of the amplitudes (or "fading" variables) is given by the random variable z, we obtain the following bound $z_{\max}$ on the value of the random variable z

$$z_{\max} = \frac{1.5GR_c}{K_{\max}\log(1/D_0)} - 1 \tag{8}$$

where $K_{\max}$ is the maximum number of users per cell and the remaining parameters are as explained earlier. Thus, the modified probability density function used in (6) is given by

$$\hat{f}_Z(z) = \frac{f_Z(z)}{F(z_{\max})} \tag{9}$$

where $F(z_{\max}) = \int_0^{z_{\max}} f_Z(z)\, dz$. Finally, the packet error probability P_{packet} can be upper bounded in terms of P_u (see [10]) as

$$P_{\text{packet}} = 1 - (1 - P_u)^L \tag{10}$$

where L is the packet length, that is, the number of sequential data bits in a packet.

III. Relation Between the Speech Quality and the Error Rate

A block diagram of the simulation scheme used to obtain the relationship between speech quality and the error rate is illustrated in Fig. 4. Since the FER derived in the previous section includes the effect of channel coding, our simulation does not employ a channel coder. For inserting errors into the coded speech, we used the error insertion device (EID), a standard tool provided by the CCITT [8]. The EID can insert bit errors or frame erasures into a speech file, with the bursty (correlation) factor of the errors or the erasures controlled according to the Gilbert–Elliott channel model. We note that the FER calculated in Section II includes the effect of error correction due to convolutional decoding. Therefore, in the simulation set up shown in Fig. 4 we only considered the FER at the input of the speech decoder. Since this setup does not include the effect of channel coding, it may be argued that the distribution of the FER's in the two cases might be different, and this may affect the final result. However, as pointed out later in Section V, it turns out that the final results are practically independent of the correlation between the FER's in the simulation scheme, so this difference will not cause any significant change in the final results.

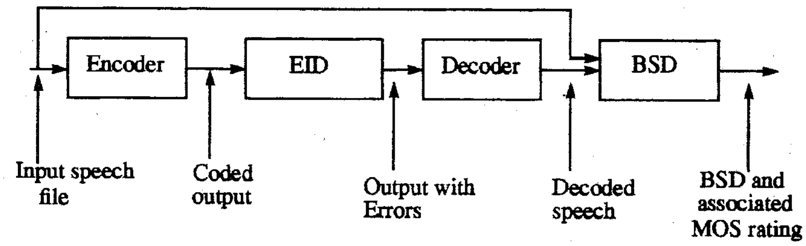


Fig. 4. A block diagram of the simulation scheme is illustrated. The decoded speech signal at the output of the speech decoder is compared with the original speech signal to calculate the bark spectral distortion.

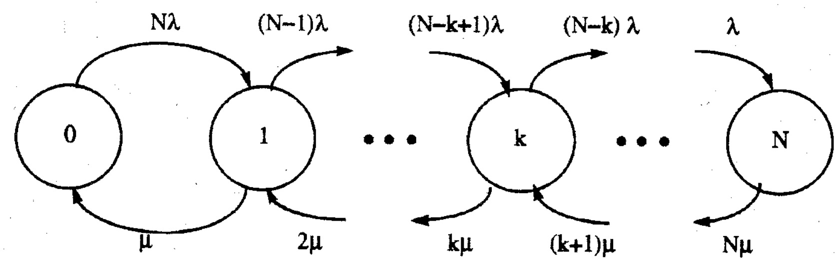


Fig. 5. The state-transition-rate diagram for the $M/M/\infty//M$ queueing model, which is used to evaluate the performance of the system.

As a speech quality measure, we used the bark spectral distortion (BSD) and the predicted mean opinion scores (MOS) obtained from the BSD values [9]. The BSD is the average squared Euclidean distance between spectral vectors of the original and reconstructed utterances (see [13]). The BSD takes into account auditory frequency warping, critical band integration, amplitude sensitivity variations with frequency, and subjective loudness. The standard error in estimating MOS scores with the measure was 0.2–0.3, with the higher accuracy for low rate coders in the range of 2.4–8 Kbps. The measure offers a more consistent assessment of the effect of incremental changes in the parameter of a speech coder than is usually obtained by the designer who relies on his or her own informal listening.

IV. System Evaluation

To evaluate the performance of a CDMA system that employs network-controlled variable rate coding we propose a new performance measure, which we call the *averaged speech quality.* In a system that employs network controlled variable rate coding, the source coder must respond to the instructions from the network controller, which in turn depend on the time-varying traffic levels in the network. Consequently, any performance measure to evaluate the system performance should include the effect of this statistical variation in traffic character. In our case, we account for this variation by using an appropriate queuing model. We analyze a heavy traffic situation and use the $M/M/\infty//M$ queueing model [14], with the assumption that the total number of users in conversation in a cell is fixed at N, while the number of active users k, that is, the users involved in a talk-spurt, has a probabilistic distribution governed by the above model (see Fig. 5). The state variable is the number of active users in the system. We assume that for each user who is not active the talk-spurts arrive according to a Poisson process with rate λ, independently of other users in the system, and that the talk-spurt durations are exponentially distributed with mean μ. In any state k, only the $(N-k)$ users who are currently silent can initiate new talk spurts. The arrival rate λ of talk-spurts from a silent user is the reciprocal of the average duration S of the silence interval. Thus, the arrival rate into the system in state k is given by

$$\lambda_k = \begin{cases} (N-k)\lambda, & \text{for } 0 \le k \le N; \\ 0, & \text{for } k > N \end{cases} \tag{11}$$

where $\lambda = 1/S$.

Similarly, the service rate in state k is conditioned on the fact that k users are involved in a talk-spurt. The service rate $1/\mu$ of a talking user is the reciprocal of the average talk-spurt duration T. Thus, the service rate of the system in state k is given by

$$\mu_k = \begin{cases} k\mu, & \text{for } 0 \le k \le N; \\ 0, & \text{for } k > N \end{cases} \tag{12}$$

where $1/\mu = T$.

Fig. 6 shows the relation between a talk-spurt and a silence interval. The ratio λ/μ can also be expressed in terms of the voice activity factor ν, as

$$\frac{\lambda}{\mu} = \frac{\nu}{1-\nu} \tag{13}$$

where

$$\nu = \frac{T}{S+T}. \tag{14}$$

In accordance with the $M/M/\infty//M$ queueing model [14], the state probability P_k that there are k active users in the

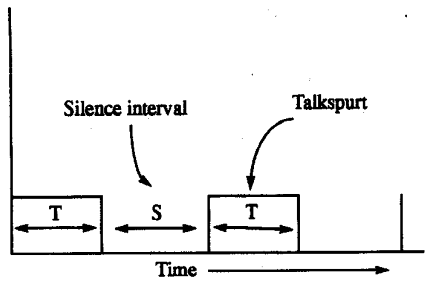


Fig. 6. Relation between a talk-spurt duration and a silence interval. Each talk-spurt is followed by a silence interval. In our model, a talk-spurt duration is exponentially distributed with mean $\mu = 1/T$, and each silence interval is exponentially distributed with mean $1/\lambda = S$.

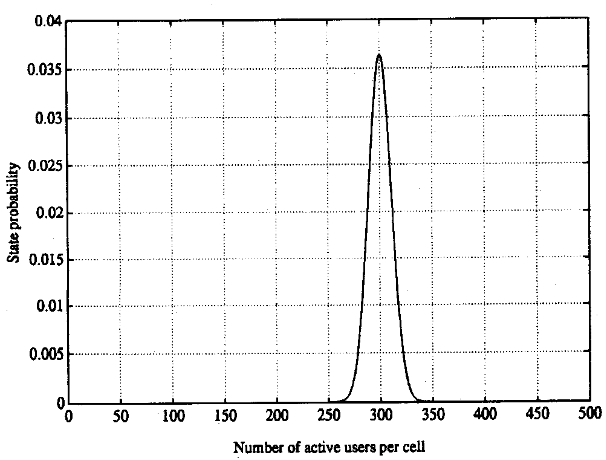


Fig. 7. State probability distribution of the number of active users per cell, derived as per (15).

system (which in our case is a cell) is given by

$$P_k = \begin{cases} {}^N C_k \left(\frac{\lambda}{\mu}\right)^k \left(1 + \frac{\lambda}{\mu}\right)^{-N} & \text{for } 0 \le k \le N; \\ 0, & \text{for } k > N. \end{cases} \quad (15)$$

From (13), this probability is a function only of the voice activity factor ν and of the total number of users N in conversation. Fig. 7 shows an example of the state probability distribution, when there are $N = 500$ users in the cell and the voice activity factor is $\nu = 0.6$.

Although we used Fig. 1 to illustrate the concept of network controlled variable rate coding, we point out that it depicts only a *static* situation, because it shows how the speech quality may vary with the number of active users in the system. In reality, the nature of a cell's loading is dynamic, in the sense that the number of active users is constantly changing, even if the total number of users N in the cell remains roughly constant (in heavy traffic conditions). Using our model, the number of active users k has a probabilistic distribution given by (15). Since the speech quality experienced by a user now varies constantly, we define a new performance measure, called the *averaged speech quality*, which we will use to evaluate the system performance. This measure is obtained by weighting the speech quality $Q(k)$, experienced by the users when there are k active users in the cell, by the probability P_k that k users are active. Therefore, the averaged speech quality experienced by a user is given by

$$\overline{Q} = \sum_{k=1}^{N} Q(k) P_k. \quad (16)$$

We note that $Q(k)$ can in fact be any speech quality measure such as segmental SNR, BSD, or MOS. As explained earlier in Section II, to characterize $Q(k)$ we have used the BSD measure and the predicted MOS values. It should be noted that the performance measure $\overline{Q}$ contains the effect both of voice activity detection and of network-controlled variable rate coding.

V. Results

The experimental results presented here are derived using the system parameters of the CDMA standard of the TIA. Fig. 8 shows an example of the relation between FER (upper bound) and the number of active users in a system for the two cases that we considered, namely, when all users employ a high rate coder and when all users employ a low rate coder. Although the QCELP coder was intended as a (source-controlled) variable rate coder, for our experiment we modified it to work at the two fixed rates of 9.6 Kbps and 4.8 Kbps, which correspond, respectively, to our high rate coder and to our low rate coder. Thus, the processing gains for Fig. 8 are 12.5 MHz/9.6 Kbps = 1302.08 for the higher rate and 12.5 MHz/4.8 Kbps = 2604.16 for the lower rate. As per the CDMA standard of the TIA, the downlink employs the best convolutional code of rate half (i.e., $R_c = 1/2$) and constraint length nine, whose transfer function [15] is given by

$$T(D) = 11D^{12} + 50D^{14} + 286D^{16} + 1630D^{18} + \cdots. \quad (17)$$

We have used only the first eight terms (four out of eight of whose coefficients are zero), because for reasonable values of the effective bit energy to noise density ratio, the higher order terms do not contribute significantly to the sum. The plots in Fig. 8 were obtained by the method outlined in Section II. We first calculated the upper threshold value D_o, by solving the equation $T(D) = 1$ for the transfer function of (17). Next for each user population $K_{\max}$, we used (8) to evaluate $z_{\max}$ and renormalized the probability density function $f_Z(z)$ as per (9). The bound on first error-event probability P_u was calculated using (6). Finally, the FER, P_{packet} (recall that modulo our assumption, of ignoring the effect of the single error correcting block decoder in the receiver, the packet error probability and the FER are identical in our case), for each user population was obtained from (10).

Figs. 9 and 10 show an example of the relation between speech quality and frame erasure rate. The plots in these two figures were obtained as explained in Section III (see Fig. 4). The EID module, which inserts erasures into frames of the coded speech, allows the user to program into it the required FER. We therefore inserted frame erasures into the coded speech at several different FER's and for each FER obtained the quality of the decoded speech by using the BSD module. (The BSD value for each FER was obtained

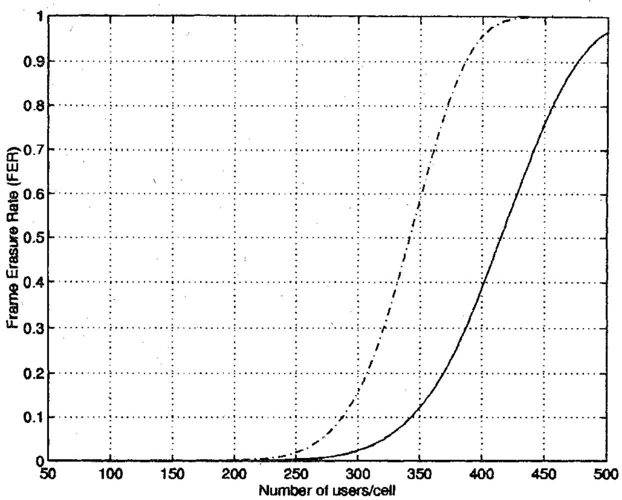


Fig. 8. FER versus number of active users per cell for two alternative bit rates, 9.6 and 4.8 Kbps, respectively. Legend: FER (full rate) = dash-dot curve and FER (half rate) = solid curve.

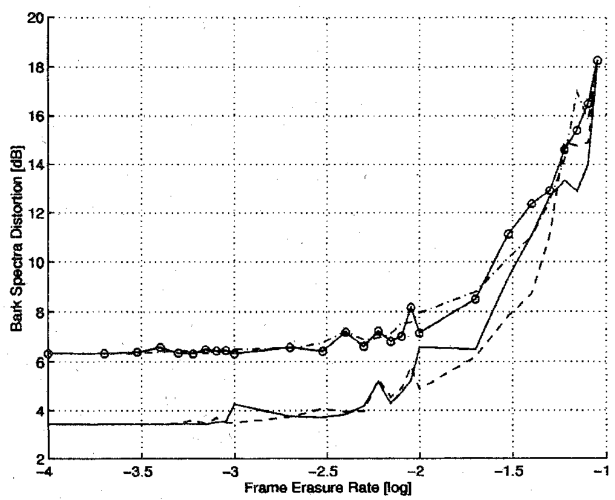


Fig. 9. BSD versus frame erasure rate for the case of two alternative bit rates, 9.6 and 4.8 Kbps, respectively. For each rate, two curves are shown, one with random frame erasures (abbreviated "random" in the legend) and one with correlated frame erasures (abbreviated "correlated" in the legend). Legend: full rate (random) = solid curve, full rate (correlated) = dashed curve, half rate (random) = solid curve with circles, and half rate (correlated) = dash-dot curve. It can also be seen that highly correlated frame erasures (a bursty factor = 0.8) give practically the same distortion as uncorrelated frame erasures (a bursty factor = 0.0), thus justifying our choice of working only with uncorrelated frame erasures.

by averaging over several speech files.) Fig. 9 shows BSD as a function of the FER in a cell, and Fig. 10 shows the predicted MOS scores, obtained from the corresponding BSD values, as a function of the FER in a cell. We obtained these graphs for bursty factors of 0.0 and 0.8, representing completely random frame erasures and highly correlated frame erasures, respectively. Even though it is expected that in an actual cellular environment the frame erasures at a receiver output would be correlated, these graphs indicate that the slight difference in performance between the two bursty factors is not significant. Considering the correlation of erasures

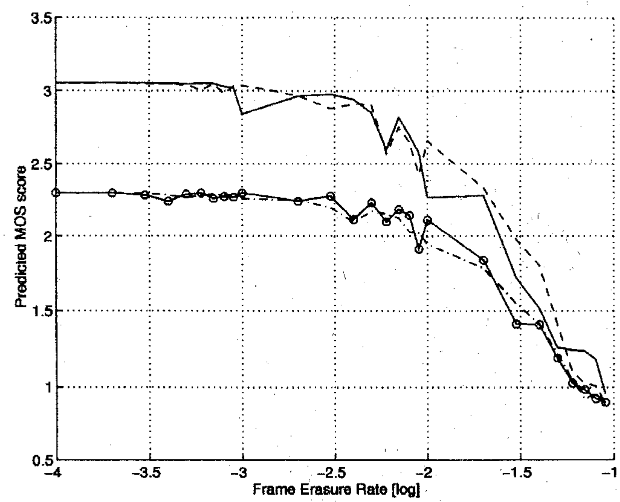


Fig. 10. MOS versus frame erasure rate for the case of two alternative bit rates, 9.6 and 4.8 Kbps, respectively. For each rate, two curves are shown, one with random frame erasures and one with correlated frame erasures. The legend and associated explanation for the curves is the same as that in Fig. 9. As in Fig. 9, we see that correlated frame erasures give the same results as uncorrelated frame erasures.

and its effect on speech quality is also an interesting issue. However, it is not our main purpose here. For our subsequent calculations we have used a correlation factor of 0.0, because the correlation of frame erasures does not affect the validity of our results, since using a bursty factor of 0.8 also gives similar final results.

In conjunction with Fig. 8, Figs. 9 and 10 give Figs. 11 and 12, respectively, which show the relation between speech quality and the number of active users in the cell. In both graphs, we have the expected crossover threshold (255 in this case). Thus, the simple network control can be implemented by switching all users to the coding rate of 4.8 Kbps beyond this threshold. Figs. 13 and 14 show the averaged MOS score, with and without the simple network control, for voice activity factors of 0.4 and 0.6, respectively. To derive the graphs shown in Figs. 13 and 14, we set the MOS score to zero when the total number of users exceeded the maximum for which MOS scores were available from Fig. 12. This is justified on account of the observation that the MOS score has already dropped very close to zero when the number of users per cell reaches the limit shown in Fig. 12. We note that actually calculating these values and using them for deriving the latter figures would only improve the averaged voice quality and therefore improve the results further.

From Figs. 13 and 14 we see that there is an improvement in speech quality for a fixed user population. Alternatively, if an acceptable threshold of the averaged MOS value was set at a nominal value of 2.0 we see that there is an increase in network capacity. Though a threshold of 2.0 might appear low at first sight, we consider it a reasonable threshold here because it is known that the MOS values predicted from the corresponding BSD values are lower than the actual MOS scores by a bias value of about 0.3 (for QCELP) (see [13]). At the crossover threshold, a system with the network-controlled variable rate speech coding method gives about a 7% increase in capacity

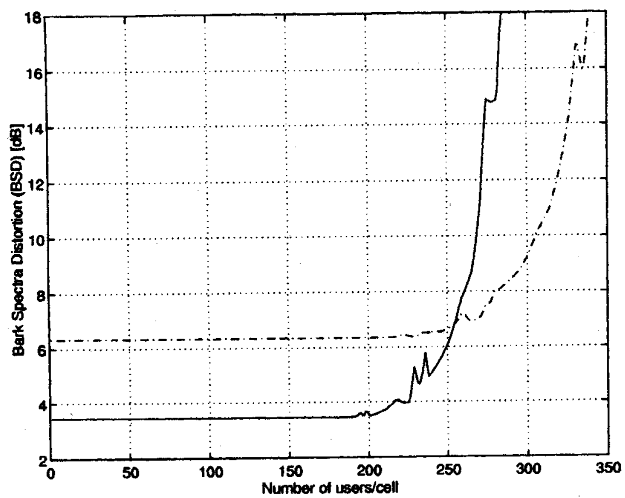


Fig. 11. Speech quality (BSD) versus the number of active users per cell for the case of two alternative bit rates, 9.6 Kbps (or "full-rate") and 4.8 Kbps (or "half-rate"), respectively. Legend: BSD (full rate) = dash-dot curve and BSD (half rate) = solid curve.

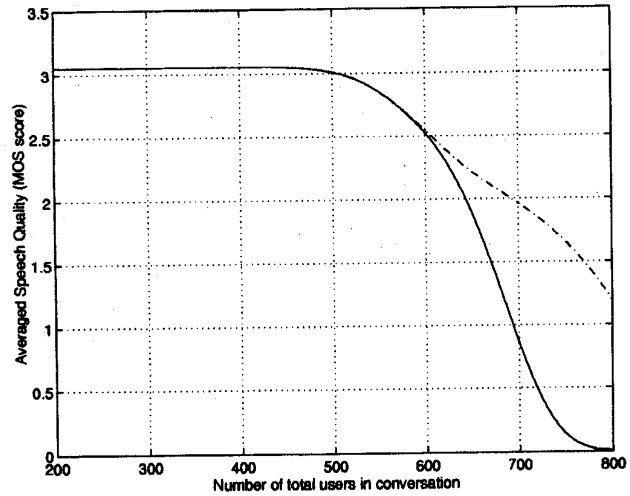


Fig. 13. Averaged speech quality $\overline{Q}$ versus total number of users in conversation N, for a voice activity factor $\nu = 0.4$ and two alternative bit rates. Legend: $\overline{Q}$ (full rate) = dash-dot curve and $\overline{Q}$ (half rate) = solid curve.

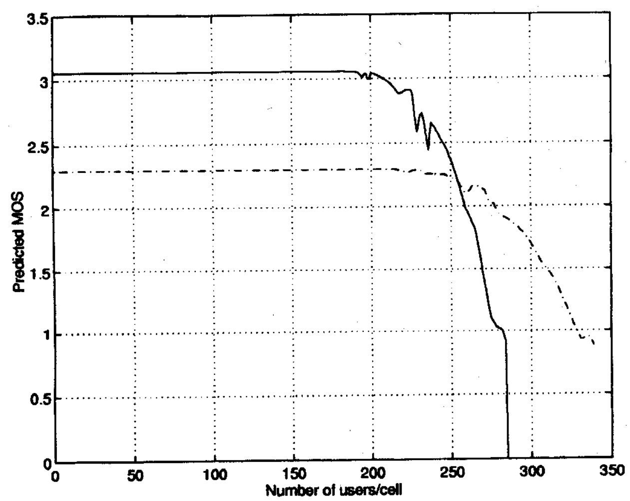


Fig. 12. Predicted MOS versus number of active users per cell for the case of two alternative bit rates, 9.6 and 4.8 Kbps, respectively. Legend: MOS (full rate) = dash-dot curve and MOS (half rate) = solid curve.

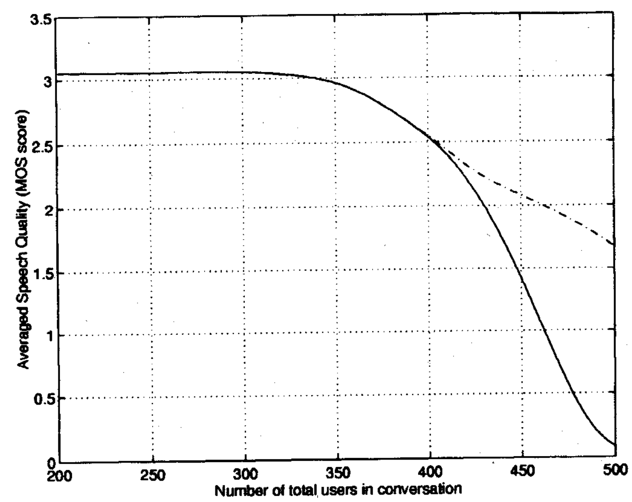


Fig. 14. Averaged speech quality $\overline{Q}$ versus total number of users in conversation N, for a voice activity factor $\nu = 0.6$ and two alternative bit rates. The legend is the same as that in Fig. 13.

over a fixed rate system for voice activity factors of 0.4 and 0.6, respectively.

We also see from Fig. 12 that under high traffic conditions employment of a higher performance low-rate coder will provide better system operation by providing a higher speech quality for all users. This is simply because the predicted MOS curve for a better low rate coder would be above the curve of the prototype low rate coder used by us, and such a coder will deliver a higher speech quality beyond the crossover threshold at which the network switches to this coder.

VI. Discussion

We proposed the use of network control of speech bit rate and a simple strategy for implementing such a scheme in a CDMA environment. The averaged speech quality performance measure was introduced to evaluate the network-controlled system. The effectiveness of the method was demonstrated with this quality measure. The degree of improvement by the network-controlled method needs further study, since the speech quality measure, the predicted MOS based on the BSD, is not perfect, and our calculations still include some approximations. For instance, frame erasure rates are not actual values but rather only upper bounds. Also the effect of error distribution has not been considered in depth. We believe, however, that the qualitative nature of our results lends strong support to the effectiveness of network-controlled variable rate coding and highlights it as an interesting area for further research.

In this paper, we confined our focus to the downlink (from base to mobile), where the base station can easily coordinate the coding rate of all users without any significant delay. In the uplink, on the other hand, we expect some delay in the switching of the coding rate at mobiles, due to the propagation delay on the control channel on which the base station's instructions would travel. The effect of this delay is of interest for further study.

As pointed out before, there is a trade-off between the speech quality resulting from the coding rate itself and the speech quality resulting from the interference in the system. In the network control scheme proposed in this paper, the network directs *all transmitters* to use a lower rate during periods of heavy traffic. Due to the trade-off, however, it is also possible that only a fraction of users can be assigned a lower rate under a certain heavy traffic condition, which should be achievable by a sophisticated network controller. Such a scheme would be capable of giving a further improvement in speech quality (or capacity, for a fixed quality) by use of network-control.

ACKNOWLEDGMENT

The authors would like to thank Prof. R. Iltis for his valuable discussions on CDMA error probabilities and for suggesting the renormalization approach that was used in Section II. They would also like to thank R. Walton of GTE Laboratories for information on GTE's implementation of QCELP, K. Srinivasan for modifying the QCELP coder for their needs, and E. Paksoy for help with the BSD/MOS program.

REFERENCES

[1] A. Gersho and E. Paksoy, "Variable rate speech coding for cellular networks," in *Speech and Audio Coding for Wireless and Network Applications,* B.S. Atal, V. Cuperman, and A. Gersho, Eds. Norwell, MA: Kluwer, 1993, pp. 77–84.

[2] A. DeJaco, W. Gardner, P. Jacobs, and C. Lee, "QCELP: The North American CDMA digital cellular variable rate speech coding standard," in *Proc. IEEE Workshop Speech Coding for Telecom.,* Sainte Adele, Quebec, Oct. 1993, pp. 5–6.

[3] E. Paksoy, K. Srinivasan, and A. Gersho, "Variable bit rate CEIP coding of speech with phonetic classification," *European Trans. Telecommun.,* vol. 5, no. 5, pp 591–602, Sept.–Oct. 1994.

[4] A. Gersho, "Advances in speech and audio compression," *Proc. IEEE,* vol. 82, no. 6, pp. 900–918, June 1994.

[5] A. S. Spanias, "Speech coding: A tutorial review," *Proc. IEEE,* vol. 82, no. 10, Oct. 1994, pp. 1541–1582.

[6] Qualcomm, Inc., "Digital cellular system cdma-analog dual mode mobile station-base station compatibility standard," Mar. 5, 1992.

[7] L. B. Milstein, T. S. Rappaport, and R. Barghouti, "Performance evaluation for cellular CDMA," *IEEE J. Select. Areas Commun.,* vol. 10, no. 4, pp. 680–689, May 1992.

[8] G. L. Stuber *et al.,* "Capacity of direct-sequence CDMA for cellular radio," in *Proc. Int. Conf. Selected Topics in Wireless Commun.,* Vancouver, Canada, June 25–26, 1992, pp. 199–202.

[9] M. B. Pursley and D. J. Taipale, "Error probabilities for spread-spectrum packet radio with convolutional codes and Viterbi decoding," *IEEE Trans. Commun.,* vol. 35, no. 1, pp. 1–12, Jan. 1987.

[10] M. B. Pursley, "Performance evaluation for phase-coded spread-spectrum multiple-access communication—Part 1, system analysis," *IEEE Trans. Commun.,* vol. 26, no. 4, pp. 328–337, Mar. 1978.

[11] A. J. Viterbi, "Convolutional codes and their performance in communication systems," *IEEE Trans. Commun.,* vol. 19, no. 5, pp. 751–772, Oct. 1971.

[12] *CCITT Software Tool Library Manual,* Report of Study Group XV, COM-XV-R 87-E, Geneva, May 1992.

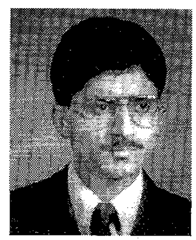
[13] S. Wang, A. Sekey, and A. Gersho, "An objective measure for predicting subjective quality of speech coders," *IEEE J. Select. Areas Commun.,* vol. 10, no. 5, pp. 819–829, June 1992.

[14] L. Kleinrock, *Queueing Theory,* vol. I. New York: Wiley, 1974.

[15] J. P. Oldenwalder, "Optimal decoding of convolutional codes," Ph.D. dissertation, Dept. Elec. Eng., Univ. of California, Los Angeles, 1970.

Tetsuji Kawashima (M'92) received the B.S. degree in communication engineering from Osaka University, Japan, in 1987.

Since then, he has been with Sony Corp., Japan, and had been involved in the development of high performance magnetooptical disk drives and is now developing a high-density CD-ROM drive. He was a Visiting Researcher at University of California at Santa Barbara from 1992 to 1993, where he studied wireless networks.



Vishal Sharma (S'94) received the B.Tech. degree in electrical engineering from the Indian Institute of Technology, Kanpur, India, in 1991, and the M.S. (signals and systems) and M.S. (computer engineering) degrees, in electrical and computer engineering, in 1993 from the University of California, Santa Barbara, where he is currently pursuing the Ph.D. degree.

During the summer of 1992 he worked at the Digital Technology Research Laboratory in Motorola's Corporate Research and Development Center, Schaumburg, IL, where he investigated algorithms for real-time resizing of decompressed video sequences. His research interests include communication networks, and parallel and distributed computing, and more specifically protocol design for high-speed networks and the analysis of switching schemes for multiprocessor systems.

Mr. Sharma is a member of ACM SIGCOMM and ACM SIGDOC.

Allen Gersho (S'58–M'64–SM'78–F'82) received the B.S. degree from Massachusetts Institute of Technology, Cambridge, in 1960 and the Ph.D. degree from Cornell University, Ithaca, NY, in 1963.

From 1963 to 1980, he was with Bell Laboratories. He is currently Professor of Electrical and Computer Engineering at the University of California, Santa Barbara (UCSB). His current research interests include signal compression methodologies and algorithm development for speech, audio, image, and video coding. He holds patents on speech coding, quantization, adaptive equalization, digital filtering, and modulation and coding for voiceband data modems. He is co-author, with R.M. Gray, of the book *Vector Quantization and Signal Compression* (Norwell, MA: Kluwer, 1992) and co-editor of two books on speech coding.

Dr. Gersho served as a member of the Board of Governors of the IEEE Communications Society from 1982–85 and is a member of various IEEE technical, award, and conference management committees. He has served as Editor of the IEEE COMMUNICATIONS MAGAZINE and Associate Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS. He received NASA "Tech Brief" awards for technical innovation in 1987, 1988, and 1992. In 1980, he was co-recipient of the Guillemin–Cauer Prize Paper Award from the Circuits and Systems Society. He received the Donald McClennan Meritorious Service Award from the IEEE Communications Society in 1983, and in 1984 he was awarded an IEEE Centennial Medal. In 1992, he was co-recipient of the 1992 Video Technology Transactions Best Paper Award from the IEEE Circuits and Systems Society.