An Error-Protected Speech Recognition System for Wireless Communications

Vijitha Weerackody, Wolfgang Reichl, Alexandros Potamianos
Bell Labs, Lucent Technologies, Murray Hill, NJ 07974, USA

Abstract

uture wireless multimedia terminals will have a variety of applications that require speech recognition capabilities. In this paper, we consider a robust distributed speech recognition system where representative parameters of the speech signal are extracted at the wireless terminal and transmitted to a centralized automatic speech recognition (ASR) server. We propose several unequal error protection schemes for the ASR bit stream and demonstrate the satisfactory performance of these schemes for typical wireless cellular channels. In addition, a "soft-feature" error concealment strategy is introduced at the ASR server that uses "soft-outputs" from the channel decoder to compute the marginal distribution of only the reliable features during likelihood computation at the speech recognizer. This soft-feature error concealment techniques reduces the ASR error rate by up to four times for certain channels. Also considered is a channel decoding technique with source information that improves ASR performance. uture wireless multimedia terminals will have a variety of applications that require speech recognition capabilities. In this paper, we consider a robust distributed speech recognition system where representative parameters of the speech signal are extracted at the wireless terminal and transmitted to a centralized automatic speech recognition (ASR) server. We propose several unequal error protection schemes for the ASR bit stream and demonstrate the satisfactory performance of these schemes for typical wireless cellular channels. In addition, a "soft-feature" error concealment strategy is introduced at the ASR server that uses "soft-outputs" from the channel decoder to compute the marginal distribution of only the reliable features during likelihood computation at the speech recognizer. This soft-feature error concealment techniques reduces the ASR error rate by up to four times for certain channels. Also considered is a channel decoding technique with source information that improves ASR performance. F

I. Introduction

Automatic speech recognition (ASR) over wireless networks is important for next generation wireless multimedia systems [11]. A variety of spoken dialogue system exist today that utilize ASR technology, e.g., personal assistants, speech portals, travel reservation, stock quotes. The number of application being written specifically for the car (hands-free ASR) and for wireless devices is also increasing. Introducing a robust spoken dialogue interface to wireless terminals will enhance existing applications and help create new ones. High speech recognition accuracy for a variety of channel and noise conditions is essential for the success of ASR applications and services. Our goal in this paper, is to investigate the degradation in speech recognition performance under typical wireless channel conditions and propose error protection and concealment strategies that improve performance.

For most automatic services, access to databases and transactions which are executed on a networked server are required in addition to speech recognition. It is therefore advantageous to have a distributed speech recognition approach in such applications, rather than using an automatic speech recognition (ASR) unit at the mobile terminal. In a distributed speech recognition system, a small client program running in the device extracts and transmits

representative parameters of the speech signal from the mobile terminal over the wireless network to a multiuser speech recognition server. The alternative approach of performing speech recognition locally on the device significantly increase computations, power and memory requirements for the device, and limits portability across languages and application domains. With todays technology only speech recognition systems with very limited vocabulary, e.g., speaker-trained name dialing, can reside on the handset, while a great majority of applications resides on the network. In this paper, we adopt and investigate a distributed approach to ASR.

The speech parameters obtained using a regular speech coding algorithm are not necessarily the best parameters for speech recognition purposes. In addition, speech coders usually spend a significant amount of bits for the transmission of the excitation or LPC-residual signal, while this information is not useful for speech recognition. In this paper, we extract and transmit speech parameters that are specifically optimized for speech recognition.

The effects of various speech coding algorithms on automatic speech recognition (ASR) performance has been studied by several authors [7], [6], [16], [19]. In [16], [19], [10], [15], [9], severe ASR performance degradation was observed for a distributed wireless speech recognition system, especially in the case of transmission errors that occur in bursts. Because of rapid fluctuations of received signal strength the mobile radio environment can be a very difficult channel for data transmission. Therefore, for the transmission of ASR parameters, a specialized channel error protection scheme is necessary to improve bandwidth and power efficiency. The channel error protected speech parameters form a speech recognition codec located at the wireless terminal and the basestation. Our work is geared towards building an efficient speech recognition codec for a wide range of different channel conditions. In addition, we want to avoid retransmission of speech parameters in case of transmission errors which introduces additional delay in the system response and reduces the spectral efficiency.

From the speech signal we extracted representative parameters appropriate for speech recognition purposes and quantized these parameters to give a source bit rate of 6 kb/s. It was determined that the bit stream obtained from the speech parameters have different sensitivity levels for transmission errors. In this paper, we propose several error protection schemes that give unequal levels of error protection to different segments of the bit stream. The overall bit rate of the coded bit stream is 9.6 kb/s. We have conducted experiments to examine this codec over a wide variety of wireless channels, such as, Gaussian and various correlated Rayleigh channels, and demonstrated the satisfactory performance of the system for a typical speech recognition task, even in the case of adverse channel conditions.

In this paper "soft-outputs" from the channel decoder are used to improve the performance of the speech recognition system. Specifically, the confidence level for each decoded bit is obtained and this is used to estimate the confidence in ASR features and weight the importance of each feature in the speech recognition algorithms. This novel "soft-feature" decision is shown to produce dramatic improvements in ASR performance. Also, we have observed a residual correlation in some of the bits in the quantized source bit stream. This correlation is utilized in the channel decoder to improve the performance of the ASR system.

The organization of this paper is as follows. In Section II, the quantization scheme used for the speech recognition features is presented. In Section III, several error protection schemes are proposed that give unequal levels of error protection for different segments of the bit stream. Soft-feature error concealment for distributed speech recognition is introduced in Section IV. In Section V, we evaluate the performance of the speech recognition codec for a wide variety of channels: Gaussian and Rayleigh fading with different mobile speeds. Finally, channel decoding with source information is presented in Section VI.

II. Speech Parameters and Quantization

Many available speech recognition systems use cepstral features for signal parameterization. It is a compact and robust speech representation, well suited for distance based classifiers and may be calculated from a mel-filterbank

Feature Component	$e, c_1, c_2, c_3, c_4, c_5$	$c_6, c_7, c_8, c_9, c_{10}, c_{11}$	c_{12}
Bits	6	4	0

TABLE I

BITS ALLOCATION FOR DIFFERENT FEATURE COMPONENTS

analysis or the linear prediction approach (LPC) [21]. The acoustic features for speech recognition used in this study are the 12 cepstral coefficients, $c_1, c_2, ..., c_{12}$, calculated every 10 ms based on a LPC analysis of order 10 and the signal energy, e. The signal sampling rate is 8000 Hz and a Hamming window with 240 samples is used. These features form a 13-dimensional vector every 10 ms, which is the acoustic input to the automatic speech recognition system.

For data transmission purposes all 13 features are scalar-quantized. A simple non-uniform quantizer is used to determine the quantization cells. The quantizer uses the empirical distribution function as the companding function, so that samples are uniformly distributed in the quantization cells. The algorithm is a simple non-iterative approximation to Lloyd's algorithm [18], which does not necessarily minimize quantization noise. A similar quantization scheme for distributed speech recognition can be found in [5]. Better performance may be achieved using a k-means type of algorithm applied to the entire feature vector (vector quantization) [23]. Note that the error protection and concealment algorithms proposed in the next sections are valid for different quantization schemes.

An empirical analysis of the effects of different bit allocation schemes on speech recognition performance can be found in Section V-A. The bit allocation scheme used in all experiments in this paper is shown in Table I. Six bits were allocated for each of the energy e and the most significant cepstrum $c_1 ... c_5$ features, while four bits were assigned to each of $c_6 ... c_{11}$. Empirical tests showed no significant performance degradation for the evaluated task by replacing the last (12th) cepstral coefficient, c_{12}, with its fixed precalculated mean. This means that there is not much information relevant to the speech recognition process in c_{12} and thus no bits where allocated to c_{12}. At the receiver, c_{12} is simply restored to its fixed precalculated average value, and the standard 13-dimensional feature vector is used during recognition. The total number of bits for this bit allocation scheme is 60 bits per 10 ms frame. This requires an uncoded data rate of 6 kb/s to be transmitted over the wireless channel which will be the data rate used throughout this paper.

III. TRANSMISSION SYSTEM

The 60 bits in a 10 ms speech frame require different levels of error protection. Unequal error protection (UEP) schemes for speech coding applications have been extensively examined in the literature as well as in the standards [14], [8], [25]. In this work we examined several UEP schemes and three schemes that gave better performance gains will be presented. The performance of the UEP schemes were based on the experimental work given in Section V.

As shown in the previous section the data rate for the quantized speech parameters is 6 kb/s. In this work we consider a 9.6 kb/s data rate for the coded signal with binary differential phase shift keying (DPSK) modulation format. This is one of the data rates used in the North American cellular standard IS-95 [26]. The channel overhead introduced at 9.6 kb/s data rate is reasonable and if lower coded bit rates are required trellis coded modulation schemes with higher order modulations may be considered. Note that we are using a differential modulation technique to simplify the demodulation process.

In slow fading channels it is useful to have a large interleaver to improve the system performance. However, large interleavers introduce delays and this may not be desirable in some realtime applications. In this work we have

UEP Level	Speech Bits	Error Protection
L1_1	$e^0(n), e^1(n), c_1^0(n), c_2^0(n), c_3^0(n), c_4^0(n), c_5^0(n)$	(12,7) cyclic code rate 1/2 conv. code
L1_2	$e^2(n), c_1^1(n), c_2^1(n), c_3^1(n), c_4^1(n), c_5^1(n)$	rate 1/2 conv. code
L2	$e^3(n), e^4(n), c_1^2(n), c_1^3(n), c_2^2(n), c_2^3(n), ...$ $c_6^0(n), c_6^1(n), c_7^0(n), c_7^1(n), ..., c_{11}^0(n), c_{11}^1(n)$	rate 2/3 conv. code
L3	$e^5(n), c_1^4(n), c_1^5(n), ..., c_5^4(n), c_5^5(n),$ $c_6^2(n), c_6^3(n), c_7^2(n), c_7^3(n), ..., c_{11}^2(n), c_{11}^3(n)$	no code

TABLE II

SPEECH BIT ASSIGNMENT FOR DIFFERENT UEP LEVELS IN UEP1

chosen an 80 ms frame, or 8 speech frames, for interleaving and channel coding purposes. The total interleaving and deinterleaving delay is 160 ms and this is can be tolerated in wireless speech recognition applications. The 12 parameters that have to be protected in a 10 ms speech frame are the energy parameter $e(n)$ and the 11 cepstral coefficients $c_1(n), c_2(n), ..., c_{11}(n)$, where n denotes the speech frame index. Obviously, the more significant bits of the above parameters should have better channel error protection. In addition, as discussed in Sections V-A and V-B, it was determined experimentally that the energy parameter $e(n)$ is the most sensitive to quantization noise as well as random transmission errors followed by $c_1(n), ..., c_5(n)$ and then $c_6(n), ..., c_{11}(n)$. The channel coded bit rate is 9.6 kb/s, therefore, the total coded bits in a 80 ms channel encoded frame is 768.

A. Unequal Error Protection Scheme 1 – UEP1

In this case we consider three levels of channel error protection denoted by L1, L2 and L3. Furthermore, to give a higher level of error protection to the most significant bits of L1, L1 is separated to two levels: L1_1 and L1_2. The assignment of the bits for different unequal error protection (UEP) levels is shown in Table II. In this notation $e^0(n), e^1(n), ...$ denote the bits of $e(n)$ in decreasing order of significance. As seen from the table the number of bits per speech frame in L1, L2 and L3 are 13, 24 and 23, respectively. In this case L1_1 contains the bits that are determined to be the most important 7 bits and these are protected using an outer (12,7) cyclic code [17] in addition to the inner convolution code. In this application the (12,7) cyclic code is used only to detect errors which is useful in error concealment at the receiver; however, with additional receiver complexity it is possible to use this code for error correction as well. L1_2 contains the next 6 important bits. We employ a rate 1/2, memory 8 code on L1 level bits and, thus, the total number of coded bits for the 8 speech frames for L1 level is 288.

The Level L2 contains the next 24 important bits and we use a rate 2/3 rate compatible punctured convolutional (RCPC) code [12]. The total number of coded L2 level bits for the 8 speech frames including the 8-bit tail is 300. The least important bits are in L3 and these are transmitted without any channel coding. In order to maintain the total bits after coding in 8 speech frames to 768, 4 bits are further punctured from the coded L2 level bits. Channel coding is done so that L1_1 level bits are followed by L1_2 and then L2. Note that because of the RCPC code the rate 1/2 code is not terminated, therefore, those bits of L1_2 that are separated from L2 level by less than a decoding depth of the channel code will not be subjected to the usual rate 1/2 mother code. At the channel encoder input the L1_2 level bits for the 8 speech frames, $n, (n+1), ...(n+7)$, are arranged in the following manner: $e^2(n), e^2(n+1), ..., e^2(n+7); c_1^1(n), c_1^1(n+1), ..., c_1^1(n+7); .....; c_5^1(n), c_5^1(n+1), ..., c_5^1(n+7)$. As stated previously, we have determined that the coefficients $c_1(n)$ are more significant than $c_5(n)$ and, therefore, this

UEP Level	Speech Bits	Error Protection
L1_1	$e^0(n), e^1(n), c_1^0(n), c_2^0(n), c_3^0(n), c_4^0(n), c_5^0(n)$	(12,7) cyclic code rate 1/2 conv. code
L1_2	$c_6^0(n), c_7^0(n), c_8^0(n), c_9^0(n), c_{10}^0(n), c_{11}^0(n)$	rate 1/2 conv. code
L2	$e^2(n), e^3(n), e^4(n), c_1^1(n), c_1^2(n), c_1^3(n), ..., c_5^1(n), c_5^2(n), c_5^3(n)$ $c_6^1(n), c_7^1(n), c_8^1(n), c_9^1(n), c_{10}^1(n), c_{11}^1(n)$	rate 2/3 conv. code
L3	$e^5(n), c_1^4(n), c_1^5(n), ..., c_5^4(n), c_5^5(n),$ $c_6^2(n), c_6^3(n), c_7^2(n), c_7^3(n), ..., c_{11}^2(n), c_{11}^3(n)$	no code

TABLE III

SPEECH BIT ASSIGNMENT FOR DIFFERENT UEP LEVELS IN UEP3

bit arrangement will assign bits of lower significance toward the end of the L1_2 frame which will be subjected to a less powerful code than the usual rate 1/2 mother code.

B. Unequal Error Protection Scheme 2 – UEP2

In this subsection we remove the outer code from the L1_1 level bits in the UEP scheme discussed earlier. This allows us to apply a stronger code on L2 level bits. However, the L1_1 level bits will not enjoy the same strong error protection because of the absence of error detection and correction capability of the outer code. The bit assignment for L1, L2 and L3, except for the error protection, is the same as in Table II. With a rate 1/2 convolutional code the total coded bits in 8 speech frames from L1 level bits is 208. For the 200 L2 bits (including the 8-bit tail) we use a rate 1/2 code with 24 bits punctured to give 376 coded bits. Then, with the 184 L3 uncoded bits the total coded bits in 8 speech frames is 768. The L1_2 bits are arranged as in Section III-A with L1_1 bits preceding L1_2 bits.

C. Unequal Error Protection Scheme 3 – UEP3

Since it was determined that the feature components, $e(n), c_1(n), c_2(n), c_3(n), c_4(n), c_5(n)$, are the most important, in the previous error protection schemes we used 2 MSBs of each one of these components in L1. However, the MSBs of $c_6(n), c_7(n), c_8(n), c_9(n), c_{10}(n), c_{11}(n)$ are important parameters as well. In this subsection we rearrange the bits so that the MSBs of all the feature components are now grouped in L1. The bit arrangement is shown in Table III. As seen from this table L1_1 bits are the same as in UEP1 and they are protected by a (12,7) outer code and a rate 1/2, memory 8, inner code similar to UEP1. As in UEP1, a rate 2/3 code is applied on L2 bits and L3 bits are not coded.

D. Transmission System Model

Denote by $a(n)$ the speech bits at the input to the channel encoder and $b(n)$ the channel encoder output. $b(n)$ is interleaved over 768 symbols which occurs in 80 ms and then differentially encoded to give $u(n) = d(n)d(n-1)$, where $d(n)$ is the interleaver output. The baseband equivalent received signal can be written as

$$y(n) \quad = \quad A\beta(n)u(n) + \nu(n) \tag{1}$$

where A is the transmit amplitude, $\beta(n)$ is the complex channel gain and $\nu(n)$ is the additive white Gaussian noise (AWGN) component. For a Rayleigh fading channel $\beta(n)$ is a correlated complex Gaussian variable with $E\{\beta(n)\beta^*(n+k)\} = J_0(2\pi\frac{v}{\lambda}kT)$, where v, λ and T are the mobile speed, wavelength of the RF carrier wave and

the symbol interval duration, respectively. At the receiver, $y(n)$ is first differentially decoded, deinterleaved, and then Viterbi decoded. The output of the Viterbi decoder, $\hat{a}(n)$, is then sent to the speech recognition unit.

IV. Soft-Feature Error Concealment

To overcome the detrimental effects of transmission errors common error concealment strategies include the repetition of previously received frames or parameter interpolation. These techniques may help to repair random bit errors but may fail for errors occurring in bursts, which are very likely in fading channels. In this section we consider a novel error concealment technique which is based on "soft-outputs" from the channel decoder to the ASR unit. In this case we use an algorithm that maximizes the *a posteriori* probability (MAP) [3] which gives the *a posteriori* probability of each decoded bit. The ASR unit utilizes this information to give improved performance gains.

For each of the 12 decoded speech feature components, the receiver generates an additional value giving the confidence of correctly decoding that component. In our case, we generate two confidence bits for each of the 12 features; the first and second bit corresponding to the first and second MSB of each feature. Specifically, suppose $\hat{a}(n)$ is the relevant MSB bits at the channel decoder output. The MAP decoder gives the probability $p_i(n) = Prob\{\hat{a}(n) = i\}, i = 0, 1$, where $p_0(n) + p_1(n) = 1$. Let us denote a threshold, $T(> 0.5)$, then the confidence level $\Lambda_i(n) = 1$ if $p_i(n) > T$; $\Lambda_i(n) = 0$ otherwise. With this assignment when the confidence value is close to 1 the corresponding bit is correct with a very high probability and when the confidence value is close to 0 the transmitted bit is represented by an erasure. These 1-bit quantized confidence values, $\Lambda_i(n)$, for each of the two MSBs of the 12 feature components, are sent to the ASR unit together with the channel decoded bit stream.

In the results presented in Section V, we do not use the MAP algorithm given in [3] for channel decoding. Instead the correct value of the confidence value of the channel decoder output is assumed to be available at the receiver. In order to generate $\Lambda_i(n)$, the channel decoder output $\hat{a}(n) = i$ is examined: if this is correct set $\Lambda_i(n) = 1$, $\Lambda_i(n) = 0$ otherwise. Note that this approach may give rise to better results than what could be obtained from a practical MAP decoding algorithm. However, the MAP algorithm together with the appropriate threshold, T, will give reasonably accurate estimates of the confidence values. More work is needed to study the effects on speech recognition performance when using approximate rather than exact confidence values.

The proposed error concealment strategy discards the transmitted features which are probably erroneous and uses only the reliable ones for likelihood computations at the speech recognizer. A reduced feature vector is used based only on the components that have a high confidence level. In an hidden Markov model (HMM) based speech recognition system, the observed feature vectors are modeled by state-specific probability distributions $p(x|s)$, where x is the feature vector and s is the state of the model. Usually a mixture of Gaussian densities is used for each state of the phoneme (or triphone) specific HMM [24]. In this case, the reduced distribution for the reliable part of the feature vector is the marginal determined by integrating over all unreliable components:

$$p(x_{rel}|s) = \int p(x|s) dx_{unrel}. \tag{2}$$

where x_{rel}, x_{unrel} is the reliable and unreliable components of the feature vector. Using the marginal distribution of the reliable components for HMM likelihood computation is one of the techniques for improving robustness of speech recognizers in noisy conditions, often labeled as the "missing feature theory" [4]. For speech recognition in noise, labeling unreliable spectral features can be a challenging task, while in our application the reliability of each feature is provided by the channel decoder. With diagonal covariance Gaussian mixture modeling, the reduced likelihood function can be easily calculated by dropping unreliable components from the full likelihood computation [4]. This approach requires little modification in existing speech recognition systems. In addition, feature components in the likelihood computation can also be weighted by their confidence values. In this case, continuous confidence

values between 0 and 1 would be used and the contribution of each feature to the likelihood computation would be scaled by its confidence. In applying this error concealment approach, the ASR features are used in a "soft" way, each component is weighted by its confidence.

The soft-feature strategy used in this paper is as follows: (i) for energy and cepstrum features, if the first or the second bit was received with confidence value equal to zero do not use it in the likelihood computation (marginalize according to Eq. (2)), (ii) for 'delta' and 'delta-delta' features (smooth first and second derivatives of the energy and cepstrum features), if the first or the second bit of any of the features in the window used for the delta computation has been received with confidence value zero, then, do not use the delta feature in the likelihood computation. Five and seven frame windows are used for the delta and delta-delta computation, respectively. For details of how the deltas are computed from the original feature set see [22]. For more details on the soft feature decoding see [20].

V. Experimental Results

The performance of the ASR system for various transmission channel conditions and error protection schemes was evaluated on an isolated word speech recognition task, where people were asked to spontaneously answer questions about their mother language, country of birth etc. The database was collected over the public telephone network. A total of 4387 utterances were used for the system evaluation. The vocabulary size was 23 different words. This test set consists of speakers from all over the United States with large dialect diversity and a significant amount of non-native speakers.

The 12 LPC-derived cepstral coefficients, the signal energy, and the first- and second- order time derivatives of these components were used as acoustic features for speech recognition. The cepstral mean for each utterance was calculated and removed before recognition. The cepstral coefficients and the signal energy are calculated at the mobile terminal and transmitted to the basestation. They are reconstructed at the receiver, augmented with the confidence values for soft-feature error concealment, and sent to the network based speech recognition server where the first- and second-order time derivatives are generated.

The acoustic models for speech recognition were trained on a collection of English speech databases collected over the public telephone network. The speech recognizer is based on continuous density HMMs and the Bell Labs recognition engine [27]. The acoustic units are state-clustered triphone models, having three emitting states and a left-to-right topology [24].

A. Quantization

The baseline word-error-rate (WER) for this task (without quantization and transmission errors) was 6.8%. The relatively high error rate is due to the noisy conditions, the usage of speaker-phones, and hesitations and filled pauses in the data (spontaneous speech).

Using the proposed quantization scheme, which gives 60 speech bits per 10 ms speech frame (as shown in Table I), the WER increases to 7.1%. This small performance degradation is due to the relatively simple scalar quantization of the feature vector components. Note that there was no performance degradation when the number of quantization bits for $c_6 \ldots c_{11}$ was reduced from 6 to 4 bits. Performance loss could be reduced or eliminated by increasing the bit rate or by applying more sophisticated vector quantization schemes commonly used for wireless transmission of speech parameters. For example, in [23], a loss-less vector quantization scheme is proposed that operates at 4 kb/s. Scalar quantization schemes similar to the one proposed here were investigated in [5]. In this work, we concentrate on the effects of transmission errors and concealment strategies on ASR performance and the issue of optimal quantization is not investigated further. The conclusions about error protection and concealment obtained from this study are mostly valid for more complex quantization strategies.

B. Speech Recognition with Transmission Errors

In order to investigate the sensitivity of different feature components to transmission errors, additive white Gaussian noise was introduced to the bit stream at different signal-to-noise (SNR) levels. These experiments were done on a subset of the final ASR test data and showed that the signal energy bits are very sensitive to bit errors. At 5 dB SNR level for the energy (without any noise in the other components), the ASR WER increased from 6.5 to 6.8% and for 3 dB SNR level the WER increased sharply to 12.2%. Note that a 3 dB SNR level translates to approximately a 2% error in the energy bits. This shows the high sensitivity of the speech recognizer to bit errors in the energy component. A 5 dB SNR level in one of the cepstral coefficient did not affect the error rate; however, a 3 dB SNR level for one of the lower order cepstral coefficients, $c_1(n), ..., c_5(n)$, resulted in a moderate performance degradation. A 3 dB SNR level in one of the higher order coefficients, $c_6(n), ..., c_{11}(n)$, showed no appreciable increase in error rate and we conclude that these feature vectors are not very sensitive to random bit errors. From these experiments and from the speech recognition literature [22], it was clean that the relative significance of the the speech feature set for the speech recognition task is: energy $e(n)$ followed by $c_1(n)$, $c_2(n)$ and the rest of the cepstrum coefficients. The relative significance was taken into account when designing the error protection schemes in Section III.

C. Speech Recognition with Error Protection Schemes

1) Gaussian Channels: In the first set of experiments, the performance of the recognition system is evaluated for a Gaussian channel using the different error protection schemes proposed in Section III. The transmission system operates at 9.6 kb/s rate for all error protection schemes and a 900 MHz carrier frequency was used for the simulations. The WER for the ASR and the bit-error-rates (BER) for the Gaussian channel are given in Tables IV and V, respectively. The results show the performance of the UEP scheme, the UEP scheme in the presence of error concealment ("UEP+conc."), and the UEP scheme with soft-features ("UEP+soft.") as discussed in Section IV. The WER without any transmission errors is at 7.1%; error rates over 20% are considered very high for this application.

Several simple concealment strategies were investigated, where an erroneous sub-frame is replaced by its previously transmitted error-free sub-frame. For the results given in Table IV the seven L1_1 bits listed in the first row of Table II were used as the sub-frame for error concealment. Errors in the current sub-frame are detected using the (12,7) outer code. In the case of an error, the current sub-frame is replaced by the previously transmitted sub-frame provided that the previous sub-frame is error-free. Since the UEP2 error protection scheme does not employ an outer code no error concealment technique was investigated for UEP2. It can be seen from Table IV that this simple error concealment technique has a negligible effect on the WERs. It was observed that for UEP1 at 2 dB SNR level, only about 0.001% speech frames were concealed using this technique and at higher SNRs this number is extremely small. So the effect of concealment is negligible at SNR levels higher than 2 dB. At 1 dB SNR, concealment was used on 0.012% of the speech frames; however, as seen from Table V, the BERs for L1_2, L2 and L3 are very high and cause recognition errors, thus, rendering the concealment technique ineffective for speech recognition purposes. Note that for the Gaussian channel, if the error bursts introduced by the channel decoder are ignored, the bit errors tend to be random. This makes the errors in the current sub-frame approximately independent of the errors in the previous sub-frame which enables the error concealment technique to work effectively. For a Rayleigh fading channel, however, where the bit errors tend to be in bursts, this error concealment technique is even less desirable than for a Gaussian channel.

As shown in Table IV, the UEP2 scheme gives the best performance gains. At 2 dB SNR, the WER for UEP2 is 10.3% whereas the WERs for UEP1 and UEP3 are 32.3% and 48.7%, respectively. For the L1_1 bits all three error

Error Protection	SNR			
Scheme	4 dB	3 dB	2 dB	1 dB
UEP1	7.4	8.7	32.3	81.9
UEP1 + conc.	7.4	8.7	32.3	81.6
UEP1 + soft.	7.5	8.9	27.2	57.8
UEP2	7.4	7.6	10.3	49.6
UEP2 + soft.	7.4	7.6	8.8	21.6
UEP3	7.4	10.4	48.7	89.5
UEP3 + conc.	7.4	10.4	48.6	89.2
UEP3 + soft.	7.4	8.1	11.2	37.1

TABLE IV

WERS (%) FOR DIFFERENT SNRS FOR A GAUSSIAN CHANNEL. "UEP+CONC." DENOTES UNEQUAL ERROR PROTECTION WITH CONCEALMENT, WHILE "UEP+SOFT." DENOTES UNEQUAL ERROR PROTECTION WITH SOFT-FEATURES.

protection schemes give similar performances as shown in Table V, However, for L2 level bits, UEP2 outperforms other schemes by more than 1 dB. L1_2 bits also give better BERs for UEP2. Note that L3 level bits are transmitted without any channel coding. From the WER and BER results in Tables IV and V, respectively, one may deduce that for satisfactory performance of the ASR (WER less than 20%) the necessary BERs are: $10^{-2} - 10^{-3}$ for L1_1 and L1_2; about 10^{-2} for L2; and about 10^{-1} for L3.

ASR performance results with soft-feature concealment ("soft.") are also listed in Table IV. For UEP2 at 1dB SNR, the WER reduces significantly from 49.6% to 21.6% with soft-features. In general, this technique improves the recognition rate dramatically for low SNR levels (up to a factor of four). As stated previously, the correct one-bit confidence level, Λ_i=0 or 1, $i = 0, 1$ were assumed to be known in these experiments. The confidence levels, are generated at the receiver by examining the decoder output and setting Λ_i=1 if the output is correct and Λ_i=0 otherwise. At very low SNRs it may be difficult to obtain the correct values of $\Lambda_i(n)$ automatically. Therefore, the WER reduction shown in Table IV are too optimistic for low SNRs under realistic conditions.

Finally, note that the WER reduction when using soft-feature concealment are better for UEP3 than UEP1, although, while, UEP2 is the best error protection scheme in terms of absolute WER (either with or without soft-features). This can be explained as follows. As seen from Tables II and III the MSBs, $(c_6^0 - c_{11}^0)$, for UEP1 and UEP3 are in L2 and L1_2, respectively. The L1_2 level bits will have a lower BER. Therefore, the confidence level, $\Lambda_i(n)$ will be higher for L1_2 than L2. Since the soft-feature error concealment depends on the confidence values, it gives the best performance gains for UEP3.

2) Rayleigh Fading Channels: Next we consider a Rayleigh fading channel at mobile speeds of 10 km/h, 50 km/h and 100 km/h. The ASR WERs and the BERs for this case are shown in Tables VI and VII, respectively. It can be seen that the best error protection is given by UEP2 followed by UEP1 and UEP3. At 10 dB SNR, the UEP2 scheme gives satisfactory ASR performance (7.3-10.2% WER) even at slow speeds. From comparing the results for the Gaussian and Rayleigh fading channels in Tables IV-VII, it can be seen that comparable speech recognition performance corresponds to BERs that are higher for the Rayleigh fading case. This is because errors in fading channels occur in bursts and features that correspond to large segments of speech are corrupted. This results in lower WERs for the Rayleigh fading channel at the same BER or, equivalently, higher BER at the same WER.

It can be seen from Table VII that for UEP2 at 10 km/h the BERs for L1_1, L1_2 and L2 are very similar. This is again due to the bursty nature of errors in the Rayleigh fading channel. At high speeds, for example at 100 km/h

Error Protection Scheme	SNR 4 dB	3 dB	2 dB	1 dB
UEP1, UEP3 – L1_1	0.00000	0.00015	0.00680	0.08614
UEP1, UEP3 – L1_2	0.00001	0.00082	0.02021	0.14799
UEP1, UEP3 – L2	0.00105	0.02637	0.18020	0.38158
UEP1, UEP3 – L3	0.03905	0.06537	0.09846	0.13627
UEP2 – L1_1	0.00000	0.00013	0.00553	0.07137
UEP2 – L1_2	0.00000	0.00019	0.00823	0.10646
UEP2 – L2	0.00001	0.00051	0.01582	0.14167
UEP2 – L3	0.03901	0.06537	0.09844	0.13625

TABLE V

BERS FOR DIFFERENT SNRS FOR A GAUSSIAN CHANNEL.

and 5 dB SNR, this effect is less pronounced and the BERs are quite different for L1_1, L1_2 and L2. As stated earlier, the simple frame-repetition error concealment technique is not effective in this case; however, the concealment technique using soft-features gives significant performance gains. With soft-feature concealment, as in the Gaussian channel case, 10-20% absolute gain in WER or 1-2 dB gain in SNR can be achieved. The WER gains are greater for low SNR values (WER reduction up to a factor of three).

VI. CHANNEL DECODING WITH SOURCE INFORMATION

A key objective of source encoding is to minimize correlation between output symbols. In many practical source coders, however, there is some residual correlation left at the output. This residual correlation can be exploited at the channel decoder to improve the decoding process [13], [2], [1]. For this application, it was observed that the speech parameters in the n^{th} speech frame, $e(n), c_1(n), c_2(n), ..., c_{11}(n)$, are correlated with the corresponding parameters in the previous speech frame. In this section, the residual correlation between the speech frames is being exploited to enhance the channel decoding process.

Our approach is similar to the channel decoding scheme presented in [1]. For this scheme to work, the availability of correct channel state information is assumed at the receiver. In this section, we will consider binary phase shift keying (BPSK) modulation rather than the binary DPSK scheme considered in previous sections, so that indeed the channel state information is available at the receiver. In this case, $a(n)$ and $d(n)$ are, respectively, the input and output of the channel encoder and the received signal as given in Eq. (1) is $y(n) = A\beta(n)d(n) + \nu(n)$, where $\beta(n)$ and the channel SNR are assumed to be available at the receiver. After coherent demodulation the input to the channel decoder is $z(n) = \mathrm{Re}\{\beta^*(n)y(n)\}$. The channel decoding algorithm optimizes the following criterion:

$$\max_{\{a(n)\}} p(\{z(n)\}, \{d(n)\}) \tag{3}$$

where $\{a(n)\}$ denotes the sequence of symbols $a(n)$. Let us assume $\{a(n)\}$ is a Markov process and is uniformly distributed, then it can be shown that Eq. (3) can be expressed as

$$\max_{\{a(n)\}} \sum_{n,i} \left\{ \sigma^2 \ln\{p(a(n)|a(n-1))\} - z^i(n)d^i(n) \right\} \tag{4}$$

where $\sigma^2 = E\{|\nu(n)|^2\}$, $d^i(n)$ now denote the i^{th} coded bit for input $a(n)$ and $z^i(n)$ is the channel decoder input for $d^i(n)$. This is similar to the Viterbi algorithm with the path metric modified by $\sigma^2 \ln\{p(a(n)|a(n-1))\}$. Intuitively,

Speed	Error Protection	SNR			
[km/h]	Scheme	15 dB	10 dB	7 dB	5 dB
10	UEP1	7.6	11.0	22.0	36.8
	UEP1 + conc.	7.6	11.0	21.7	36.4
	UEP1 + soft.	7.5	9.9	16.8	28.5
	UEP2	7.4	10.2	17.3	28.3
	UEP2 + soft.	7.3	8.7	12.9	19.9
	UEP3	7.8	12.5	25.1	41.9
	UEP3 + conc.	7.7	12.6	24.9	41.4
	UEP3 + soft.	7.3	9.2	14.3	24.1
50	UEP1	7.2	8.0	15.3	34.8
	UEP1 + conc.	7.2	8.0	15.2	34.9
	UEP1 + soft.	7.2	7.8	12.7	24.6
	UEP2	7.1	7.7	10.6	21.0
	UEP2 + soft.	7.1	7.4	8.7	13.9
	UEP3	7.3	8.9	18.2	41.8
	UEP3 + conc.	7.3	8.9	18.2	41.4
	UEP3 + soft.	7.2	7.6	10.0	16.9
100	UEP1	7.4	7.4	11.8	34.9
	UEP1 + conc.	7.4	7.4	11.8	37.0
	UEP1 + soft.	7.4	7.4	11.2	25.5
	UEP2	7.3	7.3	8.7	16.7
	UEP2 + soft.	7.3	7.2	7.7	11.1
	UEP3	7.2	7.6	15.8	44.8
	UEP3 + conc.	7.2	7.6	15.8	44.8
	UEP3 + soft.	7.2	7.4	8.6	15.0

TABLE VI

WERS (%) FOR DIFFERENT MOBILE SPEEDS AND SNRS FOR A FADING CHANNEL. "UEP+CONC." DENOTES UNEQUAL ERROR PROTECTION WITH CONCEALMENT, WHILE "UEP+SOFT." DENOTES UNEQUAL ERROR PROTECTION WITH SOFT-FEATURES.

the above criterion gives more weight to the *a priori* information, $p(a(n)|a(n-1))$, in very noisy conditions, and relies more on the channel decoder input, $z^i(n)$, for low noise channel conditions.

The above decoding procedure is applied to the MSBs of the speech parameters, $e^0(n)$, $c_1^0(n), ..., c_{11}^0(n)$. For the task outlined in Section V, a high correlation was observed for the MSBs. The transition probabilities for these coefficients are depicted in Table VIII. The transition probabilities are obtained by averaging over all the utterances for this task. These probabilities can be made available at the receiver. For the MSBs of the 12 speech parameters, it was observed that $p(a(n)=0) = p(a(n)=1) = 0.5$ and $p(a(n)=0|a(n-1)=0) = p(a(n)=1|a(n-1)=1)$, where $a(n)$ represents any of the MSBs in the n^{th} speech frame.

Speed (km/h)	Error Protection Scheme	SNR 15 dB	10 dB	7 dB	5 dB
10	UEP1, UEP3 – L1_1	0.00115	0.01305	0.04558	0.09497
	UEP1, UEP3 – L1_2	0.00159	0.01621	0.05463	0.11013
	UEP1, UEP3 – L2	0.00496	0.03865	0.10908	0.19098
	UEP1, UEP3 – L3	0.01446	0.04288	0.07863	0.11376
	UEP2 – L1_1	0.00107	0.01195	0.04202	0.08754
	UEP2 – L1_2	0.00133	0.01442	0.04941	0.10200
	UEP2 – L2	0.00161	0.01614	0.05456	0.10986
	UEP2 – L3	0.01447	0.04285	0.07862	0.11374
50	UEP1, UEP3 – L1_1	0.00001	0.00131	0.01433	0.05480
	UEP1, UEP3 – L1_2	0.00002	0.00217	0.02163	0.07542
	UEP1, UEP3 – L2	0.00026	0.01279	0.07813	0.18947
	UEP1, UEP3 – L3	0.01445	0.04285	0.07853	0.11365
	UEP2 – L1_1	0.00000	0.00110	0.01263	0.04851
	UEP2 – L1_2	0.00001	0.00161	0.01691	0.06331
	UEP2 – L2	0.00001	0.00208	0.02081	0.07350
	UEP2 – L3	0.01448	0.04286	0.07853	0.11365
100	UEP1, UEP3 – L1_1	0.00000	0.00011	0.00465	0.03383
	UEP1, UEP3 – L1_2	0.00000	0.00028	0.00931	0.05458
	UEP1, UEP3 – L2	0.00001	0.00483	0.06174	0.19370
	UEP1, UEP3 – L3	0.01472	0.04309	0.07869	0.11380
	UEP2 – L1_1	0.00000	0.00009	0.00389	0.02873
	UEP2 – L1_2	0.00000	0.00014	0.00585	0.04129
	UEP2 – L2	0.00000	0.00023	0.00818	0.05213
	UEP2 – L3	0.01471	0.04302	0.07862	0.11371

TABLE VII

BERS FOR DIFFERENT MOBILE SPEEDS AND SNRS FOR A RAYLEIGH FADING CHANNEL. L1_1, L1_2, L2 AND L3 ARE THE DIFFERENT ERROR PROTECTION LEVELS.

Component (MSB)	e^0	c_1^0	c_2^0	c_3^0	c_4^0	c_5^0	c_6^0	$c_7^0 - c_{11}^0$
$p(a(n) = 0 \mid a(n-1) = 0)$	0.95	0.86	0.80	0.80	0.75	0.75	0.78	0.73

TABLE VIII

CONDITIONAL PROBABILITIES FOR THE MSBS OF FEATURE COMPONENTS IN SUCCESSIVE SPEECH FRAMES. $a(n)$ IS THE MSB AND n THE SPEECH FRAME INDEX.

These transition probabilities were used in the decoding algorithm in Eq. (4) for the error protection scheme UEP2 defined in Section III-B. There are 8 speech frames in a single channel coding and interleaving frame; therefore, for

Speed	Channel Decoding	SNR			
[km/h]	Scheme	5 dB	3 dB	1.5 dB	0 dB
10	UEP2	11.5	17.2	25.1	37.0
	UEP2 + src.	9.3	13.2	19.1	28.3
	UEP2 + soft.	9.4	13.4	17.9	26.1
	UEP2 + src. + soft.	8.2	10.2	13.0	17.8
50	UEP2	8.2	10.6	16.7	31.6
	UEP2 + src.	7.8	9.0	11.9	21.0
	UEP2 + soft.	7.6	8.8	12.0	19.6
	UEP2 + src. + soft.	7.7	8.3	9.7	13.1
100	UEP2	7.4	8.7	13.0	28.7
	UEP2 + src.	7.4	8.0	10.1	17.7
	UEP2 + soft.	7.4	8.0	10.0	17.4
	UEP2 + src. + soft.	7.4	7.9	9.2	12.5

TABLE IX

CHANNEL DECODING WITH AND WITHOUT SOURCE INFORMATION. WERS (%) FOR DIFFERENT SPEEDS AND SNRS FOR A RAYLEIGH FADING CHANNEL. "UEP+SRC." DENOTES UNEQUAL ERROR PROTECTION WITH SOURCE INFORMATION, WHILE "UEP+SRC.+SOFT." DENOTES UNEQUAL ERROR PROTECTION WITH SOURCE INFORMATION AND SOFT-FEATURES.

the 8 speech frames the L1 level bits are arranged in the following manner:

$$e^0(n), e^0(n+1), ..., e^0(n+7); e^1(n), e^1(n+1), ..., e^1(n+7);$$
$$c_1^0(n), c_1^0(n+1), ..., c_1^0(n+7); .....; c_5^0(n), c_5^0(n+1), ..., c_5^0(n+7).$$

With this bit arrangement the Markov transitions that exists in the bit stream can be exploited by the decoding algorithm. Note that $c_1^0(n)$ and $c_1^0(n+1)$ are correlated with the transition probability depicted in Table VIII; however, the transitions from one parameter to another, e.g., $c_1^0(n+7)$ to $c_2^0(n)$, are not correlated. The L2 level bits, the MSBs of $c_6(n), ..., c_{11}(n)$ given in the third row of Table II, are also correlated with the corresponding bits of the previous speech frame. The L2 level bits for the 8 speech frames are arranged as

$$c_6^0(n), c_6^0(n+1), ..., c_6^0(n+7); c_7^0(n), c_7^0(n+1), ..., c_7^0(n+7); c_{11}^0(n), c_{11}^0(n+1), ..., c_{11}^0(n+7);$$

Again, the successive bits of each speech parameter will exhibit a Markov-type transition and this is incorporated in the decoding algorithm given by Eq. (4).

Tables IX and X depict the WER and the corresponding BER for channel decoding with and without *a priori* source information. Note that a Rayleigh fading channel with BPSK modulation scheme is assumed. It can be seen that decoding with source information gives a 10-20% absolute improvement in WER, with greater gains at slower speeds and lower SNR levels. When combined with the soft-feature error concealment technique the overall improvement in the WER is about 30-40%. For channel decoding with source information, significant improvements in BER for the L1_1 level and some gains for L1_2 and L2 are shown in Table X. At 10 km/h and 0 dB SNR, the improvement in BER for L1_1 is more than 1.5 dB. Note that the L1_2 level does not have *a priori* source information to improve its decoding process. However, since the L1_2 level bits are placed between L1_1 and L2 bits before the channel encoder, an improvement in the BER for L1_2 is also observed. The BER gains for L2 level bits are quite small because of the relatively small transition probabilities for these bits.

Speed (km/h)	Channel Decoding Scheme	SNR 5 dB	3 dB	1.5 dB	0 dB
10	UEP2 – L1_1	0.01818	0.04164	0.07323	0.12010
	– L1_2	0.02281	0.05156	0.08903	0.14418
	– L2	0.02601	0.05763	0.09837	0.15661
	– L3	0.06064	0.08704	0.11138	0.13916
	UEP2 + src. – L1_1	0.00871	0.02079	0.03771	0.06411
	– L1_2	0.01875	0.04295	0.07561	0.12469
	– L2	0.02305	0.05150	0.08828	0.14173
	– L3	0.06064	0.08704	0.11138	0.13916
50	UEP2 – L1_1	0.00240	0.01190	0.03395	0.08200
	– L1_2	0.00356	0.01700	0.04673	0.10981
	– L2	0.00480	0.02147	0.05682	0.12704
	– L3	0.06056	0.08691	0.11120	0.13895
	UEP2 + src. – L1_1	0.00086	0.00439	0.01317	0.03447
	– L1_2	0.00242	0.01219	0.03466	0.08486
	– L2	0.00397	0.01817	0.04857	0.11018
	– L3	0.06056	0.08691	0.11120	0.13895
100	UEP2 – L1_1	0.00029	0.00337	0.01649	0.06054
	– L1_2	0.00048	0.00549	0.02575	0.08882
	– L2	0.00079	0.00822	0.03510	0.10930
	– L3	0.06049	0.08683	0.11114	0.13897
	UEP2 + src. – L1_1	0.00007	0.00108	0.00558	0.02201
	– L1_2	0.00029	0.00352	0.01712	0.06248
	– L2	0.00064	0.00675	0.02906	0.09210
	– L3	0.06049	0.08683	0.11114	0.13897

TABLE X

CHANNEL DECODING WITH AND WITHOUT SOURCE INFORMATION (DENOTED "+ SRC."). BERS FOR DIFFERENT MOBILE SPEEDS AND SNRS FOR A RAYLEIGH FADING CHANNEL.

VII. CONCLUDING REMARKS

In this paper, a speech recognition codec was proposed for distributed ASR over wireless channels. The relevant speech parameters were extracted at the wireless terminal, error protected and transmitted over a 9.6 kb/s wireless channel. Since the speech recognizer has different levels of sensitivity to errors in each of the speech parameters, three unequal error protection schemes were examined and it was shown that the second scheme (UEP2) gives the best ASR performance gains. It was also shown, that acceptable WERs of about 10% can be obtained using UEP2 for a Gaussian channel at 2 dB SNR, and for a Rayleigh fading channel at 10 km/h and 10 dB SNR. We demonstrated that the simple error concealment technique which repeats the previously received error-free sub-frame is not effective for fading channels. However, the soft-feature error concealment technique reduces the error rate up to a factor of four, depending on channel conditions. Finally, a channel decoding technique was presented, which makes use of the

residual correlations that exists in the source bit stream, and demonstrated the significant WER reduction that can be obtained at low SNRs and slower mobile speeds.

REFERENCES

[1] S. A. Al-Semari, F. Alajaji, and T. Fuja, "Sequence MAP Decoding of Trellis Codes for Gaussian and Rayleigh Channels," *IEEE Trans. on Veh. Technology*, vol. 48, pp. 1130–1140, July 1999.

[2] F. Alajaji, N. Phamdo, and T. Fuja, "Channel Codes That Exploit the Residual Redundancy in CELP-Encoded Speech," *IEEE Trans. on Speech and Audio Processing*, vol. 4, pp. 325–336, Sept 1996.

[3] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal Decoding of Linear Codes for Minimizing Symbol Error Rate," *IEEE Trans. on Inform. Theory*, vol. 20, pp. 284–287, March 1976.

[4] M. Cooke, P. Green, L. Josifovski, and A. Vizinho, "Robust ASR with Unreliable Data and Minimal Assumption," in *Proceedings, Robust Methods for Speech Recognition in Adverse Conditions*, (Tampere, Finland), pp. 195–198, 1999.

[5] V. Digalakis, L. Neumeyer, and M. Perakakis, "Quantization of cepstral parameters for speech recognition over the world wide web," in *Proc. Internat. Conf. on Acoust., Speech, and Signal Process.*, (Seattle, Washington), May 1998.

[6] S. Dufour, C. Glorion, and P. Lockwood, "Evaluation of the Root-Normalized Front-End (RN⊥LFCC) for Speech Recognition in Wireless GSM Network Environments," in *1996 International Conference on Acoustics, Speech and Signal Processing*, (Atlanta, Georgia), pp. 77–80, 1996.

[7] S. Euler and J. Zinke, "The Influence of Speech Coding Algorithms on Automatic Speech Recognition," in *1998 International Conference on Acoustics, Speech and Signal Processing*, (Adelaide, Australia), pp. 621–624, 1994.

[8] European Telecommunications Standardization Institute. *GSM Recommendation*, 1988.

[9] L. Fissore, F. Ravera, and C. Vair, "Speech Recognition Over GSM: Specific Features and Performance Evaluation," in *Robust Methods for Speech Recognition in Adverse Conditions*, (Tempere, Finland), pp. 127–130, 1999.

[10] A. Gallardo-Antolin, F. D. de Maria, and F. Valverde-Albacete, "Avoiding Distortions Due to Speech Coding and Transmission Errors in GSM ASR Tasks," in *1999 International Conference on Acoustics, Speech and Signal Processing*, (Phoenix, Arizona), 1999.

[11] P. Haavisto, "Speech Recognition for Mobile Communications," in *Proceedings, Robust Methods for Speech Recognition in Adverse Conditions*, (Tampere, Finland), pp. 15–18, 1999.

[12] J. Hagenauer, "Rate-Compatible Punctured Convolutional Codes (RCPC Codes) and their Applications," *IEEE Trans. on Communications*, vol. 36, pp. 389–400, April 1988.

[13] J. Hagenauer, "Source-Controlled Channel Decoding," *IEEE Trans. on Communications*, vol. 43, pp. 2449–2457, September 1995.

[14] J. Hagenauer, N. Seshadri, and C.-E. W. Sundberg, "The performance of Re-Compatible Punctured Convolutional Codes for Digital Mobile Radio," *IEEE Trans. on Communications*, vol. 38, July 1990.

[15] L. Karray, A. B. Jelloun, and C. Mokbel, "Solutions for Robust Recognition Over the GSM Cellular Network," in *1998 International Conference on Acoustics, Speech and Signal Processing*, (Seattle, Washington), pp. 261–264, 1998.

[16] B. T. Lilly and K. K. Paliwal, "Effects of Speech Coders on Speech Recognition Performance," in *ICSLP'96*, (Philadelphia, Pennsylvania), pp. 2344–2347, 1996.

[17] S. Lin and D. J. Costello, *Error Control Coding: Fundamentals and Applications*. Prentice Hall, 1994.

[18] S. P. Lloyd, "Least Squares Quantization of PCM," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 129–136, 1982.

[19] C. Mokbel, L. Mauuary, L. Karray, D. Jouvet, J. Monne, J. Simonin, and K. Bartkova, "Towards Improving ASR robustness for PSN and GSM telephone applications," *Speech Communications*, vol. 23, pp. 141–159, 1997.

[20] A. Potamianos, V. Weerackody, and W. Reichl, "Soft-feature decoding for distributed speech recognition over wireless channels," *in preparation*, 2000.

[21] L. R. Rabiner and B.-H. Juang, *Fundamentals of Speech Recognition*. Englewood Cliffs, NJ: Prentice–Hall, 1993.

[22] L. R. Rabiner and B.-H. Juang, *Fundamentals of Speech Recognition*. Englewood Cliffs, NJ: Prentice–Hall, 1993.

[23] G. N. Ramaswamy and P. S. Gopalakrishnan, "Compression of acoustic features for speech recognition in network environments," in *Proc. Internat. Conf. on Acoust., Speech, and Signal Process.*, (Seattle, Washington), May 1998.

[24] W. Reichl and W. Chou, "Decision Tree State Tying Based on Segmental Clustering for Acoustic Modeling," in *1998 International Conference on Acoustics, Speech and Signal Processing*, (Seattle, Washington), 1998.

[25] Telecommunications Industry Association. *EIA/TIA Interim Standard, Cellular System Dual-Mode Mobile Station Base Station Compatibility Standard IS-54B, EIA/TIA*, 1992.

[26] TIA/EIA Interim Standard 95: *Mobile Station-Base Station Compatibility Standard for Dual-Mode Wideband Spread Spectrum Cellular Standard*, July 93.

[27] Q. Zhou and W. Chou, "An Approach to Continuous Speech Recognition Based on Layered Self-Adjusting Decoding Graph," in *1997 International Conference on Acoustics, Speech and Signal Processing*, (Munich, Germany), pp. 1779–1782, 1997.