Low-Bitrate Distributed Speech Recognition for Packet-Based and Wireless Communication

Alexis Bernard, *Student Member, IEEE*, and Abeer Alwan, *Senior Member, IEEE*

***Abstract*—In this paper, we present a framework for developing source coding, channel coding and decoding as well as erasure concealment techniques adapted for distributed (wireless or packet-based) speech recognition. It is shown that speech recognition as opposed to speech coding, is more sensitive to channel errors than channel erasures, and appropriate channel coding design criteria are determined. For channel decoding, we introduce a novel technique for combining at the receiver soft decision decoding with error detection. Frame erasure concealment techniques are used at the decoder to deal with unreliable frames. At the recognition stage, we present a technique to modify the recognition engine itself to take into account the time-varying reliability of the decoded feature after channel transmission. The resulting engine, referred to as weighted Viterbi recognition, further improves recognition accuracy. Together, source coding, channel coding and the modified recognition engine are shown to provide good recognition accuracy over a wide range of communication channels with bitrates of 1.2 kbps or less.**

***Index Terms*—Automatic speech recognition, distributed speech recognition (DSR), joint channel decoding-speech recognition, soft decision decoding, weighted Viterbi algorithm, wireless and packet (IP) communication.**

I. Introduction

In DISTRIBUTED speech recognition (DSR) systems, speech features are acquired by the client and transmitted to the server for recognition. This enables low power/complexity devices to perform speech recognition. Applications include voice-activated web portals, menu browsing and voice-operated personal digital assistants.

This paper investigates channel coding, channel decoding, source coding and speech recognition techniques suitable for DSR systems over error prone channels (Fig. 1). The goal is to provide high recognition accuracy over a wide range of channel conditions with low bitrate, delay and complexity for the client.

Wireless communications is a challenging environment for speech recognition. The communication link is characterized by time-varying, low signal-to-noise ratio (SNR) channels. Previous studies have suggested alleviating the effect of channel errors by adapting acoustic models [1] and automatic speech recognition (ASR) front-ends [2] to different channel conditions, or by modeling GSM noise and holes [3]. Other studies analyzed the effect of random and burst errors in the GSM bitstream for remote speech recognition applications [4]. Finally, [5] and [6] evaluate the reliability of the decoded feature to provide robustness against channel errors. Similarly, packet switched networks constitute a difficult environment. The communication link in IP based systems is characterized by packet losses, mainly due to congestion at routers. Packet loss recovery techniques including silence substitution, noise substitution, repetition and interpolation [7]–[9].

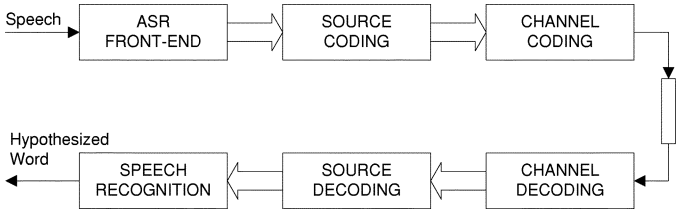


Fig. 1. Block diagram of a remote speech recognition system.

In terms of source coding for DSR, there are three possible approaches. The first approach bases recognition on the decoded speech signal, after speech coding and decoding. However, it is shown in [10]–[12] that this method suffers from significant recognition degradation at low bitrates. A second approach is to build a DSR engine based on speech coding parameters without re-synthesizing the speech signal [13]–[16]. The third approach performs recognition on quantized ASR features, and provides a good tradeoff between bitrate and recognition accuracy [17]–[20]. This paper presents contributions in several areas of DSR systems based on quantized ASR features.

In the area of *channel coding*, it is first explained and experimentally verified that speech recognition, as opposed to speech coding, is more sensitive to channel errors than channel erasures. Two types of channels are analyzed, independent and bursty channels. Second, efficient channel coding techniques for error detection based on linear block codes are presented.

In the area of *channel decoding*, the merits of soft and hard decision decoding are discussed, and a new technique for performing error detection with soft decision decoding is presented. The soft decision channel decoder, which introduces additional complexity only at the server, is shown to outperform the widely-used hard decision decoding.

In the area of *speech recognition*, the recognition engine is modified to include a time-varying weighting factor depending on the quality of each decoded feature after transmission over time-varying channels. Following frame erasure concealment,

Manuscript received September 25, 2001; revised August 7, 2002. This work was supported in part by the NSF, HRL, STM, and Broadcom through the University of California Micro Program. Portions of this work were presented at the IEEE International Conference on Acoustics, Speech and Signal Processing, Salt Lake City, UT, May 7–11, 2001, and the Eurospeech conference in Aalborg, Denmark, September 3–7, 2001. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Harry Printz.

The authors are with the Speech Processing and Auditory Perception Laboratory, Electrical Engineering Department, University of California, Los Angeles, CA 90095-1594 USA (e-mail: abernard@icsl.ucla.edu; alwan@icsl.ucla.edu).

Digital Object Identifier 10.1109/TSA.2002.808141

an estimate of the quality of the substituted features is taken into account using a weighted Viterbi recognizer (WVR). Together, erasure concealment and WVR improves robustness of the DSR system against channel noise, extending the range of channel conditions over which wireless or internet-based speech recognition can be sustained.

Source coding, channel coding, and speech recognition techniques are then combined to provide high recognition accuracy over a large range of channel conditions for two types of speech recognition features: perceptual linear prediction (PLP) and Mel frequency cepstral coefficients (MFCC).

This paper is organized as follows. Section II analyzes the effect of channel errors and erasures on recognition accuracy. Section III provides a description of the channel encoders used to efficiently protect the recognition features. In Section IV, different channel decoding techniques are presented. Section V presents the weighted Viterbi recognition (WVR) algorithm. Techniques alleviating the effect of erasures using WVR are proposed in Section VI. Finally, Section VII illustrates the performance of the overall speech recognition system applied to quantized PLP and MFCC features.

II. Effect of Channel Erasures and Errors

In this section, we study how channel errors and erasures affect the Viterbi speech recognizer. We then present techniques for minimizing recognition degradation due to transmission of speech features over noisy channels.

Throughout this paper, speech recognition experiments consist of continuous digit recognition based on 4 kHz bandwidth speech signals. Training is done using speech from 110 males and females from the Aurora-2 database [18] for a total of 2200 digit strings. The feature vector consists of PLP or Mel frequency cepstral coefficients with the first and second derivatives. As specified by the Aurora-2 ETSI standard [18], hidden Markov (HMM) word models contain 16 states with 6 mixtures each, and are trained using the Baum–Welch algorithm assuming a diagonal covariance matrix. Recognition tests contain 1000 digit strings spoken by 100 speakers (male and female) for a total of 3241 digits.

A. Effect of Channel Erasures and Errors on DSR

The emphasis in remote ASR is recognition accuracy and not playback. Recognition is made by computing feature vectors' likelihood time and by selecting the element in the dictionary that most likely produced that sequence of observations. The nature of this task implies different criteria for designing channel encoders and decoders than those used in speech coding/playback applications.

The likelihood of observing a given sequence of features given a hidden Markov model is computed by searching through a trellis for the most probable state sequence. The Viterbi algorithm (VA) presents a dynamic programming solution to find the most likely path through a trellis. For each state j, at time t, the likelihood of each path is computed by multiplying the transition probabilities a_{ij} between states and the output probabilities $b_j(\boldsymbol{o}_t)$ along that path. The partial likelihood $\phi_{j,t}$ is computed efficiently using the following recursion:

$$\phi_{j,t} = \max_i [\phi_{i,t-1} a_{ij}] b_j(\boldsymbol{o}_t). \tag{1}$$

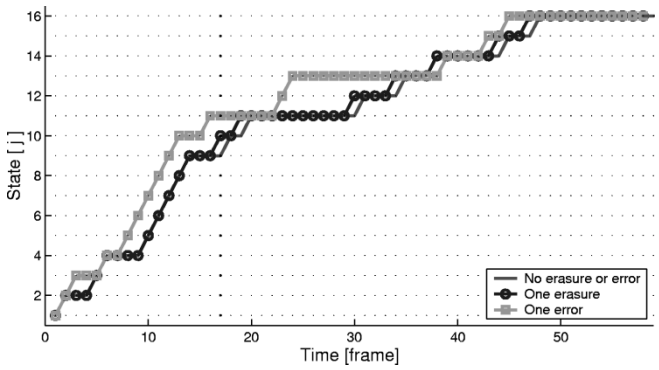


Fig. 2. Illustration of the consequences of a channel erasure and error on the most likely paths taken in the trellis by the received sequence of observations, given a 16-state word digit model. The erasure and error occur at frame number 17.

The probability of observing the N_F-dimensional feature $\boldsymbol{o}_t$ is

$$b_j(\boldsymbol{o}_t) = \sum_{m=1}^{N_M} c_m \frac{1}{\sqrt{(2\pi)^{N_F}|\boldsymbol{\Sigma}|}} \cdot \exp\left(-\tfrac{1}{2}(\boldsymbol{o}_t - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{o}_t - \boldsymbol{\mu})\right) \tag{2}$$

where N_M is the number of mixture components, c_m is the mixture weight, and the parameters of the multivariate Gaussian mixture are its mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

Fig. 2 analyzes the effect of a channel error and erasure in the VA. Assume first a transmission free of channel errors. The best path through the trellis is the line with no marker. Assume now that a channel *error* occurs at time t. The decoded feature is $\hat{\boldsymbol{o}}_t$ as opposed to $\boldsymbol{o}_t$ and the associated probabilities for each state j may differ considerably ($b_j(\hat{\boldsymbol{o}}_t) \neq b_j(\boldsymbol{o}_t)$), which will disturb the state metrics $\phi_{j,t}$. A large discrepancy between $b_j(\hat{\boldsymbol{o}}_t)$ and $b_j(\boldsymbol{o}_t)$ can force the best path in the trellis to branch out from the error-free best path. Consequently, many features may be accounted for in the overall likelihood computation using the state model $\hat{j}$ instead of the correct state model j, which will once again modify the probability of observation since $b_{\hat{j}}(\boldsymbol{o}_{t+k}) \neq b_j(\boldsymbol{o}_{t+k})$.

On the other hand, channel *erasures* have little effect on likelihood computation. State metrics are not disturbed since the probability of the missing observation cannot be computed. Also, note that not updating the state metrics ($\phi_{j,t} = \phi_{j,t-1}$) is not as likely to create a path split between the best paths with and without an erasure as a channel error. Hence, channel erasures typically do not propagate through the trellis.

B. Simulations of Channel Erasures and Errors

In this section, we simulate the effects of channel erasures and channel errors on DSR.

Fig. 3 illustrates the effect of randomly inserted channel erasures and errors in the communication between the client and the server. The feature vector transmitted consists of 5 PLP cepstral coefficients, enough to represent two observable peaks in

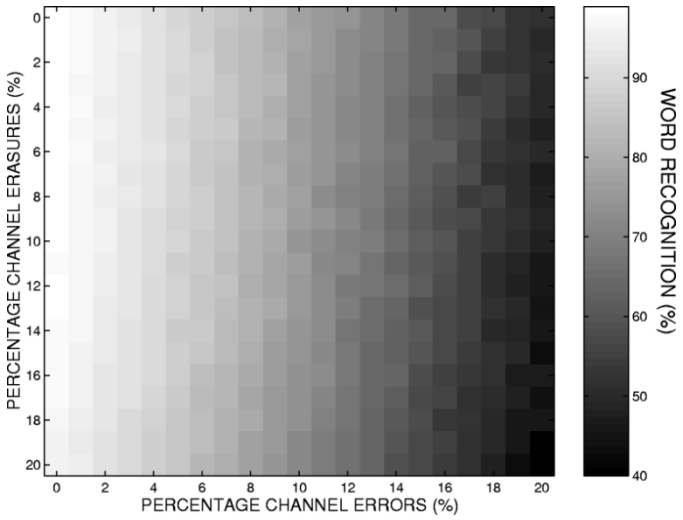


Fig. 3. Simulation of the effect of channel erasures and errors on continuous digit recognition performance using the Aurora-2 database and PLP features. Recognition accuracies are represented in percent on a gray scale.

the perceptual spectrum and the spectral tilt. Erasures are simulated by removing the corresponding frame from the observation sequence. Channel errors, on the other hand, are simulated by replacing the feature vector with another vector, chosen randomly according to the statistical distribution of the features. This simulation technique has the merit of being independent of the source coding algorithm. It is valid especially for low-bitrate quantization schemes, which are highly sensitive to channel errors.

Fig. 3 shows that channel errors, which propagate through the trellis, have a disastrous effect on recognition accuracy, while the recognizer is able to operate with almost no loss of accuracy with up to 15% of channel erasures. This confirms results obtained in [19] for isolated digit recognition based on PLP coefficients and in [5] for MFCCs. Note that computation of the temporal derivatives at the receiver accentuates error propagation.

The results indicate that a very important attribute of any channel encoder designed for remote recognition applications should be error detection more than error correction. Sections III and IV present innovative techniques to maximize error detection capabilities of linear block codes suitable for DSR applications. For the remainder of this section, we assume that all transmission errors are detected and replaced by erasures. Models for erasure channels are presented next.

C. Gilbert–Elliot Models for Erasure Channels

Two types of erasure channels are analyzed. In the first type, channel erasures occur independently. In the second type, channel erasures occur in bursts, which is typically the case for correlated fading channels in wireless communication or IP based communication systems, where fadings or network congestion may cause a series of consecutive packets to be dropped.

For independent-erasure channels, erasures are inserted randomly with a given probability. A classic model for bursty channels is the Gilbert–Elliot model [21], in which the transmission is modeled as a Markov system where the channel is assigned one of two states: *good* or *bad*. With such a model characterized by the state transition probabilities P_{GB} and P_{BG}, there is a probability $P_G = P_{BG}/(P_{BG} + P_{GB})$ to be in the good state and a probability $P_B = P_{GB}/(P_{GB} + P_{BG})$ to be in the bad state. If the probabilities of channel erasures are P_{E_G} and P_{E_B} for the good and bad state, respectively, the overall average probability of erasure is: $P_E = P_G P_{E_G} + P_B P_{E_B}$.

TABLE I
GILBERT–ELLIOT TEST CHANNELS (PROBABILITIES IN %)

P_{GB}/P_{BG}	2.5/20	2.5/15	5/20	2.5/10	1.25/5	5/15	10/20
P_B	11.1	14.3	20.0	20.0	20.0	25.0	33.3
P_E	9.7	12.3	16.8	16.8	16.8	20.7	27.3
$\overline{L_b}$	4.0	5.6	4.0	9.0	19.0	5.6	4.0

Throughout this paper, P_{E_G} will be considered to be equal to 0.01 and P_{E_B} is set to 0.80. Different types of bursty channels are analyzed, depending on P_{GB} and P_{BG}, which in turn determine how bursty the channel is. Table I summarizes the properties of the bursty channels studied, including the probability (in percent) of being in the bad state (P_B), the overall probability of erasure, (P_E), and the average length (in frames) of a burst of erasures $(\overline{L}_b)$.

The Gilbert–Elliot model parameters are selected based on values reported in the literature on Gilbert models for packet-based (IP) networks [22], [23] and wireless communication channels [24]–[26].

III. CHANNEL CODING FOR DSR SYSTEMS

The analysis in Section II indicates that the most important requirement for a channel coding scheme for DSR is low probability of undetected error (<0.5%) and large enough probability of correct decoding (>90%). This section presents techniques to detect most channel errors. Corrupted frames are then ignored (erased) and frame erasure concealment techniques presented in Section VI can be applied.

For *packet-based* transmission, frames are typically either received or lost, but not in error. Frame erasures can be detected by analyzing the ordering of the received packet and there is no need for sophisticated error detection techniques.

With *wireless communication*, transmitted bits $\boldsymbol{x}$ are altered during transmission. Based on the values of the received bits $\boldsymbol{y}$, the receiver can either correctly decode the message (CD for correct decoding), detect a transmission error (ED for error detection) or fail to detect such error (UE for undetected error).

Since the number of source information bits necessary to code each frame can be very low (6–40 bits/frame) for efficient ASR feature coding schemes [19], linear block codes are favored over convolutional or trellis codes for delay and complexity considerations, as well as for their ability to provide error detection for each frame independently.

A. Error Detecting Linear Block Codes

An (N, K) linear block code maps K information bits into N bits $(N > K)$. The larger the number of redundancy bits $(R = N-K)$, the larger the minimum distance $(d_{\min})$ between any two of the 2^K valid codewords. In order to guarantee the best possible recognition rate over a wide range of channel conditions, a combination of different block codes is used. More information bits (K) are used for high SNR channels while more redundancy bits (R) are used for low SNR channels.

For good channel conditions, Single Error Decoding (SED) codes, which detect any one bit error in the N bits received codeword, are sufficient. A minimum Hamming distance of $d_{\min} = 2$ is sufficient to form an SED code. However, when there are 2 errors among the N received bits, SED codes may fail to detect the error. To increase channel protection, Double Error Detection (DED) codes are utilized. Any linear block code with $d_{\min} = 3$ can be used to correct single error events [Single Error Correcting (SEC) code] or to detect all one and two-bit error events (DED). For our application, since residual channel errors degrade recognition accuracy more significantly than channel erasures, all codes with $d_{\min} = 3$ will be used as DED codes. Finally, codes with $d_{\min} = 4$ will be used as Triple Error Detecting (TED) codes as opposed to SEC/DED codes.

B. Search for Good Codes

Exhaustive searches over all possible linear block codes were run for all dimensions of interest, i.e., $7 \le K \le 10$ and $10 \le N \le 12$, in order to find the codes with the best distance spectrum. For the particular case of $R = 1$, i.e., a $(K+1, K)$ code, $d_{\min} = 2$, the parity matrix P of dimension $1 \times K$ of the code is $P = [1, 1, \cdots, 1, 1]$. The parity matrices P and the minimum Hamming distance $d_{\min}$ for all other codes of interest are given in Table II. Parity matrices are given in hexadecimal notation.

IV. Channel Decoding for DSR Systems

For wireless communications, information bits x_i are transmitted and distorted by the channel $y_i = \alpha(t) \cdot x_i + n(t)$, where $\alpha(t)$ is the complex channel gain and $n(t)$ is the additive white Gaussian noise (AWGN) component. For Rayleigh fading channels, α is Rayleigh distributed. For AWGN channels, $\alpha(t) = 1$. Depending on whether the actual values of the received bits or only their signs are used, the channel decoder is said to perform *soft* or *hard* decision decoding, respectively.

For a discrete memoryless channel, the *likelihood* of receiving the vector $\boldsymbol{y}$ (N bits) given that the codeword $\boldsymbol{x}_m$ was transmitted is given by

$$p(\boldsymbol{y}|\boldsymbol{x}_m) = \prod_{j=1}^{N} p(y_j|x_{mj}) \qquad (0 \le m \le 2^K - 1). \tag{3}$$

A. Hard Decision Decoding

Transmission channels followed by *hard decision* decoding act like a binary symmetric channel (BSC). For AWGN and Rayleigh fading channels, the cross probability of the equivalent BSC is $p = Q(\sqrt{\alpha^2(2E_b/N_0)})$, where E_b denotes the average energy per bit, N_0 is the average noise energy $(\sigma^2 = N_0/2)$ and $Q(x) = \int_x^\infty (1/\sqrt{2\pi})\, e^{-(z^2/2)}\, dz$ is the tail integral of the normal Gaussian distribution. If channel noise statistics are stationary over the transmission of the N-bits codeword, the BSC cross probability is a constant and (3) becomes

$$p(\boldsymbol{y}|\boldsymbol{x}_m) = p^{d_H}(1-p)^{N-d_H}, \qquad (0 \le m \le 2^K - 1)$$

where d_H is the Hamming distance between $\boldsymbol{y}$ and $\boldsymbol{x}_m$. Maximizing $p(\boldsymbol{y}|\boldsymbol{x}_m)$ is equivalent to minimizing the *Hamming* distance d_H between $\boldsymbol{y}$ and $\boldsymbol{x}_m$.

Fig. 4(a) shows a two-dimensional example for decoding a (2, 1) linear block code. The valid codevectors are shown in dark circles. Assume the $(+1, +1)$ codevector was transmitted. If the soft received bits end up in the second or fourth quadrant, the resulting received codevector after bit thresholding is equally distant, in terms of Hamming distance, from two valid codewords. No decision can be made and an erasure is declared (ED). If the received symbol is in the first or third quadrant, the codeword is correctly (CD) or incorrectly decoded (UE for undetected error), respectively.

Typically, hard decision decoding suffers a 2 dB loss compared to soft decision decoding for AWGN channels and about half the diversity for multi-path communications [27].

B. Soft Decision Decoding

Consider next a *soft decision* memoryless channel where the channel input is ± 1 and the channel output is a real number with Gaussian statistics. Specifically, the stationary channel is specified by

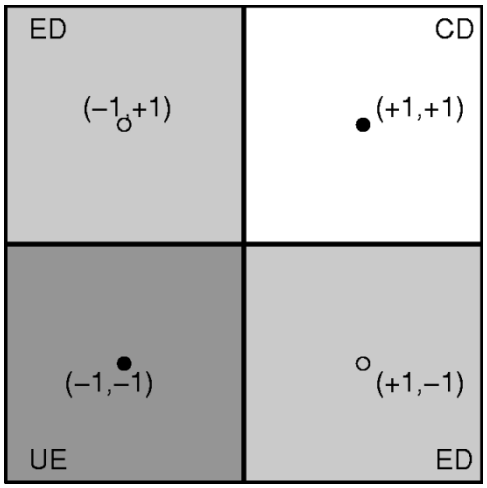
$$p(\boldsymbol{y}|\boldsymbol{x}_m) = \frac{1}{(\sqrt{\pi N_0})^N} \exp\left(-\sum_{j=1}^{N} \frac{(y_j - x_{mj})^2}{N_0}\right). \tag{4}$$

Maximizing $p(\boldsymbol{y}|\boldsymbol{x}_m)$ is equivalent to minimizing the squared *Euclidean* distance ${d_E}^2 = \sum_{j=1}^{N} (y_j - x_{mj})^2$ between $\boldsymbol{y}$ and $\boldsymbol{x}_m$.

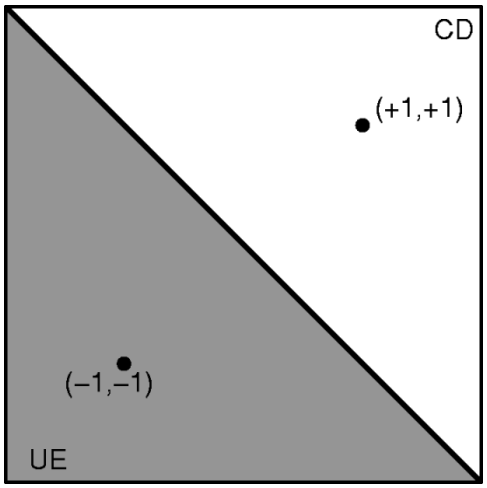
Fig. 4(b) is an example of soft decision decoding for the same (2, 1) code. The maximum likelihood decoder chooses its output to be the codeword for which the Euclidean distance between the received vector $\boldsymbol{y}$ and the codeword $\boldsymbol{x}_m$ is minimum.

TABLE II
Description of the Linear Block Codes Used for Channel Coding Speech Recognition Features

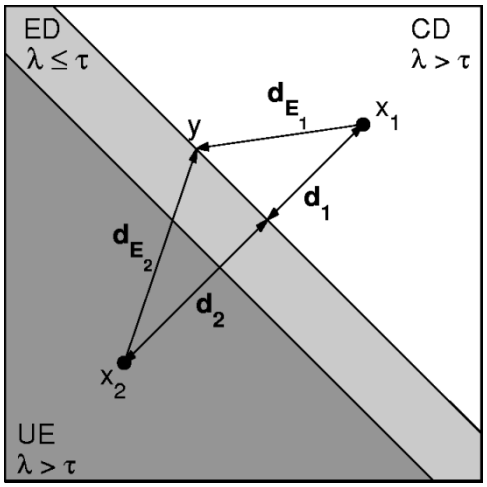
(N,K)	R	P	$d_{\min}$	Type
(12,10)	2	1,1,1,2,2,2,3,3,3,3	2	SED
(12,9)	3	1,2,3,3,4,5,5,6,7	2	SED
(12,8)	4	3,5,6,9,A,D,E,F	3	DED
(10,8)	2	1,1,1,2,2,3,3,3	2	SED
(12,7)	5	07,0B,0D,0E,13,15,19	4	TED
(10,7)	3	1,2,3,4,5,6,7	2	SED



(a)



(b)



(c)

Fig. 4. Illustration of the different decoding strategies. (a) Hard decoding, (b) soft decoding, and (c) λ-soft decoding.

With soft decision decoding, $P_{ED} = 0$, allowing only for correct or erroneous decoding. Consequently, both P_{CD} and P_{UE} increase, which ultimately decreases recognition performance. We propose in the following section a technique to combine the advantage of soft decision decoding with the error detection capability of hard decision decoding.

C. Modified Soft Decision Decoding (λ-Soft)

In order to accept a decision provided by the soft decoder, one would like to evaluate the probability that the decoded codevector was the one transmitted. Such *a posteriori* probability is given by

$$p(\hat{\boldsymbol{x}} = \boldsymbol{x}_m | \boldsymbol{y}) = \frac{\prod_{j=1}^{N} \exp\left[-\frac{(y_j - x_{mj})^2}{N_0}\right]}{\sum_{m'=0}^{2^K-1} \prod_{j=1}^{N} \exp\left[-\frac{(y_j - x_{m'j})^2}{N_0}\right]}$$

which is complex and requires the knowledge of N_0, which is difficult to evaluate.

Another solution is to perform error detection based on the ratio of the likelihoods of the two most probable codevectors. Assuming that all codewords are equiprobable, the ratio of the likelihoods of the two most probable vectors $\boldsymbol{x_1}$ and $\boldsymbol{x_2}$ (the two closest codevectors from the received vector $\boldsymbol{y}$ at Euclidean distances d_{E_1} and d_{E_1} from $\boldsymbol{y}$) is given by

$$\frac{P(\boldsymbol{y}|\boldsymbol{x} = \boldsymbol{x_1})}{P(\boldsymbol{y}|\boldsymbol{x} = \boldsymbol{x_2})} = \exp\left(\frac{d_{E_2}^2 - d_{E_1}^2}{N_0}\right) \tag{5}$$

$$= \exp\left(\frac{D^2}{N_0} \cdot \frac{d_2 - d_1}{D}\right) \tag{6}$$

where D is the Euclidean distance between the two closest codevectors $\boldsymbol{x_1}$ and $\boldsymbol{x_2}$, while d_1 and d_2 are the distances from the projection of the received codevector $\boldsymbol{y}$ to the line joining $\boldsymbol{x_1}$ and $\boldsymbol{x_2}$. The important factor in (6) is

$$\lambda = \frac{d_2 - d_1}{D}. \tag{7}$$

If $\lambda = 0$, both codevectors are equally probable and the decision of the Maximum-Likelihood (ML) decoder should be rejected. If $\lambda = 1$ ($d_1 = 0$, $d_2 = D$), correct decision is almost guaranteed since the block codes used are chosen according to channel conditions so that the minimum Euclidean distance between any two codevectors is at least several times as large as the expected noise ($D^2/N_0 \gg 1$).

Fig. 4(c) shows an example of λ-soft decision decoding the same (2, 1) code. Error detection can be declared when λ is smaller than a threshold τ. Classic soft decision decoding is a particular case of modified soft decision decoding with $\lambda = 0$. The area for error detection grows as λ increases.

D. Comparison of Channel Decoding Performances

For comparison, consider the (10, 7) SED block code of Table II over an independent Rayleigh fading channel at 5 dB SNR. Hard decoding yields $P_{UE} = 0.3\%$, $P_{ED} = 30.2\%$ and $P_{CD} = 69.5\%$. These numbers are insufficient to provide good recognition results. With soft decision decoding, on the other hand, the probability of undetected errors is too large ($P_{UE} = 2.6\%$).

Fig. 5 illustrates the performance of the λ-soft decision decoding schemes for the same code over the same channel for different values of λ. Note first that λ-soft decision decoding with $\lambda = 0$ corresponds to classic soft decision decoding. With increasing λ, however, one can rapidly reduce P_{UE} to the desired

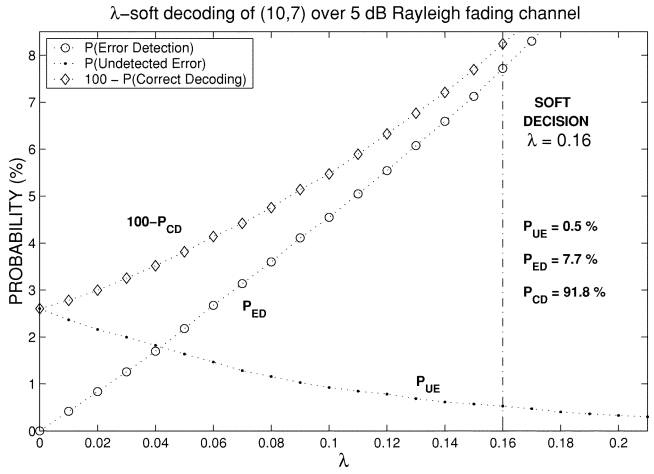


Fig. 5. Illustration of the probability of correct detection (P_{CD}), error detection (P_{ED}) and undetected error (P_{UE}) as a function of the parameter λ when using λ-soft decision decoding of the (10, 7) DED linear block code over an independent Rayleigh fading channel at 5 dB SNR.

values, while still keeping P_{CD} large enough and usually above that of hard decision decoding. For instance, with $\lambda = 0.16$, we have $P_{UE} = 0.5\%$, $P_{ED} = 7.7\%$ and $P_{CD} = 91.8\%$, which results in good recognition accuracy. Note that when P_{UE} decreases, P_{CD} decreases as well, which indicates that a tradeoff must be found.

The probabilities (correct decoding, undetected error and error detection) for the block codes designed for different independent Rayleigh fading channel SNRs are listed in Table III. The value $\lambda = 0.16$ is experimentally found appropriate to keep the number of undetected errors small while the probability of correct decoding remains high.

Note that soft decoding is made at the cost of the additional complexity of computing Euclidean distances for all 2^K codewords. However, note that channel decoding is done at the server, where the complexity of the recognizer prevails.

E. Recognition Experiments

Commonly used ASR features include spectral features such as Mel-Frequency Cepstral Coefficients (MFCCs) and Linear Prediction Cepstral Coefficients (LPCCs). LPCCs can be extracted from a standard linear prediction model or from a Perceptual Linear Prediction model (PLP) [28] which models human auditory perception, and provides good recognition accuracy with a low-dimensional feature vector. This section analyzes recognition results for source and channel coding of PLP features. Results for MFCCs will be presented in Section VII.

In [19] and [29], it is shown that an efficient representation of the PLP spectrum for quantization is using the line spectral frequencies (LSF) of the linear prediction system, to exploit their high inter- and intra-frame correlation. Quantizing LSFs also yields a better representation of the low-order cepstral coefficients, more important for speech recognition. Finally, error sensitivity of the LSFs to quantization noise depends on the LSF order. Appropriate weighting is performed when designing the vector quantizer and during the VQ search.

TABLE III
PROBABILITY OF CORRECT DETECTION (P_{CD}), ERROR DETECTION (P_{ED}) AND UNDETECTED ERROR (P_{UE}) USING HARD, SOFT AND λ-SOFT ($\lambda = 0.16$) DECODING ON RAYLEIGH FADING CHANNELS. $P_{ED} = 0$ FOR SOFT DECODING

Code	SNR	P_{CD}			P_{ED}		P_{UE}		
(N,K)	(dB)	Hard	Soft	λ-soft	Hard	λ-soft	Hard	Soft	λ-soft
(10,9)	10	88.4	97.9	89.3	11.1	10.6	0.6	2.1	0.1
(10,8)	9	86.0	98.8	94.0	13.8	5.8	0.3	1.2	0.1
(10,8)	8	82.6	98.3	92.4	17.0	7.4	0.4	1.7	0.2
(12,9)	7	75.4	98.3	93.0	24.2	6.7	0.3	1.7	0.3
(12,8)	6	70.2	99.3	96.2	29.8	3.8	0.0	0.7	0.0
(12,8)	5	65.0	98.7	94.3	35.0	5.6	0.1	1.3	0.2
(12,8)	4	58.8	97.6	91.6	41.1	8.1	0.1	2.4	0.3
(12,7)	3	52.3	98.6	94.3	47.7	5.5	0.0	1.4	0.2
(12,7)	2	45.1	97.3	91.3	54.9	8.3	0.1	2.7	0.4

TABLE IV
RECOGNITION ACCURACY AFTER LSF QUANTIZATION OF THE PLP COEFFICIENTS USING THE AURORA-2 DATABASE

Bits/Frame	7	8	9	10
Recognition Accuracy (%)	97.07	98.05	98.31	98.48

The five LSFs are computed and quantized every 10 ms using vector quantizers operating at 7 to 10 bits per frame. The receiver decodes the LSFs, derives the LP coefficients from the LSFs, and the cepstral coefficients from the LP coefficients. Predictive VQ and interpolation, used in [19] to further reduce the bitrate, are not used here because they increase sensitivity to transmission errors. Table IV reports recognition results after quantization at different bitrates.

Table V presents recognition accuracy after transmission of the quantized LSFs over an independent Rayleigh fading channel whose equivalent bit error rate ranges from 0.25% to 7%. Depending on the channel conditions, different channel encoders are used. Overall bitrate, including source and channel coding, is 1 kbps for good channels and 1.2 kbps for bad channels.

Table V shows that the proposed technique (λ-soft), which performs soft decision based error detection, outperforms both hard and soft decision decoding. Hard decoding typically keeps P_{UE} small enough, but at the cost of too many frames being erased. Classic soft decision decoding, on the other hand, suffers from the fact that it cannot detect errors, which results in a large proportion of erroneously decoded frames.

V. WEIGHTED VITERBI RECOGNITION (WVR)

With remote recognition, reliability of the decoded features is a function of channel characteristics. When channel charac-

TABLE V
RECOGNITION ACCURACY USING HARD, SOFT AND λ-SOFT DECISION DECODING OVER RAYLEIGH FADING CHANNELS

Code	Bitrate	SNR	BER	ACCURACY(%)		
(N,K)	(kbps)	(dB)	(%)	Hard	Soft	λ-Soft
(10,10)	1.0	19.96	0.25	94.71	94.71	98.32
(10,9)	1.0	13.87	1.00	97.31	96.35	98.12
(10,8)	1.0	10.69	2.00	94.47	95.03	97.82
(11,8)	1.1	8.80	3.00	87.24	95.62	97.43
(12,8)	1.2	6.29	5.00	67.25	93.17	97.04
(12,7)	1.2	4.53	7.00	40.48	91.31	95.88

teristics degrade, one can no longer guarantee the confidence in the decoded feature. The weighted Viterbi recognizer (WVR), presented in [5], modifies the Viterbi algorithm (VA) to take into account the confidence in the decoded feature. The time-varying reliability γ_t is inserted in the VA by raising the probability $b_j(\boldsymbol{o}_t)$ to the power γ_t to obtain the following state metrics update equation:

$$\phi_{j,t} = \max_i [\phi_{i,t-1} a_{ij}][b_j(\boldsymbol{o}_t)]^{\gamma_t}. \tag{8}$$

Such weighting, also used in [30] for state duration modeling, has the advantage of becoming a simple multiplication of $\log(b_j(\boldsymbol{o}_t))$ by γ_t in the logarithmic domain often used for scaling purposes. Furthermore, note that if one is certain about the received feature, $\gamma_t = 1$ and (8) is equivalent to (1). On the other hand, if the decoded feature is unreliable, $\gamma_t = 0$ and the probability of observing the feature given the HMM state model $b_j(\boldsymbol{o}_t)$ is discarded in the VA recursive step.

Under the hypothesis of a diagonal covariance matrix $\boldsymbol{\Sigma}$, the overall probability $b_j(\boldsymbol{o}_t)$ can be computed as the product of the probabilities of observing each individual feature. The weighted recursive formula (8) can include individual weighting factors $\gamma_{k,t}$ for each of the N_F front-end features

$$\phi_{j,t} = \max_i [\phi_{i,t-1} a_{ij}] \prod_{k=1}^{N_F} [b_j(o_{k,t})]^{\gamma_{k,t}}. \tag{9}$$

VI. ALLEVIATING THE EFFECT OF ERASURES

In this section, techniques designed for coping with channel erasures are presented, regardless of whether the erasures are the result of a detected channel error or an actual channel erasure.

One method used to reduce the effect of channel transmission on recognition accuracy consists of dropping the unreliable features from the sequence of observations (e.g., [19]). The motivation is that channel errors rapidly degrade recognition accuracy, while recognizers can cope with missing segments in the sequence of observations given the redundancy of the speech signal. The drawback is that the timing information associated with them is lost. When missing frames are removed from the trellis, no state transitions are possible, and the received features might be analyzed using an inappropriate HMM state. This problem becomes more significant when erasures occur in bursts, forcing the trellis search in the same state for a long period of time, which can significantly impact recognition accuracy.

Another method is frame erasure concealment, which replaces the missing frame with an estimate, and preserves the timing information. Repetition-based concealment replaces missing frames with copies of previously-received frames, while interpolation-based concealment uses some form of pattern matching and interpolation from the neighboring frames to derive a replacement frame (e.g., [7]–[9]). Both techniques are justified by the high correlation between consecutive frames. Interpolation techniques require reception of the next valid feature vector, which may add significant delay when bursts of erasures occur.

We present and compare in the following two sections extensions to the frame dropping and repetition-based concealment techniques, whereby the confidence in the channel decoding operation or the frame erasure concealment technique is fed into the Viterbi recognizer for improved recognition performance.

A. λ-WVR Based on Channel Decoding Reliability

We introduced the WVR technique in [5] to match the recognizer with the confidence in the decoded feature after channel transmission. We present here a channel decoding reliability measurement based on the proposed λ-soft decision decoding scheme presented in Section IV. We consider both binary and continuous WVR weighting.

With *binary* weighting, the weighting coefficients γ_t can either be 0 (if the frame is lost or an error is detected) or 1 (if the frame is received). The advantage of this technique over frame dropping, where state metrics are not updated ($\phi_{j,t} = \phi_{j,t-1}$), is that the timing information of the observation sequence is conserved. State metrics are continuously updated, even when $\gamma_t = 0$, by virtue of the state transition probability matrix using $\phi_{j,t} = \max_i[\phi_{i,t-1} a_{ij}]$.

The system can be refined if a time-varying *continuous* estimate γ_t of the feature vector reliability is used. We propose the function $\gamma_t = \lambda_t^2$ to map the interval $[0, 1]$ for λ_t to the interval $[0, 1]$ for γ_t. The quadratic exponent is empirically chosen after it was shown to provide necessary statistical rejection of the uncertain frames.

Note that if hard decision decoding was employed, only binary weighting could be used. For soft decision decoding, on the other hand, both binary weighting with $\gamma_t = 0$ if $\lambda_t < \tau$ and $\gamma_t = 1$ if $\lambda_t \geq \tau$, and continuous weighting with $\gamma_t = \lambda_t^2$ can be used.

B. ρ-WVR Based on Erasure Concealment Quality

Performance of repetition techniques degrades rapidly as the number of consecutive lost frames increases. When frame losses exceed the length of a phoneme (20–100 ms or 2–10 frames), the speech signal has evolved to another sound, which no longer justifies repetition of the last correctly received feature vector.

TABLE VI
DETERMINATION OF THE WEIGHTING COEFFICIENTS FOR CONCEALMENT BASED WEIGHTED VITERBI RECOGNITION

GILBERT STATE	GOOD	BAD
Static features	$\gamma_{k,t} = \sqrt{\rho_k(t-tc)}$	
Dynamic features	$\gamma_{k,t} = 1$	$\gamma_{k,t} = 0$

In this case, it is beneficial to decrease the weighting factor $\gamma_{k,t}$ when the number of consecutively repeated frames increases. For the weighting coefficients, we propose

$$\gamma_{k,t} = \sqrt{\rho_k(t-t_c)} \tag{10}$$

where ρ_k is the time auto-correlation of the kth feature and t_c is the time instant of the last correctly received frame. Note that if there is no erasure, then $t = t_c$ and $\gamma_{k,t} = 1$.

For the case of feature vectors consisting of temporal and dynamic features (derivative and acceleration), the weighting coefficients $\gamma_{k,t}$ are computed as follows. First, the receiver determines the status of the channel. If two consecutive frames are lost/received, then it determines that the channel is bad/good. In the bad channel state, temporal features are repeated and the weighting coefficients of the dynamic features are set to zero. If the channel state is good, the dynamic features are computed and the weighting coefficients of the dynamic features are set to one. A one-sided derivative is used if one neighboring frame, on either side, is lost while still in a good channel state. This option is chosen over repeating the entire previous frame (temporal *and* dynamic features) since time-correlation of the dynamic features is significantly smaller than for the temporal features. Table VI recapitulates the weighting coefficients for ρ-WVR.

C. Comparison of the Different Techniques

Table VII(a) illustrates recognition accuracy for the different frame erasure concealment techniques applied to the independent erasure channel. Baseline recognition accuracy for erasure-free channels is 98.52%. Several observations are made. 1) After about 10–20% of independent frame erasures, recognition accuracy degrades rapidly. 2) Transmission of the binary frame erasure reliability measurement to the weighted Viterbi recognizer preserves synchronization of the VA and significantly reduces the word error rate. 3) Repetition-based frame erasure concealment, which in addition to preserving the timing also provides an approximation for the missing frame, typically outperforms binary λ-WVR. 4) Addition of the weighting coefficients $\gamma_{k,t}$ representing the quality of the feature concealment technique (10) in the Viterbi search further improves recognition performance.

These results are confirmed in Table VII(b) for the bursty Gilbert channels of Table I, for which we can make additional observations: 1) Binary WVR may outperform repetition-based erasure concealment when the average burst lengths are large. 2) Again, frame erasure concealment combined with WVR provides the best recognition results. For instance, for the Gilbert channel with $(P_{GB}, P_{BG}) = (1.25, 5)$, recognition accuracy

TABLE VII
RECOGNITION ACCURACY OVER INDEPENDENT AND BURSTY ERASURE CHANNELS USING FRAME DROPPING, FRAME DROPPING WITH BINARY λ-WVR, REPETITION ERASURE CONCEALMENT WITH AND WITHOUT CONTINUOUS ρ-WVR

Pct. Erasures	0%	10%	20%	30%	40%	50%
Frame dropping	98.52	97.19	93.51	85.49	71.23	56.33
Binary λ-WVR	98.52	98.31	98.11	97.19	96.87	94.31
Concealment	98.52	98.47	98.31	98.19	97.67	96.35
Conc. + ρ-WVR	98.52	98.52	98.47	98.39	98.11	97.61

(b) Bursty (Gilbert-Elliot) erasure channels.

Channels	2.5/20	2.5/15	5/20	2.5/10	1.25/5	5/15	10/20
Frame dropping	91.31	87.35	86.21	82.17	80.47	79.81	75.11
Binary λ-WVR	97.44	96.41	96.29	94.61	93.81	95.13	94.88
Concealment	97.50	96.62	96.82	94.50	93.38	94.42	93.91
Conc.+ρ-WVR	98.13	97.64	97.89	97.46	97.18	96.94	96.94

improves from 93.27% to 97.03%, a 71% relative word error rate (WER) reduction compared to the baseline recognition performance of 98.52%. 3) Despite average overall probability of frame erasures between 9% and 27% and average length of erasure bursts between 4 and 19 frames (see Table I), recognition accuracy approaches baseline performance.

Note that Table VII does not include results for continuous λ-WVR, which require simulations of a complete remote recognition system, including channel coding and decoding. We compare in Section VII the performance of continuous ρ-WVR and continuous λ-WVR on a complete DSR system.

VII. PERFORMANCE OF COMPLETE DSR SYSTEMS

In this section, the concepts presented above (channel coding, channel decoding and speech recognition) with their respective innovations (error detection over error correction, soft-decision based error detection and weighted Viterbi recognition) are applied to complete DSR systems. Two ASR features are analyzed, PLP and MFCC.

A. Complete DSR System for PLP Features

Table VIII presents recognition accuracy of a complete DSR system over a wide range of independent Rayleigh fading channels. Source coding is applied to the LSFs of the PLP system, with 5 to 7 bits per 20 ms frame, using the technique proposed in [19], which includes predictive coding and interpolation. Depending on the channel conditions, different linear block codes maximizing error detection are used [19] and λ-soft channel decoding is performed. The overall bit rate, including source and channel coding, is limited to 500 bps.

Two scenarios are considered. In the first scenario (λ-WVR), all the features are transmitted to the recognizer, even the unreliable ones, and the weighting coefficients ($\gamma_t = \lambda_t^2$) will lower

TABLE VIII
RECOGNITION PERFORMANCE OF CHANNEL BASED CONTINUOUS λ-WVR ($\gamma_t = \lambda_t^2$) AND CONCEALMENT BASED CONTINUOUS ρ-WVR ($\gamma_{k,t} = \sqrt{\rho_k(t - t_c)}$) OVER RAYLEIGH FADING CHANNELS USING THE AURORA-2 DATABASE AND PLP FEATURES

Block Code (N,K)	Bit Rate (bps)	SNR (dB)	BER (%)	RECOGNITION (%) Cont. λ-WVR $\gamma_t = \lambda_t^2$	ρ-WVR $\gamma_{k,t} = \sqrt{\rho_k}$
(8,7)	400	9	2.88	98.5	98.5
(8,6)	400	8	3.55	98.3	98.2
(8,6)	400	7	4.35	98.3	98.3
(10,7)	500	6	5.30	98.4	98.5
(10,7)	500	5	6.42	98.2	98.3
(10,6)	500	4	7.71	98.1	98.1
(10,6)	500	3	9.19	97.6	97.7
(10,5)	500	2	10.85	97.4	97.6

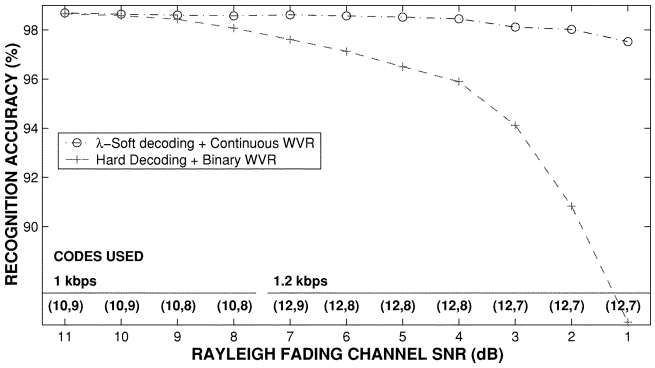


Fig. 6. Recognition accuracy after transmission of the 13 MFCCs over an independent Rayleigh fading channel.

the importance of the inaccurate ones. In the second scenario (ρ-WVR), the unreliable features (those for which $\lambda_t \leq 0.16$) are dropped and concealed with a substitution feature vector. WVR weighting coefficients are based on the quality of the concealment operation ($\gamma_{k,t} = \sqrt{\rho_k(t - t_c)}$). Table VIII indicates that in this case, no strategy consistently outperforms.

B. Complete DSR System for MFCC Features

Parts of the experiments presented above for PLP are repeated in this section for MFCC features, illustrating the generality of the source coding, channel coding and channel decoding scheme presented in the previous sections.

MFCCs are quantized using the techniques presented in [5] (first order predictive weighted VQ with two splits) with 7 to 9 bits per split and interpolation by a factor of 2 at the receiver. After channel protection, the number of bits after forward error correction is 10 or 12 bits per split, for a total of 1.0 or 1.2 kbps, depending on channel conditions.

Fig. 6 illustrates recognition accuracy after choosing for each SNR the block code that yields the best results. The superior performance of the joint soft decision decoding-Viterbi recognition scheme is confirmed for MFCC features. Recognition accuracies remain acceptable over a wide range of independent Rayleigh fading channel SNRs and using overall bit rates less than 1.2 kbps.

VIII. SUMMARY AND CONCLUSIONS

In this paper, we present a framework for developing source coding, channel coding and decoding as well as erasure concealment techniques adapted for DSR applications.

First, it is shown that speech recognition, as opposed to speech coding, is more sensitive to channel errors than channel erasures and appropriate channel coding design criteria are determined.

Efficient linear block codes for error detection are presented and a new technique for performing error detection with soft decision decoding is described. The new channel decoder, which introduces additional complexity only at the server, is proven to outperform the widely-used hard decision decoding scheme for error detection.

Once an error is detected, the corresponding frame is erased and frame erasure concealment techniques which alleviate the effect of channel transmission are discussed. We introduce the weighted Viterbi recognizer (WVR) whereby the recognizer is modified to include a time-varying weighting factor depending on the quality of each feature after transmission over time-varying channels.

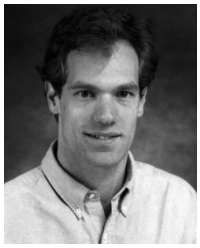
As a case study, source coding, channel coding, and speech recognition techniques are combined to provide high recognition accuracy over a large range of channel conditions for PLP based coefficients. Line spectral pairs representing the PLP spectrum are quantized using weighted vector quantization operating at 1 kbps or less. We demonstrate that high recognition accuracy over a wide range of channel conditions is possible with less than 1.2 kbps overall bitrate when using the appropriate source and channel coder, alleviation of the effect of channel erasures and the weighted Viterbi recognition engine. Similar results were also obtained for MFCCs, illustrating the generality of the proposed framework. In fact, the source and channel coding techniques presented are not restricted to the transmission of PLP based coefficients and MFCCs, and can be extended to other types of ASR feature.

Future work will include examining the effects of model size (word, phoneme, tri-phone), as well as studying the robustness of the source encoders and recognition scheme to acoustic noise.

REFERENCES

[1] T. Salonidis and V. Digalakis, "Robust speech recognition for multiple topological scenarios of the GSM mobile phone system," in *Proc. ICASSP*, May 1998, pp. 101–104.
[2] 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 *Proc. ICASSP*, vol. 1, May 1996, pp. 77–80.
[3] L. Karray, A. Jelloun, and C. Mokbel, "Solutions for robust recognition over the GSM cellular network," in *Proc. ICASSP*, vol. 1, 1998, pp. 166–170.
[4] A. Gallardo, F. Diaz, and F. Vavlerde, "Avoiding distortions due to speech coding and transmission errors in GSM ASR tasks," in *Proc. ICASSP*, May 1999, pp. 277–280.

[5] A. Bernard and A. Alwan, "Joint channel decoding—Viterbi recognition for wireless applications," in *Proc. Eurospeech*, vol. 4, Sept. 2001, pp. 2703–2706.
[6] A. Potamianos and V. Weerackody, "Soft-feature decoding for speech recognition over wireless channels," in *Proc. ICASSP*, vol. 1, May 2001, pp. 269–272.
[7] B. Milner and S. Semnani, "Robust speech recognition over IP networks," in *Proc. ICASSP*, vol. 3, June 2000, pp. 1791–1794.
[8] B. Milner, "Robust speech recognition in burst-like packet loss," in *Proc. ICASSP*, vol. 1, May 2001, pp. 261–264.
[9] C. Perkins, O. Hodson, and V. Hardman, "A survey of packet loss recovery techniques for streaming audio," *IEEE Network*, vol. 12, pp. 40–48, Oct. 1998.
[10] S. Euler and J. Zinke, "The influence of speech coding algorithms on automatic speech recognition," in *Proc. ICASSP*, 1994, pp. 621–624.
[11] B. T. Lilly and K. K. Paliwal, "Effect of speech coders on speech recognition performance," in *Proc. ICSLP*, vol. 4, Oct. 1996, pp. 2344–2347.
[12] L. Yapp and G. Zick, "Speech recognition on MPEG/audio encoded files," in *IEEE Int. Conf. Multimedia Comput. Syst.*, June 1997, pp. 624–625.
[13] J. Huerta and R. Stern, "Speech recognition from GSM parameters," in *Proc. ICSLP*, vol. 4, 1998, pp. 1463–1466.
[14] S. H. Choi, H. K. Kim, H. S. Lee, and R. M. Gray, "Speech recognition method using quantised LSP parameters in CELP-type coders," *Electron. Lett.*, vol. 34, no. 2, pp. 156–157, Jan. 1998.
[15] H. K. Kim and R. V. Cox, "Feature enhancement for a bitstream-based front-end in wireless speech recognition," in *Proc. ICASSP*, vol. 1, May 2001, pp. 241–243.
[16] H. K. Kim and R. Cox, "Bitstream-based feature extraction for wireless speech recognition," in *Proc. ICASSP*, vol. 1, 2000, pp. 21–24.
[17] V. Digalakis, L. Neumeyer, and M. Perakakis, "Quantization of cepstral parameters for speech recognition over the World Wide Web," *IEEE J. Select. Areas Commun.*, vol. 17, pp. 82–90, Jan. 1999.
[18] D. Pearce, "Enabling new speech driven services for mobile devices: An overview of the ETSI standards activities for distributed speech recognition front-ends," in *Proc. Applied Voice Input/Output Soc. Conf.*, May 2000.
[19] A. Bernard and A. Alwan, "Source and channel coding for remote speech recognition over error-prone channels," in *Proc. ICASSP*, vol. 4, May 2001, pp. 2613–2616.
[20] G. Ramaswamy and P. Gopalakrishnan, "Compression of acoustic features for speech recognition in network environments," in *Proc. ICASSP*, vol. 2, May 1998, pp. 977–980.
[21] E. N. Gilbert, "Capacity of burst noise channel," *Bell Syst. Tech. J.*, vol. 39, pp. 1253–1265, Sept. 1960.
[22] M. Yajnik, S. Moon, J. Kurose, and D. Towsley, "Measurement and modeling of the temporal dependence in packer loss," in *INFOCOM*, vol. 1, Mar. 1999, pp. 345–352.
[23] C. Boulis, M. Osterndorf, E. Riskin, and S. Otterson, "Graceful degradation of speech recognition performance over packet-erasure network,", Sept. 2002.
[24] J. S. Swarts and H. C. Ferreira, "On the evaluation and application of Markov channel models in wireless communications," in *Proc. Vehicular Technology Conf.*, vol. 1, Sept. 1999, pp. 117–121.
[25] T. Tao, J. Lu, and J. Chuang, "Hierarchical Markov model for burst error analysis in wireless communications," in *Vehicular Technology Conf.*, May 2001, pp. 2843–2847.
[26] A. Konrad, B. Zhao, D. Joseph, and R. Ludwig, "A Markov-based channel model algorithm for wireless networks," in *ACM Work. on Model., Anal. and Simul. of Wireless and Mobile Syst.*, July 2001, pp. 28–36.
[27] J. Proakis, *Digital Communications*. New York: McGraw-Hill, 1995.
[28] H. Hermansky, "Perceptual linear predictive (PLP) analysis of speech," *J. Acoust. Soc. Amer.*, vol. 87, no. 4, pp. 1738–1752, Apr. 1990.
[29] A. Bernard, "Source and channel coding for speech transmission and remote speech recognition," Ph.D. dissertation, Univ. California, Los Angeles, 2002.
[30] N. B. Yoma, F. R. McInnes, and M. A. Jack, "Weighted Viterbi algorithm and state duration modeling for speech recognition in noise," in *Proc. ICASSP*, vol. 2, May 1998, pp. 709–712.



Alexis Bernard (S'97) was born in Brussels, Belgium, in 1973. He received the B.S. degree in electrical engineering from the Université Catholique de Louvain (UCL), Louvain-La-Neuve, Belgium, in 1996 and the M.S. and Ph.D. degrees in electrical engineering from the University of California, Los Angeles (UCLA), in 1998 and 2002, respectively.

In 1994–1995, he was an Exchange Scholar at the Katholieke Universiteit Leuven (KUL), Belgium. He has worked as an Intern for several companies, including Alcatel Telecom in Antwerp, Belgium (1997) and Texas Instruments in Dallas, TX (1999 and 2000). He is now a Member of Technical Staff at the DSP Solutions R&D Center of Texas Instruments. His current research interests include source and channel coding, with focus on speech transmission and distributed speech recognition as well as information theory, communications and noise robust speech recognition.

Dr. Bernard is a recipient of the Belgian American Educational Foundation fellowship.

Abeer Alwan (S'82–M'85–SM'00) received the Ph.D. degree in electrical engineering from the Massachusetts Institute of Technology (MIT), Cambridge, in 1992.

Since then, she has been with the Electrical Engineering Department at UCLA as an Assistant Professor (1992–1996), Associate Professor (1996–2000), and Professor (2000–present). She established and directs the Speech Processing and Auditory Perception Laboratory at UCLA. Her research interests include modeling human speech production and perception mechanisms and applying these models to speech-processing applications such as automatic recognition, compression, and synthesis. She is an editor-in-chief of *Speech Communication*

Dr. Alwan is an elected member of Eta Kappa Nu, Sigma Xi, Tau Beta Pi, the New York Academy of Sciences, and the IEEE Signal Processing Technical Committees on Audio and Electroacoustics and on Speech Processing. She served as an elected member on the Acoustical Society of America Technical Committee on Speech Communication from 1993 to 1999. She is the recipient of the NSF Research Initiation Award (1993), the NIH FIRST Career Development Award (1994), the UCLA-TRW Excellence in Teaching Award (1994), the NSF Career Development Award (1995), and the Okawa Foundation Award in Telecommunications (1997).