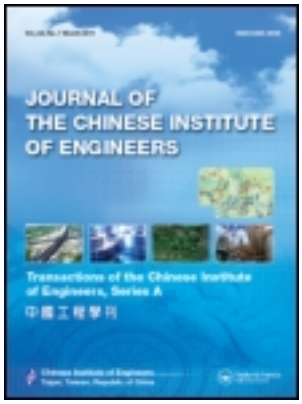


Journal of the Chinese Institute of Engineers
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tcie20

Channel-optimized error mitigation for distributed speech recognition over wireless networks

Cheng-Lung Lee [a] & Wen-Whei Chang [b]

[a] Department of Communications Engineering , National Chiao-Tung University , 1001 University Road, Hsinchu 300, Taiwan, R.O.C.

[b] Department of Communications Engineering , National Chiao-Tung University , 1001 University Road, Hsinchu 300, Taiwan, R.O.C. Phone: 886-3-5731826 E-mail:

Published online: 04 Mar 2011.

To cite this article: Cheng-Lung Lee & Wen-Whei Chang (2009) Channel-optimized error mitigation for distributed speech recognition over wireless networks, Journal of the Chinese Institute of Engineers, 32:1, 45-51, DOI: 10.1080/02533839.2009.9671481

To link to this article: http://dx.doi.org/10.1080/02533839.2009.9671481

CHANNEL-OPTIMIZED ERROR MITIGATION FOR DISTRIBUTED SPEECH RECOGNITION OVER WIRELESS NETWORKS

Cheng-Lung Lee and Wen-Whei Chang*

ABSTRACT

This paper investigates the error mitigation algorithms for distributed speech recognition over wireless channels. A MAP symbol decoding algorithm which exploits the combined a priori information of source and channel is proposed. This is used in conjunction with a modified BCJR algorithm for decoding convolutional codes based on sectionalized code trellises. Performance is further enhanced by the use of the Gilbert channel model that more closely characterizes the statistical dependencies between channel bit errors. Experiments on Mandarin digit string recognition task indicate that our proposed mitigation scheme achieves high robustness against channel errors.

Key Words: channel error mitigation, distributed speech recognition.

I. INTRODUCTION

The increasing use of mobile and IP networks for speech communication has lead to distributed speech recognition (DSR) systems being developed (ETSI ES 202 212 v1.1.1., 2003). The basic idea of DSR consists of using a local front-end from which speech features are extracted and transmitted through a data channel to a remote back-end recognizer. F or transmission, speech features are grouped into pairs and compressed via vector quantizers (VQs) in order to meet bandwidth requirements. The VQ encoder operates by mapping a large set of input vectors into a finite set of representative codevectors. The transmitter sends the index of the nearest codevector to the receiver, while the receiver decodes the codevector associated with the received index and uses it as an approximation of the input vector. Transmitting VQ data over noisy channels changes the encoded information and consequently leads to degraded recognition performance. In the case of packet-erasure channels, several packet loss compensation techniques such as interpolation (Bernard and Alwan, 2002) and error control coding (Boulis *et al*., 2002) have been introduced for DSR. For wireless channels, joint source-channel decoding (JSCD) techniques (Peinado *et al*., 2003; Reinhold and Valentin, 2004; Fingscheidt and Vary, 2001) have been shown effective for error mitigation using source residual redundancy assisted by bit reliability information provided by the soft-output channel decoder. However, the usefulness of these techniques may be restricted because they only exploit the bit-level source correlation on the basis of a memoryless AWGN channel assumption.

In this paper, we attempt to capitalize more fully on the a priori knowledge of source and channel and then develop a DSR system with increased robustness against channel errors. The first step toward realization is to use quantizer indexes rather than single index-bits as the bases for the JSCD, since the dependencies of quantizer indexes are stronger than the correlations of the index-bits. The next knowledge source to be exploited is the channel error characteristics. Transmission errors encountered in most real communication channels exhibit various degrees of statistical dependency that are contingent on the transmission medium and on the particular modulation technique used. A typical example occurs in digital mobile radio channels, where speech parameters suffer severe degradation from error bursts due to the combined effects of fading and multipath propagation. A standard technique for robust VQ over a channel with memory is to use interleaving to render the channel memoryless and then design a decoding algorithm for the memoryless channel. This approach, however, often introduces large

*Corresponding author. (Tel: 886-3-5731826; Email: wwchang @cc.nctu.edu.tw)

The authors are with the Department of Communications Engineering, National Chiao-Tung University, 1001 University Road, Hsinchu 300, Taiwan, R.O.C.

Table 1 Entropies for DSR feature pairs

Parameter (u_t)	C_1, C_2	C_3, C_4	C_5, C_6	C_7, C_8	C_9, C_{10}	C_{11}, C_{12}	C_0, log$E$
Bits/Codeword	6	6	6	6	6	5	8
$H(u_t)$	5.75	5.71	5.68	5.80	5.82	4.85	7.33
$H(u_t\|u_{t-1})$	3.17	3.42	3.85	4.14	4.25	3.64	3.46

decoding delays and does not utilize the channel memory information. Further improvement can be realized through a more precise characterization of the channel on which the decoder design is based (Kanal and Sastry, 1978). For this investigation, we focused on the two-state Markov chain model proposed by Gilbert (Gilbert, 1960). This model has several practical advantages over the Gaussian channel (Peinado *et al*., 2003) and binary Markov channels (Wang and Moayeri, 1993). First, the Gilbert model is relatively simple and can characterize a wide range of digital channels, as evidenced by its applicability to performance analysis of various error control schemes (Drukarev and Yiu, 1986). Second, as we shall see later, the channel transition probabilities of the Gilbert model have a recursive formula that can be represented in terms of model parameters.

II. DSR TRANSMISSION SYSTEM

The standard, ETSI ES 202 212, describes the speech processing, transmission, and quality aspects of a DSR system. The local front-end consists of a feature extraction algorithm and an encoding scheme for speech input to be transmitted to a remote recognizer. Each speech frame is represented by a 14-dimension feature vector containing log-energy *logE* and 13 Mel-frequency cepstral coefficients (MFCCs) ranging from C_0 to C_{12}. For the cepstral analysis speech signals are sampled at 8 kHz and analyzed using a 25 ms Hamming window with 10 ms frame shift. These features are further compressed based on a split vector codebook where the set of 14 features is split into 7 subsets with two features in each. Each feature pair is quantized using its own codebook. MFCCs C_1 to C_{10} are quantized with 6 bits each pair, (C_{11}, C_{12}) is quantized with 5 bits, and (C_0, *logE*) is quantized with 8 bits. Two quantized frames are grouped together and protected by a 4-bit cyclic redundancy check creating a 92-bit frame-pair packet. Twelve of these frame-pairs are combined and appended with overhead bits resulting in an 1152-bit multiframe packet representing 240 ms of speech. Multiframe packets are concatenated into a bit-stream for transmission via a data channel with an overall data rate of 4800 bits/s.

This work is devoted to channel error mitigation for DSR over burst error channels. Fig. 1 gives the block diagram of the transmission scheme for each

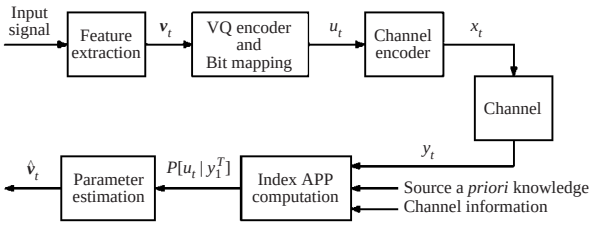


Fig. 1 Transmission scheme for each DSR feature pair

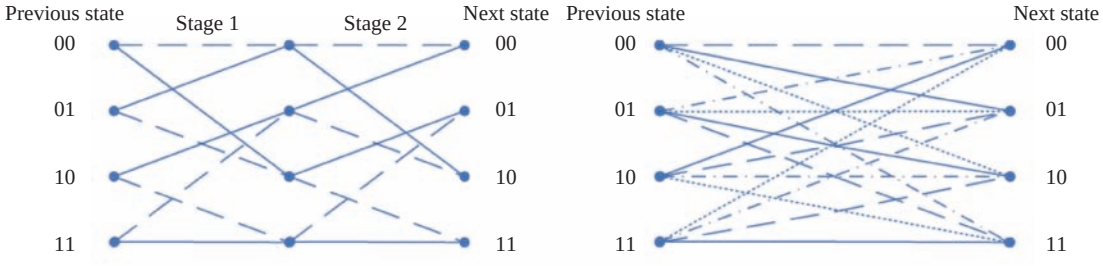
DSR feature pair. Suppose at time t, the input vector v_t is quantized to obtain a codevector $c_t \in \{c^{(i)}, i = 0, 1, \cdots, 2^k - 1\}$ that, after bit mapping, is represented by a k-bit combination $u_t = (u_t(1), u_t(2), \cdots, u_t(k))$. Each bit combination u_t is assigned to a quantizer index $i \in \{0,1, \cdots, 2^k - 1\}$ and we write for simplicity $u_t = u_t^i$ to denote that u_t represents the i-th quantizer index. Due to constraints on coding complexity and delay, the VQ encoder exhibits considerable redundancy within the encoded index sequence, either in terms of a non-uniform distribution or in terms of correlation. If only the non-uniform distribution is considered and the indexes are assumed to be independent of each other, the redundancy is defined as the difference between the index length k and the entropy given by

$$H(u_t) = -\sum_{u_t} P(u_t) \cdot \log_2 P(u_t) . \tag{1}$$

If inter-frame correlation of indexes is considered by using a first-order Markov model with transition probabilities $P(u_t|u_{t-1})$, the redundancy is then defined as the index length k and the conditional entropy given by

$$H(u_t|u_{t-1}) = -\sum_{u_t}\sum_{u_{t-1}} P(u_t, u_{t-1}) \cdot \log_2 P(u_t|u_{t-1}) . \tag{2}$$

Table 1 shows the index lengths and entropies for the seven feature pairs of the ETSI DSR frond-end. For each column in Table 1, the probabilities $P(u_t)$ and $P(u_t|u_{t-1})$ have to be estimated in advance from a training speech database. From it we see that the DSR index sequence is better characterized by a first-order Markov process. For error protection individual index-bits are fed into a binary convolutional encoder consisting of M shift registers. The register shifts



(a) Bit-level trellis diagram (b) Merged trellis diagram

Fig. 2 Trellis diagrams used for (a) the encoder and (b) the MAP decoder

one bit at a time and its state is determined by the M most recent inputs. After channel encoding, the code-bit combination corresponding to the quantizer index u_t is denoted by $x_t = (x_t(1), x_t(2), \cdots, x_t(n))$ with the code rate $R = k/n$.

One of the principal concerns in transmitting VQ data over noisy channels is that channel errors corrupt the bits that convey information about quantizer indexes. Assume that a channel's input x_t and output y_t differ by an error pattern e_t, so that the received bit combination is $y_t = (y_t(1), y_t(2), \cdots, y_t(n))$ in which $y_t(l) = x_t(l) \oplus e_t(l)$, $l = 1, 2, \cdots, n$, and $\oplus$ denotes the bitwise modulo-2 addition. At the receiver side, the JSCD decoder will find the most probable transmitted quantizer index given the received sequence. The decoding process starts with the formation of an *a posteriori* probability (APP) for each of the possibly transmitted indices $u_t = i$, which is followed by choosing the index value $\hat{i}$ that corresponds to the maximum *a posteriori* (MAP) probability for that quantizer index. Once the MAP estimate of the quantizer index is determined, its corresponding codevector becomes the decoded output $\hat{\boldsymbol{v}}_t = \boldsymbol{c}^{(\hat{i})}$. The APP is the probability that a decoded index $u_t = i$ can be derived from the joint probability $P(u_t^i, s_t, y_1^T)$, where s_t is the channel encoder state at time t and $y_1^T = (y_1, y_2, \cdots, y_T)$ is the received sequence from time $t = 1$ through some time T. We have chosen the length $T = 24$ in compliance with the ETSI bit-streaming format, where each multiframe message packages speech features from 24 frames. Proceeding in this way, the symbol APP can be obtained by summing the joint probability over all encoder states, as follows:

$$P(u_t = i \,|\, y_1^T) = \sum_{s_t} \frac{P(u_t^i, s_t, y_1^T)}{P(y_1^T)}, \; i = 0, 1, \cdots, 2^k - 1. \tag{3}$$

III. MODIFIED BCJR ALGORITHM

Depending upon the choice of the symbol APP calculator, a number of different MAP decoder implementations can be realized. For the transmission scheme with channel coding, a soft-output channel decoder can be used to provide both decoded bits and their reliability information for further processing to improve the system error performance. The most well-known soft-output decoding algorithm is the BCJR algorithm (Bahl *et al*., 1974) that was devised to minimize the bit error probability. This algorithm is a trellis-based decoding algorithm for both linear block and convolutional codes. The derivation presented in Bahl *et al*. led to a forward-backward recursive computation on the basis of a bit-level code trellis. In a bit-level trellis diagram, there are two branches leaving each state and every branch represents a single index-bit. Proper sectionalization of a bit-level trellis may result in useful trellis structural properties (Lin and Costello, 2004) and allow us to devise MAP decoding algorithms which exploit bit-level as well as symbol-level source correlations. To advance with this, we propose a modified BCJR algorithm which parses the received code-bit sequence into blocks of length n and computes the APP for each quantizer index on a symbol-by-symbol basis. Unlike a conventional BCJR algorithm that decodes one bit at a time, our scheme proceeds with decoding the quantizer indexes in a frame as nonbinary symbols according to their index length k. By parsing the code-bit sequence into n-bit blocks, we are in essence merging k stages of the original bit-level code trellis into one. As an example, we illustrate in Fig. 2 two stages of the bit-level trellis diagram of a rate 1/2 convolutional encoder with generator polynomial $(5, 7)_8$. The solid lines and dashed lines correspond to the input bits of 0 and 1, respectively. Fig. 2 also shows the decoding trellis diagram when two stages of the original bit-level trellis are merged together. In general, there are 2^k branches leaving and entering each state in a k-stage merged trellis diagram. Having defined the decoding trellis diagram as such, there will be one symbol APP corresponding to each branch which represents a particular quantizer index $u_t = i$. For convenience, we say that the sectionalized trellis diagram forms a finite-state machine defined by its state transition function $S(u_t^i, s_t)$ and output function $X(u_t^i,$

s_t). Viewed from this perspective, the code-bit combination $x_t = X(u_t^i, s_t)$ is associated with the branch from state s_t to state $s_{t+1} = S(u_t^i, s_t)$ if the corresponding quantizer index at time t is $u_t = i$.

We next modified the BCJR algorithm based on sectionalized trellis to exploit the combined a priori information of source and channel. We begin our development of the modified BCJR algorithm by rewriting the joint probability in Eq. (3) as follows:

$$P(u_t^i, s_t, y_1^T) = \alpha_t^i(s_t)\beta_t^i(s_t), \tag{4}$$

where $\alpha_t^i(s_t) = P(u_t^i, s_t, y_1^t)$ and $\beta_t^i(s_t) = P(y_{t+1}^T | u_t^i, s_t, y_1^t)$. For the MAP symbol decoding algorithm, the forward and backward recursions are to compute the following metrics:

$$\begin{aligned}\alpha_t^i(s_t) &= \sum_{s_{t-1}} \sum_j P(u_t^i, s_t, u_{t-1}^j, s_{t-1}, y_t, y_1^{t-1}) \\ &= \sum_{s_{t-1}} \sum_j \alpha_{t-1}^j(s_{t-1})\gamma_{i,j}(y_t, s_t, s_{t-1}),\end{aligned} \tag{5}$$

$$\begin{aligned}\beta_t^i(s_t) &= \sum_{s_{t+1}} \sum_j P(u_{t+1}^j, s_{t+1}, y_{t+1}, y_{t+2}^T | u_t^i, s_t, y_1^t) \\ &= \sum_{s_{t+1}} \sum_j \beta_{t+1}^j(s_{t+1})\gamma_{j,i}(y_{t+1}, s_{t+1}, s_t),\end{aligned} \tag{6}$$

in which

$$\begin{aligned}&\gamma_{i,j}(y_t, s_t, s_{t-1}) \\ &= P(u_t^i, s_t, y_t | u_{t-1}^j, s_{t-1}, y_1^{t-1}) \\ &= P(s_t | u_{t-1}^j, s_{t-1}, y_1^{t-1}) P(u_t^i | s_t, u_{t-1}^j, s_{t-1}, y_1^{t-1}) \\ &\quad \cdot P(y_t | u_t^i, s_t, u_{t-1}^j, s_{t-1}, y_1^{t-1}).\end{aligned} \tag{7}$$

Having a proper representation of the branch metric $\gamma_{i,j}(y_t, s_t, s_{t-1})$ is the critical step in applying MAP symbol decoding to error mitigation and one that conditions all subsequent steps of the implementation. As a practical manner, several additional factors must be considered to take advantage of source correlation and channel memory. First, making use of the sectionalized structure of a decoding trellis, we write the first term in Eq. (7) as

$$\begin{aligned}&P(s_t | u_{t-1}^j, s_{t-1}, y_1^{t-1}) = P(s_t | u_{t-1}^j, s_{t-1}) \\ &= \begin{cases} 1, & s_t = S(u_{t-1}^j, s_{t-1}) \\ 0, & \text{othertwise}. \end{cases}\end{aligned} \tag{8}$$

The next knowledge source to be exploited is the residual redundancy remaining in the DSR features. Assuming that the quantizer index is modelled as a first-order Markov process with transition probabilities $P(u_t | u_{t-1})$, the second term in Eq. (7) is reduced to

$$P(u_t^i | s_t, u_{t-1}^j, s_{t-1}, y_1^{t-1}) = P(u_t = i | u_{t-1} = j). \tag{9}$$

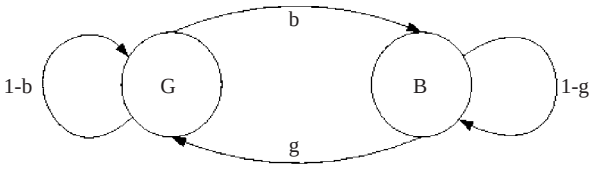


Fig. 3 Gilbert channel model

In addition to source a priori knowledge, specific knowledge about the channel memory must be taken into consideration. There are many models describing the correlation of bit error sequences. If no channel memory information is considered, which means that the channel bit errors are assumed to be random, the third term in Eq. (7) is reduced to

$$\begin{aligned}&P(y_t | u_t^i, s_t, u_{t-1}^j, s_{t-1}, y_1^{t-1}) \\ &= P(y_t | x_t = X(u_t^i, s_t)) \\ &= P(e_t) = \varepsilon^l (1-\varepsilon)^{n-1},\end{aligned} \tag{10}$$

where ε is the channel bit error rate (BER) and l is the number of ones occurring in the error pattern e_t. When intraframe and interframe memory of the channel are considered, the third term in Eq. (7) becomes

$$\begin{aligned}&P(y_t | u_t^i, s_t, u_{t-1}^j, s_{t-1}, y_1^{t-1}) \\ &= P(y_t | x_t = X(u_t^i, s_t), y_{t-1}, x_{t-1} = X(u_{t-1}^j, s_{t-1})) \\ &= P(e_t | e_{t-1}).\end{aligned} \tag{11}$$

IV. PROBABILITY RECURSIONS FOR GILBERT CHANNEL

Designing a robust DSR system requires that parameterized probabilistic models be used to summarize some of the most relevant aspects of error statistics. It is apparent from previous work on channel modelling (Kanal and Sastry, 1978) that we are confronted with contrasting requirements in selecting a good model. A model should be representative enough to describe real channel behavior and yet it should not be analytically complicated. To permit theoretical analysis, we assumed that the encoded bits of DSR features were subjected to the sample error sequences typical of the Gilbert channel. The Gilbert channel model consists of a Markov chain having an error-free state G and a bad state B, in which errors occur with the probability $(1 - h)$. The state transition probabilities are b and g for the G to B and B to G transitions, respectively. The model state-transition diagram is shown in Fig. 3. The effective BER produced by the Gilbert channel is $\varepsilon = (1 - h)b/(g + b)$. Notice that in the particular case of a Gilbert model with parameter values $b = 1$, $g = 0$, $h = 1 - \varepsilon$,

the channel model reduces to a memoryless binary symmetric channel with the BER ε.

The effectiveness of the MAP symbol decoding depends crucially on how well the error characteristics are incorporated into the calculation of channel transition probabilities $P(e_t|e_{t-1})$. Although using channel memory information was previously proposed for MAP symbol decoding (Turin, 2001), the emphasis was placed upon channels with no interframe memory. When only access to the intraframe memory is available, it was shown that the channel transition probabilities of the Gilbert channel have closed-form expressions that can be represented in terms of model parameters $\{h, b, g\}$. Under such conditions, we can proceed with the MAP symbol decoding in a manner similar to the work of (Turin, 2001). Extensions of these results to channels with both intraframe and interframe memory have been found difficult. Recognizing this, we next develop a general treatment of probability recursions for the Gilbert channel. The main result is a recursive implementation of MAP symbol decoder being closer to the optimal for channels with memory. For notational convenience, channel bit error $e_t(l)$ will be denoted as r_m, in which the bit time m is related to the frame time t as $m = n(t-1) + l$, $l = 1, 2, \cdots, n$. Let $q_m \in \{G, B\}$ denote the Gilbert channel state at bit time m. The memory of the Gilbert channel is due to the Markov structure of the state transitions, which lead to a dependence of the current channel state q_m on the previous state q_{m-1}.

To develop a recursive algorithm, it is more convenient to rewrite the channel transition probabilities as

$$P(e_t|e_{t-1}) = \prod_{m=n(t-1)+1}^{nt} P(r_m = 1|r_{m_0}^{m-1})^{r_m} P(r_m = 0|r_{m_0}^{m-1})^{1-r_m}, \quad (12)$$

where $r_{m_0}^{m-1} = (r_{m_0}, r_{m_0+1}, \ldots, r_{m-1})$ represents the bit error sequence starting from bit $m_0 = n(t-2) + 1$. The following is devoted to a way of recursively computing $P(r_m = 1|r_{m_0}^{m-1})$ from $P(r_{m-1} = 1|r_{m_0}^{m-2})$. The Gilbert channel has two properties, $P(q_m|q_{m-1}, r_{m_0}^{m-1}) = P(q_m|q_{m-1})$ and $P(r_m|q_m, r_{m_0}^{m-1}) = P(r_m|q_m)$, which facilitate the probability recursions. By successively applying the Bayes rule and the Markovian property of the channel, we have

$$\begin{aligned} P(r_m = 1|r_{m_0}^{m-1}) &= P(r_m = 1|q_m = B, r_{m_0}^{m-1}) P(q_m = B|r_{m_0}^{m-1}) \\ &= (1-h) P(q_m = B|r_{m_0}^{m-1}), \end{aligned} \quad (13)$$

in which

$$\begin{aligned} &P(q_m = B|r_{m_0}^{m-1}) \\ &= P(q_m = B|q_{m-1} = G, r_{m_0}^{m-1}) P(q_{m-1} = G|r_{m_0}^{m-1}) \\ &\quad + P(q_m = B|q_{m-1} = B, r_{m_0}^{m-1}) P(q_{m-1} = B|r_{m_0}^{m-1}) \\ &= b \frac{P(q_{m-1} = G, r_{m-1}|r_{m_0}^{m-2})}{P(r_{m-1}|r_{m_0}^{m-2})} \\ &\quad + (1-g) \frac{P(q_{m-1} = B, r_{m-1}|r_{m_0}^{m-2})}{P(r_{m-1}|r_{m_0}^{m-2})} \\ &= b + (1-g-b) \frac{P(q_{m-1} = B, r_{m-1}|r_{m_0}^{m-2})}{P(r_{m-1}|r_{m_0}^{m-2})} \\ &= b + (1-g-b) \frac{P(r_{m-1}|q_{m-1} = B)}{P(r_{m-1}|r_{m_0}^{m-2})} \\ &\quad \cdot \frac{P(r_{m-1} = 1|r_{m_0}^{m-2})}{1-h}. \end{aligned} \quad (14)$$

V. EXPERIMENTAL RESULTS

Computer simulations were conducted to evaluate three MAP-based error mitigation schemes for DSR over burst error channels. First a bit-level trellis MAP decoding scheme BMAP is considered that uses the standard BCJR algorithm to decode the index-bits. The decoders SMAP1 and SMAP2 exploit the symbol-level source redundancy by using a modified BCJR algorithm based on a sectionalized trellis structure. The SMAP1 is designed for a memoryless binary symmetric channel, whereas the SMAP2 exploits the channel memory though the Gilbert channel characterization. The channel transition probabilities to be used for the SMAP1 is $P(e_t)$ in Eq. (10), and $P(e_t|e_{t-1})$ in Eq. (11) for the SMAP2. For purposes of comparison, we also investigated an error mitigation scheme (Peinado *et al*., 2003) which applied the concept of softbit speech decoding (SBSD) and achieved good recognition performance for AWGN and burst channels. A preliminary experiment was first performed to evaluate various decoders for reconstruction of the feature pair (C_0, *logE*) encoded with the DSR front-end. A rate $R = 1/2$ convolutional code with memory order $M = 6$ and the octal generator $(46,72)_8$ is chosen as the channel code. Table 2 presents the signal-to-noise ratio (SNR) obtained from transmission of the index-bits over Gilbert channel with BER ranging from 10^{-3} to 10^{-1}. The results of these experiments clearly demonstrate the improved performance achievable using the SMAP1 and SMAP2 in comparison to those of BMAP and SBSD. Furthermore, the improvement has a tendency to increase for noisy channels with higher BER. This indicates that the residual redundancy of quantizer indexes is better exploited at the symbol level to achieve more performance improvement. A comparison of

Table 2 SNR(dB) performance for various decoders on a Gilbert channel

BER	BMAP	SBSD	SMAP1	SMAP2
0.001	26.84	26.88	26.93	27.51
0.0025	26.37	26.51	26.56	27.10
0.0063	25.21	25.83	25.91	26.41
0.0158	22.30	22.71	23.31	25.13
0.0398	17.51	20.67	21.13	24.67
0.1	14.12	16.88	18.52	23.94

Table 3 Estimated Gilbert model parameters for GSM TCH/F4.8 data channels

CIR(dB)	1	4	7	10
g	0.001	0.01	0.02	0.05
b	0.0197	0.0034	0.0022	0.0034
h	0.7528	0.6086	0.7511	0.9403

the SMAP1 and SMAP2 also revealed the importance of matching the real error characteristics to the channel model on which the MAP symbol decoder design is based. The better performance of SMAP2 can be attributed to its ability to compute the symbol APP taking interframe and intraframe memory of the channel into consideration, as opposed to the memoryless channel assumption made in the SMAP1.

We further validate the proposed decoding algorithms for the case where error sequences were generated using a complete GSM simulation. The simulator is based on the CoCentric GSM library (*CoCentric System Studio-Referenec Design Kits*, 2003) with TCH/F4.8 data and channel coding, interleaving, modulation, a channel model, and equalization. The channel model represents a typical case of a rural area with 6 propagation paths and a user speed of 50 km/h. Further, cochannel interference was simulated at various carrier-to-interference ratios (CIR). In using the SMAP1 and SMAP2 schemes, the channel transition probabilities have to be combined with *a priori* knowledge of Gilbert model parameters which can be estimated once in advance using the gradient iterative method (Chouinard *et al.*, 1988). For each simulated error sequence, we first measured the error-gap distribution $P(0^l|1)$ by computing the probability that at least l successive error-free bits will be encountered next on the condition that an error bit has just occurred. The optimal identification of Gilbert model parameters was then formulated as the least square approximation of the measured error-gap distribution by exponential curve fitting. Table 3 gives estimated Gilbert model parameters for the GSM TCH/F4.8 data channels operating at CIR = 1, 4, 7, 10 dB. The next step in the present investigation concerned the performance degradation that may result from using the SMAP2 scheme under channel mismatch conditions. In Table 4, CIR_d refers to the

Table 4 SNR performance of the SMAP2 over the GSM data channel under channel mismatch conditions

	$CIR_a = 1$	$CIR_a = 4$	$CIR_a = 7$	$CIR_a = 10$
$CIR_d = 1$	11.86	16.68	27.02	30.25
$CIR_d = 4$	11.62	16.78	27.19	30.40
$CIR_d = 7$	11.51	16.72	27.24	30.41
$CIR_d = 10$	11.31	16.32	27.01	30.64

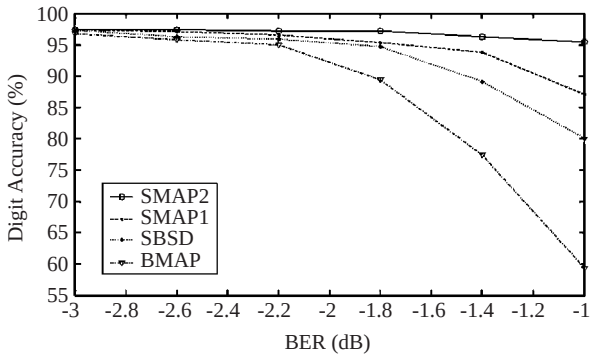


Fig. 4 Recognition performances for DSR transmission over a Gilbert channel

CIR value assumed in the design process, and CIR_a refers to the true CIR used for the evaluation. The best results are in the main diagonal of the table, where channel-matched Gilbert model parameters are used for the channel transition probability computation of Eq. (12). The performance decreases in each column below the main diagonal when the CIR_d is increased. The investigation further showed that the SMAP2 is not very sensitive to a channel mismatch between the design and evaluation assumptions.

We next considered the speaker-independent recognition of Mandarin digit strings as the task without restricting the string length. A Mandarin digit string database recorded by 50 male and 50 female speakers was used in our experiments. Each speaker pronounced 10 utterances and 1-9 digits in each utterance. The speech of 90 speakers (45 male and 45 female) was used as the training data, and the other 10 as test data. The total numbers of digits included in the training and test data were 6796 and 642, respectively. The reference recognizer is based on the HTK software package. A 38-dimension feature vector used in the recognizer consisted of 12 MFCCs, delta-MFCC, delta-delta-MFCC, delta-energy and delta-delta-energy. The digits were modelled as whole word Hidden Markov Models (HMMs) with 8 states per word and 64 mixtures for each state. In addition, a 3-state HMM was used to model pauses before and after the utterance and a one-state HMM was used to model pauses between digits. The DSR results obtained by various error mitigation algorithms for the Gilbert channel are shown in Fig. 4. It

can be seen that employing the source a priori information, sectionalized trellis MAP decoding, and channel memory constantly improves the recognition accuracy. The SMAP2 scheme performs the best in all cases, showing the importance of combining the a priori knowledge of source and channel by means of a sectionalized code trellis and Gilbert channel characterization.

VI. CONCLUSIONS

A JSCD scheme which exploits the combined source and channel statistics as an a priori information is proposed and applied to channel error mitigation in DSR applications. We first investigate the residual redundancies existing in the DSR features and find ways to exploit these redundancies in the MAP symbol decoding process. Also proposed is a modified BCJR algorithm based on sectionalized code trellises which uses Gilbert channel characterization for better decoding in addition to source a priori knowledge. Experiments on Mandarin digit string recognition indicate that the proposed decoder achieved significant improvements in recognition accuracy for DSR over burst error channels.

ACKNOWLEDGEMENT

This study was jointly supported by MediaTek Inc. and the National Science Council, Republic of China, under contract NSC 95-2221-E-009-078.

NOMENCLATURE

b	the probability for the state transition G to B
B	the bad state of the Gilbert channel
$\boldsymbol{c}_t$	the codeword of the VQ at time t
e_t	the bit-error pattern at time t
g	the probability for the state transition B to G
G	the good state of the Gilbert channel
k	the length of u_t
s_t	the state of the channel encoder
u_t	the binary index representing $\boldsymbol{c}_t$
v_t	the source vector
x_t	the code-bit combination after channel encoding
y_t	the received bit combination at the receiver
ε	bit error rate of the noisy channel

REFERENCES

Bahl, L. R., Cocke, J., Jelinek, F., and Raviv, J., 1974, "Optimal Decoding of Linear Codes for Minimizing Symbol Error Rate," *IEEE Transactions on Information Theory*, Vol. IT-20, No. 2, pp. 284-287.

Bernard, A., and Alwan, A., 2002, "Low-bitrate Distributed Speech Recognition for Packet-based and Wireless Communication," *IEEE Transactions on Speech and Audio Processing*, Vol. 10, No. 8, pp. 570-579.

Boulis, C., Ostendorf, M., Riskin, E., and Otterson, S., 2002, "Graceful Degradation of Speech Recognition Performance over Packet-erasure Networks," *IEEE Transactions on Speech and Audio Processing*, Vol. 10, No. 8, pp. 580-590.

Chouinard, J. Y., Lecours M., and Delisle, G. Y., 1988, "Estimation of Gilbert's and Fritchman's Models Parameters Using the Gradient Method for Digital Mobile Radio Channels," *IEEE Transactions on Vehicular Technology*, Vol. 37, No. 3, pp.158-166.

CoCentric System Studio-Referenec Design Kits, 2003, Synopsys, Inc., Mountain View, CA, USA.

Drukarev, A. I., and Yiu, K. P., 1986, "Performance of Error-correcting Codes on Channels with Memory," *IEEE Transactions on Communications*, Vol. COM-34, No. 6, pp. 513-521.

ETSI ES 202 212 v1.1.1. Digital Speech Recognition; Extended Advanced Front-end Feature Extraction Algorithm; Compression Algorithms; Back-end Speech Reconstruction Algorithm. November 2003.

Fingscheidt, T., and Vary, P., 2001, "Softbit Speech Decoding: a New Approach to Error Concealment," *IEEE Transaction on Speech and Audio Processing*, Vol. 9, No. 3, pp. 240-251.

Gilbert, E. N., 1960, "Capacity of a Burst-noise Channel," *The Bell System Technical Journal*, Vol. 39, No. 1, pp. 1253-1265.

Kanal, L. N., and Sastry, A. R. K., 1978, "Models for Channels with Memory and their Applications to Error Control," *Proceedings of IEEE*, Vol. 66, No. 7, pp. 724-744.

Lin, S., and Costello, D. J., 2004, *Error Control Coding*, Prentice Hall, NJ, USA.

Peinado, A. M., Sanchez, V., Perez-Cordoba, J. L., and Torre, A., 2003, "HMM-based Channel Error Mitigation and its Application to Distributed Speech Recognition," *Speech Communication*, Vol. 41, No. 2, pp. 549-561.

Reinhold, H. U., and Valentin, I., 2004, "Soft Features for Improved Distributed Speech Recognition over Wireless Networks," *Proceedings of 8th International Conference on Spoken Language Processing*, Jeju Island, Korea, pp. 2125-2128.

Turin, W., 2001, "MAP Symbol Decoding in Channels with Error Bursts," *IEEE Transactions on Information Theory*, Vol. 47, No. 5, pp. 1832-1838.

Wang, H. S., and Moayeri, N., 1993, "Modeling, Capacity, and Joint Source/Channel Coding for Rayleigh Fading Channels," *Proceedings of IEEE Vehicular Technology Conference*, Secaucus, NJ, USA, pp. 473-479.

Manuscript Received: Oct. 23, 2007
Revision Received: June 30, 2008
and Accepted: July 30, 2008