EMBEDDED JOINT SOURCE-CHANNEL CODING OF SPEECH USING SYMBOL PUNCTURING OF TRELLIS CODES

Alexis Bernard, Xueting Liu, Richard Wesel and Abeer Alwan
Electrical Engineering Department, University of California, Los Angeles
Box 951594, 405 Hilgard Avenue,
Los Angeles, CA 90095-1594

ABSTRACT

This paper presents an embedded joint source-channel coding scheme of speech. The source coder is an embedded variable bit rate perceptually based sub-band coder producing bits with different error sensitivities. The channel encoder is a Rate Compatible Punctured Trellis code (RCPT) which permits rate variability and unequal error protection by puncturing symbols. Furthermore, RCPT code design naturally incorporates large constellations, allowing high information rate per symbol. The embedded speech coder and the rate compatible puncturing of symbols provide the embeddibility of the joint coding scheme. The coder is robust to acoustic noise and produces good quality speech for a wide range of channel conditions (AWGN or fading), allowing digital transmission of speech with analog-like graceful degradation.

1. INTRODUCTION

Fixed target source bit rates impose an unnecessary constraint on transmission systems. Fixed rate error protection does not allow adaptation with channel conditions. Furthermore, equal error protection leads to a unique level of protection while the bits in the bitstream may be differently sensitive to transmission errors. In this paper, we describe a novel joint source-channel coder that allows for embedded variable bit rate speech coding with bit prioritization (Section 2), unequal error protection (Section 3) and how they combine to adapt to different channel conditions. The results are presented in Section 4 for both AWGN and Rayleigh fading channels using RCPT and Rate Compatible Punctured Convolutional code RCPC [7].

2. EMBEDDED VARIABLE BIT RATE SUB-BAND CODING WITH PERCEPTUAL BIT PRIORITIZATION

2.1 Sub-band coder

The encoder (Figure 1) is a modified version of the encoder described in [1]. The speech is first divided into 20 ms frames. An 8-channel IIR QMF filterbank divides the speech frame into 8 sub-bands which are then individually encoded. For each frame, dynamic bit allocation, according to the perceptual importance of each sub-band, is then performed. The MPEG psycho-acoustic model [2] estimates the signal to mask ratio (SMR) required in each band to mask the quantization noise. The SMR is a measure of the perceptual importance of each band. Then, a dynamic bit allocation scheme translates the SMR prescribed by the model into a bit assignment to further scalar quantize the sub-band samples using proportional allocation. The dynamic allocation of bits, which is the side-information of the coder, is transmitted with the coded bits.

Acknowledgment: This work was supported by the Rockwell/U.C. MICRO GRANT #98-170, by the NSF CAREER award 9733089 and by the USC IMSC (NSF ERC).

2.2 Dynamic bit allocation and bit error sensitivity

Dynamic bit allocation has three advantages. First, it shapes the quantization noise according to the spectrum of the speech signal. Second, it allows the same coder to work at different bit rates without any modification. A coder at a higher bit rate simply allocates the same bits as a coder operating at a lower bit rate, together with additional bits allowing marginal amelioration of the encoded speech signal. In other words, the coder operating at lower bit rate is embedded in the higher bit rates coder. Third, the bit allocation is progressive; it allocates first the bits with the most perceptual importance and ends by allocating the perceptually least important bits. This provide us with the bit prioritization necessary for later unequal error protection. Figure 2 shows an example of progressive bit allocation for the case of a coder operating at 18 kbps for a 4 kHz wide speech signal. Each frame (20 ms) is composed of 160 samples, divided into 8 sub-bands with 20 sub-band samples. Each block represents the allocation of 1 bit to all sub-band samples. Hence, the bitstream is prioritized on a 20-bits basis, selecting the blocks stage after stage in the bit allocation and from left to right. In this specific example, no bits were allocated to the 4^{th} sub-band as the signal in this sub-band is perceptually inaudible (SMR was negative). For comparison, the error sensitivity of each bit was simulated by systematically setting the same bit in error in every frame and measuring the resulting distortion of the speech signal. A frequency-weighted spectral distortion metric, similar to that in [4] was used. The frequency weighting is based on the SMR curves. As expected, the (almost) monotically decreasing nature of Figure 3 justifies the ranking of the importance of the 18 groups of bits. For instance, the 3 first groups of bits, corresponding to the crucial side information, show high distortion. Note that small variations in the distortion metric can have noticeable influence on the speech quality. The coder has an overall delay of 25 ms and offers variable bit rate (VBR) in the range from 8 kbps (MOS=2.5) to 32 kbps (MOS=5).

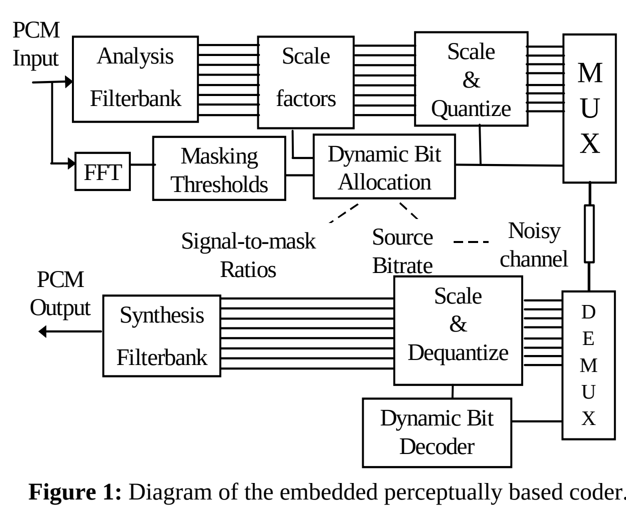

Figure 1: Diagram of the embedded perceptually based coder.

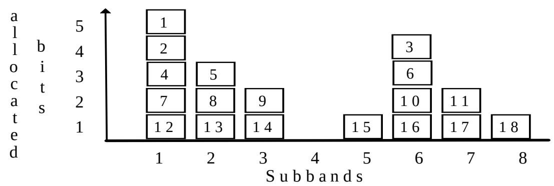

Figure 2: Bit allocation (*y axis*) and bit prioritization (*number inside the blocks*) for the 18 groups of 20 bits.

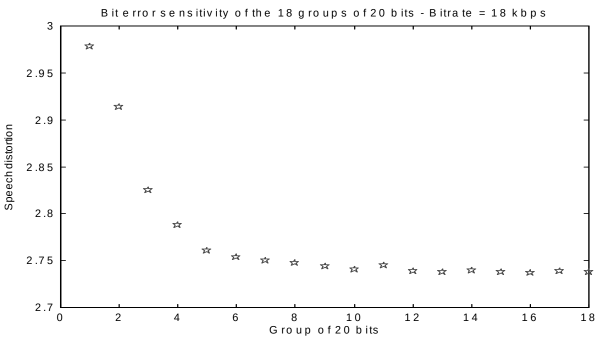

Figure 3: Bit error sensitivity analysis.

3. EMBEDDED UNEQUAL ERROR PROTECTION AND PROGRESSIVE PUNCTURING OF SYMBOLS

In an unequal error protection scheme, all the bits in the bitstream should contribute the same amount to the overall noise distortion after transmission errors. This leads to more protection for the sensitive bits and less protection or no protection at all for the least sensitive bits. Unequal error protection was introduced in [5-6] using the rate compatible punctured convolutional codes (RCPC) [7]. In [8], a punctured Reed-Solomon coder was proposed. In both cases, bits are punctured from the bitstream. We explore symbol puncturing of a trellis code.

3.1 Progressive puncturing of symbols in an 8-PSK trellis code

In this paper, a 16-state rate-1/3 8-PSK Rate Compatible Punctured Trellis code (RCPT) originally presented in [9] is used. This trellis code is structured specifically for periodic puncturing with a period of 8, and supports puncturing anywhere from 0 to 5 out of every 8 symbols. The 8-PSK labeling used is the gray labeling 0, 2, 3, 1, 5, 7, 6, 4 around the circle. The encoder generator matrix is [32 11 27] in octal notation. Bit error rates (BER) associated with the 6 different levels of puncturing (*a to f*) and the uncoded curves simulated for 8-PSK and 4-PSK for an AWGN channel and an independent Rayleigh fading channel are shown in Figures 4 and 5, respectively. The figures show the BER obtained with the 16-state rate-1/2 4-PSK RCPC code presented in [7] whose generator matrix is [23 35]. Note that puncturing 5 out of 8 symbols reduces the redundancy to zero. Even if the trellis code operates when the redundancy is reduced to zero, we choose, for complexity purposes, to leave these bits uncoded (curve *8 PSK)* instead of encoding them and then puncturing all the redundancy.

As seen in the figures, the simplicity offered by symbol puncturing is not at the cost of severe degradation in protection

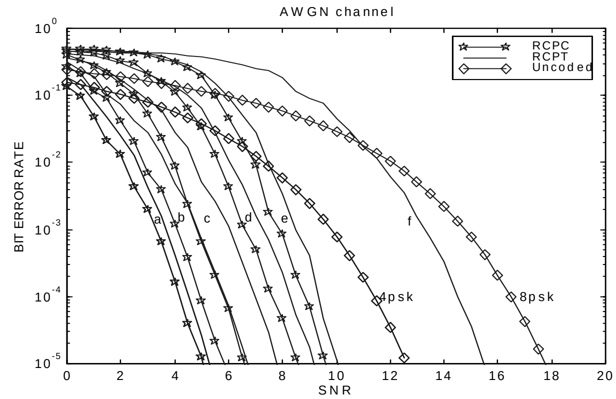

Figure 4: Bit error rate curves for the RCPC and RCPT encoding schemes under AWGN channel.

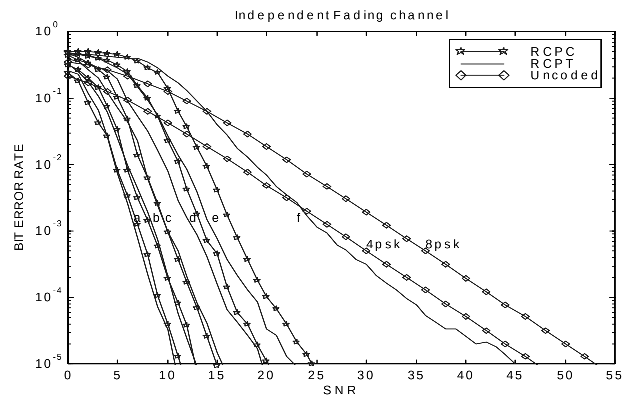

Figure 5: Bit error rate curves for the RCPC and RCPT encoding schemes under independent fading.

performances. It should be noticed that for the independent Rayleigh fading channel, some bit interleaving such as that proposed in [3] would improve the performance of the RCPC channel encoder. Indeed the effect of a fade on a 4 PSK constellation would then be spread over in the trellis. For correlated fading, symbol interleaving can be applied to the RCPT scheme and bit interleaving to the RCPC scheme.

A puncturing pattern that removes q symbols out of p symbols (where p is the puncturing period) is a p-q pattern. The per-symbol information rate associated with a p-q puncturing applied to our rate-1/3 trellis code is then given by $R= p/(p-q)$. With q ranging from 0 to 5, we see that the per-symbol information rate of our RCPT coder ranges from 1 (full protection) to 3 (uncoded). By contrast, the per-symbol information rate of the 4-PSK RCPC coder ranges only from 1 to 2 (uncoded), allowing less flexibility in the choice of unequal error protection.

Tables 1 and 2 summarize the number of bits or symbols punctured (q) the puncturing pattern, the per-symbol information rate (R), the Residual Euclidean Distance (RED), the Periodic Effective Code Length (PECL), the Periodic Product Distance (PPD) and the number of nearest neighbors (N) for both the RCPC and the RCPT coders. As expected from Figure 4 and 5, the tables confirm that RCPC performs slightly better over AWGN, at the cost of a higher traceback depth. Over fading channel, RCPT outperforms under heavy puncturing (curves *d* and *e*) despite slightly smaller PPD and PECL.

ID	q	R	RED[2]	N	Puncturing	PECL	PPD
a	0	1.00	12.58	1	00000000	5	37.4
b	1	1.14	8.34	.125	10000000	4	0.94
c	2	1.33	6.34	0.25	10001000	3	0.47
d	3	1.60	4.58	.125	10101000	2	2.32
e	4	2.00	4.58	1.5	10101010	2	2.32
f	5	2.60	1.171	1.75	11101010	1	0.33
8 PSK		3.00					

Table 1: Characteristics of the 8-PSK 16-states (υ=4) rate-1/3 puncturing period-8 RCPT. (*p* = 8 symbols)

ID	q	R	RED[2]	N	Puncturing	PECL	PPD
a	0	1.00	14	2	00000000 00000000	5	128
b	2	1.14	10	0.25	00000000 00010001	4	32
c	4	1.33	8	0.5	00000000 01010101	4	16
d	6	1.60	6	1	00000000 01110111	3	8
e	7	1.77	4	.125	00001000 01110111	2	4
4 PSK		2.00					

Table 2: Characteristics of the 4-PSK 16-states (υ=4) rate-1/2 puncturing period-8 RCPC. (*p* = 16 bits)

RCPT can offer up to 3 information bits per-symbol which is advantageous at high SNR. Also, puncturing symbols in RCPT is an easier task than puncturing bits in RCPC. Another particularly attractive feature of the RCPT coder is that, even in the presence of a deep fade or strong interference that can be considered as a form of puncturing, the coder is robust while traditional trellis codes with uncoded bits fail under such conditions. Finally, RCPT allows embedded channel coding without channel decoding the received symbols. In other words, in case of traffic congestion, the decision to not transmit (puncture) some symbols in order to decrease the symbol throughput can be made at any intermediate node. As long as the node communicates the receiver that the baud rate was decreased by puncturing symbols in the source stream and/or in the redundancy stream, decoding and reconstruction of the signal is done using the same decoder.

3.2 Dynamic and channel dependent puncturing of the bit-, symbolstream

Informal listening tests showed that side information containing the bit allocation and the scale factors must be transmitted with BER<0.05%, the next allocated groups should remain in the range 0.1%<BER<0.5% while the last groups of bits tolerate BER as high as 2%. These different tolerated BER levels can only be obtained after applying unequal error protection depending on the channel conditions.

Our goal is to design an overall channel adaptive joint source-channel coding system that provides speech quality that is consistently good over a wide range of channel conditions for a given symbol rate of 10 kbauds.

Rate kbps	Bits/ Frame	RCPT AWGN	RCPC AWGN	RCPT Fading	RCPC Fading
10	200	a_{200}	a_{200}	a_{200}	a_{200}
12	240	$a_{60}b_{40}c_{140}$	$a_{60}b_{40}c_{140}$	$a_{60}b_{40}c_{140}$	$a_{60}b_{40}c_{140}$
14	280	$b_{60}c_{80}4_{140}$	$b_{60}c_{80}d_{140}$	$b_{60}c_{80}4_{140}$	$b_{60}c_{80}d_{140}$
16	320	$b_{20}c_{60}d_{140}4_{100}$	$c_{80}d_{140}e_{40}4_{60}$	$b_{20}c_{80}d_{100}e_{120}$	$c_{80}d_{140}e_{40}4_{60}$
18	360	$c_{60}d_{120}4_{120}8_{60}$	$d_{80}e_{160}4_{120}$	$c_{20}d_{120}e_{220}$	$d_{80}e_{160}4_{120}$
20	400	$d_{80}e_{340}8_{200}$	4_{400}	$c_{20}d_{100}e_{140}f_{140}$	4_{400}
22	440	$e_{320}8_{120}$		$d_{60}e_{180}f_{140}8_{60}$	
24	480	$e_{200}f_{160}8_{120}$		$e_{200}f_{160}8_{120}$	
26	520	$e_{120}f_{160}8_{240}$		$e_{120}f_{160}8_{240}$	
28	560	$f_{320}8_{240}$		$f_{320}8_{240}$	
30	600	8_{600}		8_{600}	

Table 3: Optimal unequal error protection architecture for AWGN and Rayleigh fading channels (max.10 kbauds).

We assume that channel conditions are known to the joint encoder. In order to achieve high speech quality for different channels, the joint source-channel coder dynamically varies both the source coding bit rate and the channel coding unequal error protection. Both source and channel coders are embedded in respectively higher bit rate source encoders and higher redundancy channel encoders. The change can happen within one frame duration (20 ms). For high SNR channels, fewer bits are allocated to the channel encoder, permitting more bits for the source encoder, thus improving the speech quality. For low SNR channels, fewer bits would be allocated to the speech encoder but these bits would be more heavily protected.

The design of the source-channel coder system, leading to the highest speech quality obtainable at a given bit rate, is obtained in three steps. First, we analyze the levels of protection needed in order to obtain the aforementioned BER for the different parts of the bitstream for every particular SNR. Second, we determine the maximum source coding bit rate that can satisfy these BER conditions given the average redundancy inferred by the levels of protection required. Finally, the puncturing architecture of the bitstream is derived so that the final source-channel coding bitstream equals 20 kbps for the 4-PSK RCPC or 30 kbps for the 8 PSK RCPT. Table 3 summarizes the results of both channel encoders for source coding bit rates in steps of 2 kbps and for AWGN and fading channels. The notation $e_{320}8_{120}$ means that the 320 first bits are coded with the curve *e* and the 120 last bits are left uncoded on a 8-PSK constellation. The architectures are known by both the sender and the receiver. Hence, the side information for the channel encoder is simply the operating source coding bit rate.

4. SIMULATIONS

Figure 6 shows the quality of the different source-RCPT channel encoder pairs simulated on an independent Rayleigh fading channel (for clarity, only a few of the source coding bit rate possibilities are shown). As expected, no specific pair systematically outperforms the other pairs. At low SNR the 10 kbps source encoder with full protection outperforms, while at high SNR, the encoders with large source coding bit rates provide the least speech distortion. At every SNR, we select the source-channel system that provides the best speech quality. The overall distortion-SNR curve is simply the lower envelope of all the curves. Figure 7 shows the minimum perceptual distortion obtained for every SNR using both the RCPT and the RCPC method. The perceptual distortion measure is a modified version of the one proposed in [4] with the frequency weighting being proportional to the SMR in each band. Note that the speech distortion decreases with increasing SNR and is kept limited even at very low SNR; this would not be the case for fixed source bit rate systems with fixed and equal channel protection. Furthermore, we notice that the RCPT encoder provides identical speech quality to RCPC at intermediate SNRs and better speech quality at high SNR, due to its higher per-symbol information of the 8-PSK RCPT and the better behavior of the curves *d* and *e* in Figure 5. Indeed with a 4-PSK constellation, RCPC allows only up to 20 kbps joint source-channel bit rates, while the 8-PSK constellation of the RCPT permits 30 kbps overall bit rates. This effect is noticeable only at high SNR because at intermediate SNR the per-symbol information rates obtained from both coders are similar. Informal listening tests showed Mean Opinion Scores (MOS) ranging from MOS=3 when SNR=5 dB to MOS=5 for higher SNR. It should also be noticed that, due to the property that the source coder is noise-robust, the overall system remains robust to, for instance, road noise.

5. SUMMARY AND CONCLUSIONS

This paper describes how to combine an embedded perceptually based variable bit rate speech encoder and a rate compatible punctured trellis code to obtain high quality speech over a wide range of channel conditions. The speech encoder produces a prioritized bitstream that is then encoded against transmission errors with unequal error protection depending on the bit error sensitivity. The speech quality obtained is good at overall baud rates as low as 10 kbauds and monotically increases with the channel SNR. The joint source-channel coder is robust to acoustic noise and is capable of using different source coding rates and channel coding redundancies. Using a single embedded speech encoder/decoder and a single rate compatible punctured trellis encoder/decoder procures the joint embeddibility of the scheme. No interruption in the communication is required to switch from one configuration to another and changes take place within 20 ms. The robustness, the simplicity and the flexibility of the symbol puncturing scheme make the system suitable for mobile communications over fading channels where a deep fade or strong interference can be modeled as symbol puncturing. The bit prioritization and the embeddibility of the source-channel coder makes it suitable for packet oriented communication link where, in case of traffic congestion, certain groups of symbols can be dropped without degrading significantly the speech quality. For slowly fading channels, one could track the channel conditions and continuously adapt the joint source-channel system. Future work will include applying unequal error protection to variable bit rate hybrid source coding schemes, such as the Multi-Band CELP (MB-CELP), examining the performance over a variety of correlated fading channels and considering the networking issues involved in such a variable bit and baud rate communication link.

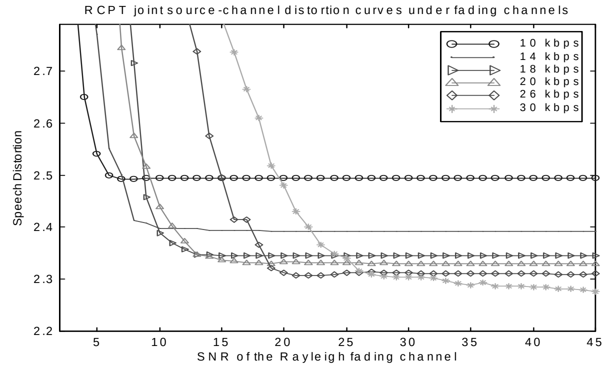

Figure 6: Distortion-SNR curves for some source coding rates under fading channels using RCPT. Overall rate = 30 kbps.

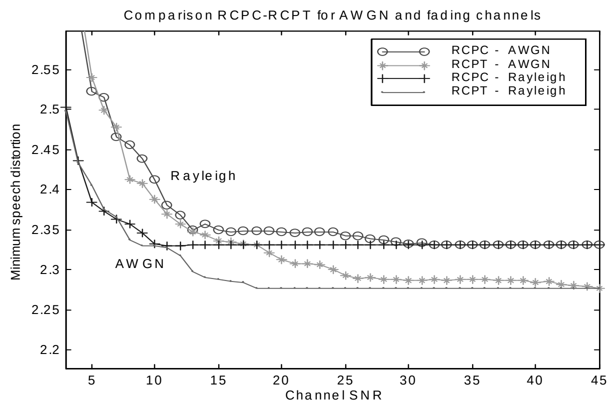

Figure 7: Comparison of joint source-channel encoding with RCPC and RCPT over AWGN and fading channels.

6. REFERENCES

[1] B. Tang, A. Shen, A. Alwan, and G. Pottie, A Perceptually-Based Embedded Sub-band Speech Coder. IEEE Trans. on Speech and Audio Proc., 5, no. 2, March 1997, pp.131-140.

[2] Stoll, G.; Brandenburg, K. The ISO/MPEG-audio codec: a generic standard for coding of high quality digital audio. ITG-Fachberichte, vol. 118, Mannheim, Germany, 1992.

[3] G. Caire, G. Taricco, E. Biglieri; Bit interleaved coded modulation, IEEE Trans. informat. theory, 44, no.3, May 98.

[4] McCree, A.; Truong, K.; Bryan George, E.; Harnwell, T.; Viswanathan, V. A 2.4 kbps MELP coder candidate for the new U.S. Federal Standard. 1996 IEEE ICASSP, pp. 200-3.

[5] Goodman, D.J.; Sundberg, C.-E. Combined source and channel coding for variable-bit-rate speech transmission. Bell System Technical Journal, 62, no.7, Sept. 1983. pp. 2017-36.

[6] Cox, R.V.; Hagenauer, J.; Seshadri, N.; Sundberg, C.-E. A sub-band coder designed for combined source and channel coding. 1988 ICASSP, NY, USA, IEEE, 1988. pp. 235-8, 1.

[7] Hagenauer, J. Rate-compatible punctured convolutional codes (RCPC codes) and their applications. IEEE Trans. on Communications, 36, no.4, April 1988. pp.389-400.

[8] Hong Shi; Ho, P.K.M.; Cuperman, V. Combined speech and channel coding for mobile radio communications. IEEE Trans. Vehicular Technology, 43, Nov. 1994, pp. 1078-87.

[9] R. Wesel, X. Liu, W. Shi. Periodic symbol puncturing of trellis code. Thirty-first Asilomar conference on signals, systems and computers, Nov. 1997, invited.