Channel Adaptive Joint Source-Channel Coding of Speech

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

Abstract

In this paper, we present a novel joint source-channel coding scheme for speech signals. The source coder is a perceptually based sub-band coder producing bits with different error sensitivities. The channel encoder is a Rate Compatible Punctured Trellis code (RCPT) that permits unequal error protection. RCPT code design naturally incorporates large constellations, allowing a high information rate per symbol. The source coder is robust to acoustic noise, adapts automatically every 20 ms and produces good quality speech for a wide range of channel conditions. The paper presents a method for finding the optimal puncturing architecture for different source bitrates and channel conditions. The resulting joint source-channel coder is suitable for situations where the channel is time varying such as in mobile communications.

1. Introduction

Fixed target source bitrates impose an unnecessary constraint on transmission systems. Fixed rate error protection does not allow adaptation with channel conditions. In addition, equal error protection leads to a unique level of protection while the bits in the bitstream may be diversely sensitive to transmission errors. In this paper, we describe a novel joint source-channel coder that combines embedded variable bit rate speech coding (Section 2) with unequal error protection (Section 3). Both techniques adapt to match different channel conditions. The results are presented in Section 4.

Acknowledgment: This work has been supported by the NSF CAREER award #NCR-9733089, Rockwell Semiconductor Systems through California MICRO grant 98-170 9733089 and by the USC IMSC (NSF ERC).

2. Variable bitrate perceptual sub-band coding with bit prioritization

2.1 Sub-band coder

The encoder (Figure 1) is a modified version of [1]. It divides the speech into 20 ms frames. An 8-channel IIR QMF filterbank divides the speech frame into 8 sub-bands that are then individually encoded. For each frame, dynamic bit allocation, according to the perceptual importance of each sub-band, is performed. The MPEG psycho-acoustic model [2] estimates the signal to mask ratio (SMR) required in each band to mask the quantization noise [3]. The SMR is the 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. The dynamic allocation of bits, as side-information of the coder, is transmitted with the coded bits.

2.2 Dynamic bit allocation and bit error sensitivity

Dynamic bit allocation has three advantages. First, it shapes the quantization noise with respect to the spectrum of the speech signal, minimizing the perceptual noise. Second, it allows the same coder to easily adapt to different bitrates. For a higher bitrate, the encoder allocates the same bits as for the lower bitrate encoder, together with additional bits allowing further improvement of the encoded speech signal. Finally, the bit allocation is progressive and first allocates the bits with the most perceptual importance. Figure 2 shows an example of progressive bit allocation for the case of a coder operating at 22 kbps for a 4 kHz wide speech signal. Each frame (20 ms) is composed of 160 samples, divided into 8 sub-bands with 20 samples per sub-band. Each block represents the allocation of 1 bit to all sub-band samples and prioritizes the bitstream on a 20-bits basis.

The error sensitivity of each bit is computed by systematically setting the same bit in error in every frame and measuring the resulting distortion of the speech signal. The distortion metric is a variation of the one proposed in [4]. The distortion is frequency weighted and the different weights depend on the SMR of the frequency bands. The monotically decreasing nature of Figure 3 validates the ranking of the importance of the 22 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 a noticeable effect on speech quality.

3. Unequal error protection and progressive puncturing

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 has been pioneered by Sundberg [5-6] using the rate compatible punctured convolutional codes (RCPC) introduced by Hagenauer [7]. Shi, Ho and Cupperman also presented a punctured Reed-Solomon encoder [8]. In both cases, they puncture bits from the bitstream. We explore symbol puncturing of a trellis code, operating jointly with a perceptually based source coder that produces bits with a wide range of error sensitivity.

3.1 Progressive puncturing of symbols on a 8-PSK trellis code

In this paper, a 64-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 9, and supports puncturing anywhere from 0 to 6 out of every 9 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 of the coder is [165 142 127]. One should note that puncturing 6 out of 9 symbols reduces the redundancy to zero. Even though this trellis code can work when the redundancy is reduced to zero, we choose, for complexity purposes, to leave these bits uncoded instead of encoding them and then puncturing all the redundancy.

Figure 4 shows the bit error rates (BER) associated with the 6 different levels of puncturing and the uncoded curve simulated for 8-PSK on an additive white Gaussian noise channel (AWGN). It also shows the BER obtained with the 64-state rate-1/2 4-PSK RCPC code presented by Hagenauer [7]. The information rates of these curves are also summarized in Tables 1 and 2.

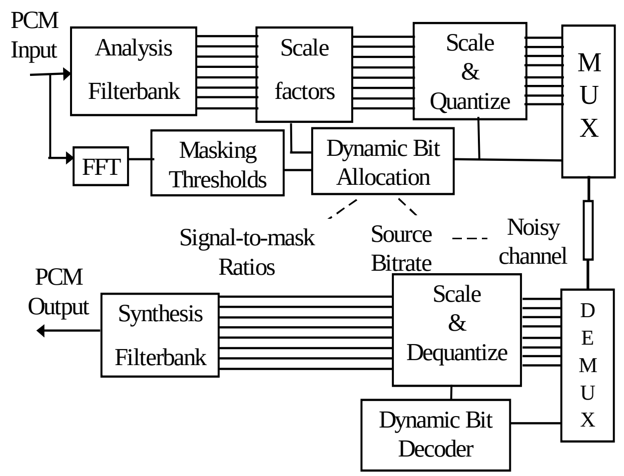

Figure 1: Diagram of the embedded perceptually based coder.

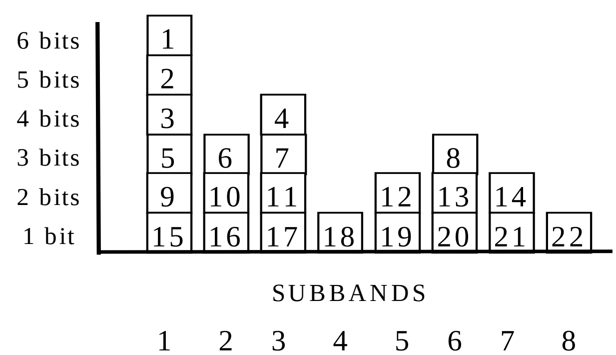

Figure 2: Bit allocation for 22 groups of 20 bits.

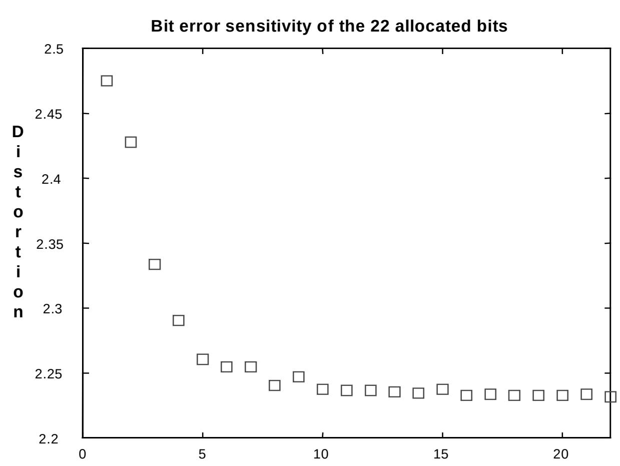

Figure 3: Average bit error sensitivities of the 22 groups of bits

An important question in the decoding of both encoders is the determination of the traceback depth L_D. For standard trellis code, Anderson [10] computed L_D as the trellis depth at which all unmerged error events have more Euclidean distance than the minimum Euclidean distance d_{free} of the trellis. For punctured trellis code, the traceback depth L_D (including branches that increase depth but add zero distance) is the one at which all unmerged incorrect paths exceed the Residual Euclidean Distance $(RED)_q$ for that puncturing pattern. The RED is defined as the Euclidean distance remaining after puncturing q symbols out of p symbols.

Such 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 6, 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 puncturing pattern, the per-symbol information rate, the residual Euclidean distance and the traceback depth L_D for both RCPC and RCPT coders. Figure 4 shows that RCPC and RCPT coders have similar BER *vs.* it allows more flexibility in the per-symbol information rate, it can offer up to 3 information bits per symbol and it has significantly lower L_D. Figure 4 shows the BER performance of both coders with $L_D=41$ but we will be using $L_D=32$ in our implementation. The RCPT coder is also less complex because puncturing symbols is an easier task than puncturing bits in RCPC. Another particularly attractive feature of the RCPT coder is that, even in the presence of 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.

3.2 Dynamic and channel dependent puncturing schemes of the bitstream

Unlike data, speech is somewhat insensitive to channel errors. Informal listening tests showed that the side information containing the bit allocation and the band scaling factors must be transmitted with a BER<0.05%, the next 4 allocated groups should remain in the range 0.1%<BER<3% while the last groups of bits can be transmitted with BER as high as 2%. These different tolerated BER levels can only be obtained after applying unequal error protection depending on the channel conditions.

Table 1: Characteristics of the 8-PSK 64-states rate-1/3 period-9 RCPT

ID	q	Information Rate/ Symbol	Residual Euclidean Distance ^{2}	Puncturing Pattern	Depth L_D
a	0	1.000	16	000000000	24
b	1	1.125	11.172	000000001	25
c	2	1.286	9.172	000010001	26
d	3	1.500	6.929	001010001	31
e	4	1.800	4.343	001010011	31
f	5	2.250	1.757	011010011	27
g	6	3.000	0.586	011010111	36
8	uncoded 8-PSK	3.000			

Table 2: Characteristics of the 8-PSK 64-states rate-1/2 period-8 RCPC

ID	bits punctured	Information Rate/ Symbol	Residual Euclidean Distance ^{2}	Puncturing Pattern	Depth L_D
A	0	1.000	20	00000000 00000000	27
B	2	1.143	14	00000000 00010001	27
C	4	1.333	12	00000000 01010101	34
D	6	1.600	8	00000000 01110111	48
E	7	1.778	6	00001000 01110111	93
4	uncoded 4-PSK	2.000			

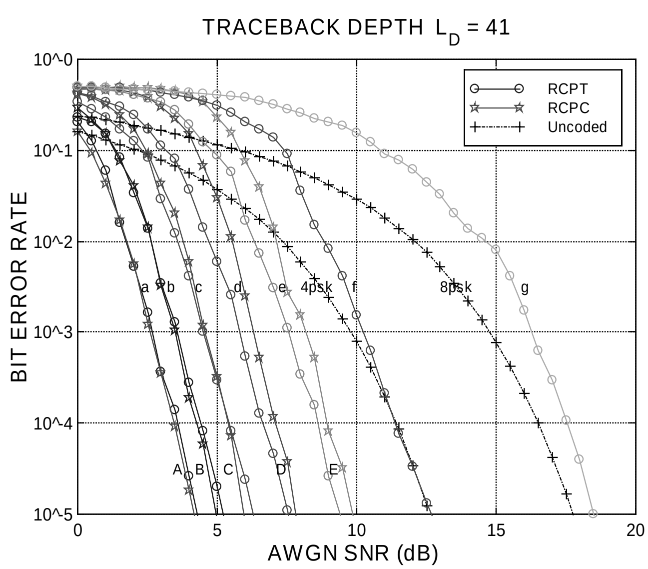

Figure 4: Bit error rate curves for both RCPT and RCPC channel encoders

Our ultimate goal is to design an overall channel adaptive joint source-channel coding system that provides speech quality that is consistently good (given the channel quality) over a wide range of channel conditions for a given symbol rate of 10 kbauds. We assume that channel conditions are known to the speech-channel encoder. In order to achieve high speech quality for different channels, the joint source-channel coder dynamically varies both the source coding bitrate and the channel coding unequal error protection. Both source and channel coders are embedded in respectively higher bitrate source encoders and higher redundancy channel encoders.

Adaptation can occur within one frame duration (20 ms). For good 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 optimal 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 bitrate 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 coded bitstream equals 20 kbps for the 4-PSK RCPC or 30 kbps for the 8 PSK RCPT. Table 3 summarizes the puncturing architecture for both channel encoders and for source coding bitrates in steps of 2 kbps.

4. Simulations and performances

Figure 5 shows the quality of the different source-channel encoder pairs simulated with $L_D=32$ on AWGN channels (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 bitrates 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 envelope of all the curves. Figure 6 shows the minimum distortion obtained for every SNR using both the RCPT and the RCPC method. Note that speech distortion is decreasing with increasing SNR and is kept limited even at very low SNR; this would not be the case for fixed source bitrate systems with fixed and equal channel protection.

Table 3: Optimal unequal error protection architecture for RCPT and RCPC

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

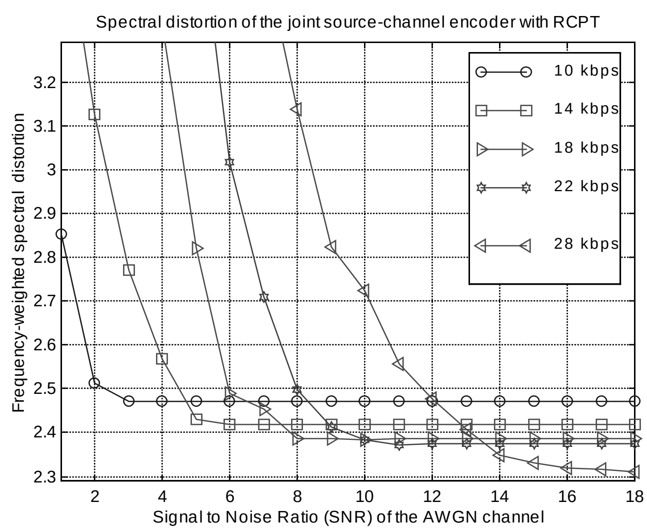

Figure 5: Spectral distortion of the different source-channel coder pairs

We notice that the RCPT encoder provides identical speech quality than the RCPC at intermediate SNR and better speech quality at high SNR, due to its higher per-symbol information rate. Indeed with a 4-PSK constellation, RCPC allows only up to 20 kbps joint source-channel bitrates, while the 8-PSK constellation of the RCPT permits 30 kbps overall bitrates. This effect is noticeable only at high SNR because at intermediate SNR the per-symbol information rates obtained from both coders are similar. At high SNR, indeed, the mutual information of a 8 PSK constellation exceeds the maximum mutual information of a 4 QAM constellation.

Informal listening tests showed Mean Opinion Scores (MOS) ranging from 3 when SNR=4 dB to 5 for high 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. Conclusions

We have shown in this paper how to combine a 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 speech encoder/decoder and a single trellis encoder/decoder. No interruption in the communication is required to switch from one configuration to another and changes take place within 20 ms.

The robustness and flexibility of the symbol puncturing scheme make this type of architecture promising for mobile communications over Rayleigh or Rician fading channels where deep fade or strong interference can be modeled as symbol puncturing. For slowly fading channels, one can even track the channel conditions and continuously adapt the joint source-channel system. This is a natural extension of the adaptive rate scheme of Goldsmith [11] and Goeckel [12] for channel where the transmitter knows the channel conditions with a relatively small tracking delay. Finally, a technique of finding the optimal puncturing schemes for different channel conditions and source bitrates has been presented.

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.

6. References

[1] B. Tang, A. Shen, A. Alwan, and G. Pottie, "A Perceptually-Based Embedded Sub-band Speech Coder", IEEE Transactions on Speech and Audio Processing, Vol. 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, 18-20 Feb. 1992.

[3] Brandenburg, K.; Sporer, " 'NMR' and 'masking' evaluation of quality using perceptual criteria", AES 11th International Conference AES Test and Measurement, NY, USA: Audio Eng. Soc., 1992. pp. 169-79.

[4] McCree, A.; Truong, K.; Bryan George, E.; Harnwell, T.P.; Viswanathan, V. "A 2.4 kbits/s MELP coder candidate for the new U.S. Federal Standard", International Conference on Acoustic, Speech and Signal Processing, 1996, NY, USA, 1996. pp. 200-3 vol. 1.

[5] Goodman, D.J.; Sundberg, C.-E. "Combined source and channel coding for variable-bit-rate speech transmission" Bell System Technical Journal, vol.62, 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", ICASSP 98, NY, USA, IEEE, 1988. pp.235-8 vol.1.

[7] Hagenauer, J. "Rate-compatible punctured convolutional codes (RCPC codes) and their applications", IEEE Transactions on Communications, vol.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 Transactions on Vehicular Technology, vol.43, (no.4), 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.

[10] Anderson, J.B.; Balachandran, K. "Decision depths of convolutional codes", IEEE Transaction on Information Theory, vol.35, (no.2), March 1989. pp.455-9.

[11] Goldsmith A., S. G. Chua, "Variable-Rate Variable Power MQAM for Fading Channels", IEEE Transactions on Communications, Oct. 1997.

[12] Goeckel D., "Adaptive Coding for Time-Varying Channels Using Outdated Fading Estimates", submitted to TCOM 99.

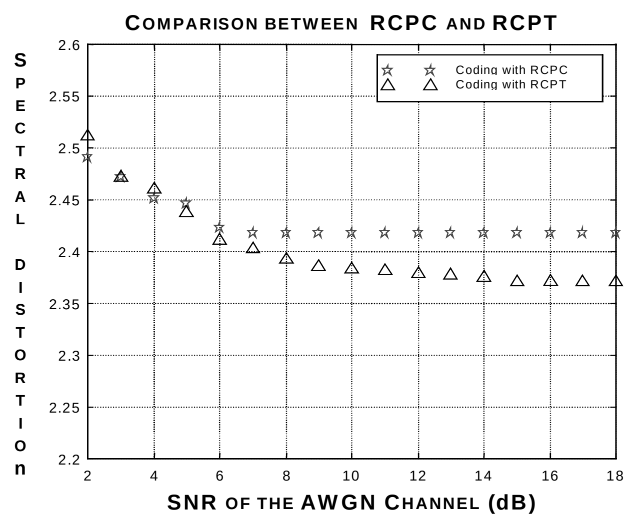

Figure 6: Minimum speech distortion at every SNR for both RCPC and RCPT