A ROBUST SPEECH CODER INCORPORATING JOINT SOURCE AND CHANNEL CODING TECHNIQUES

Philip Secker and Andrew Perkis

Department of Elec. and Comp. Engineering
University of Wollongong
Australia

ABSTRACT - This paper discusses the incorporation of joint source and channel coding techniques into a high quality 12kbps speech coder. Trellis structures are used to quantize both the Linear Predictive coefficients and the residual signal with each trellis optimised to the expected channel distortion. The resulting system shows remarkable robustness to a wide range of bit error rates, with most gain achieved at rates as high as 0.1 .

INTRODUCTION

Communication channels such as satelite radio are characterised by high bit error rates (BER) and burst error durations. With narrow to medium-band speech coding the impact of channel errors on the quality of the reconstructed speech can be quite severe. The application of low bit-rate speech coding systems to less than ideal communication channels usually requires a tradeoff between source coding and channel coding. Although, fundamentally, source coding and channel coding can be separated without loss of optimality, the possibility of combining the process into a single coding algorithm is attractive with the expectation of a less complex system without the bandwidth increase associated with channel coding. Ayanoglu and Gray (1987) presented a trellis coding algorithm that source encodes using the expected channel transition probability. This joint source and channel coding technique is attractive in low bit-rate speech coding over noisy channels where the bandwidth is not available for channel coding schemes such as foward error correction.

CODER DESCRIPTION

To investigate joint source and channel coding methods and its applicability to speech coding, a coder was developed to act as a platform on which these techniques could be applied. In its basic operation it produces high quality speech at 12kbps. The speech coder developed is an adaptive residual coder (ARC) with noise spectrum shaping as shown in Figure 1 (Svendsen, 1985). The coder works on 128 bit frames (16 ms) of 8kHz sampled speech. Specifically, joint source and channel coding has been applied to both the residual (which is transmitted at 1 bit per sample) and LPC coefficients (33 bits per frame) with pitch and gain parameters remaining unprotected.

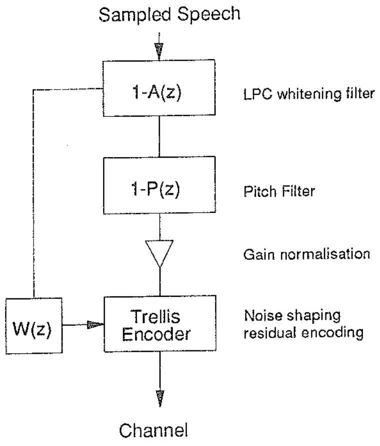


Figure 1. Basic ARC speech encoder

The coder uses 10th order LPC analysis performed on 16 ms frames. The LPC coefficients are transformed to Line Spectral Pairs (LSP) (Soong & Juang, 1984) and trellis encoded at 33 bits. The LPC filtered speech contains impulses spaced at the fundamental pitch period. These are filtered with a simple one-tap filter of the form $P(z) = 1 - bz^{-M}$ with a search for M over the range of 20 to 120 sample lags corresponding to pitch frequencies of 400 to 67 Hz. Pitch parameters are updated at twice the frame rate (64 sample frames). Simple logarithmic scaling is applied to the gain which is then scalar quantized. The residual is trellis encoded at 1 bit per sample using a noise shaping distortion measure. Straight-forward quantization of the residual with a per sample distortion such as the squared error criterion gives less than satisfactory performance. It is well known that due to the auditory masking properties of the human ear, more noise can be tolerated in the louder formant frequency regions. Thus a weighting filter is used in the distortion calculation of the form $W(z) = 1/(1 - A(z/\gamma))$ where $A(z)$ represents the forward predictor of the LPC synthesizer (Atal 1982). γ is a parameter that controls the level of noise power in the formant regions.

TRELLIS CODING

A trellis source encoder seeks to find the optimal channel sequence for the shift register decoder of Figure 2. The contents of the shift register are used to address a look-up table (or codebook) containing the reconstruction values. The coder is characterised by the length of the shift register (called the constraint length K). A bit-rate of 1 bit per sample is established with 2^K reconstruction values available. The Viterbi algorithm is used to search the trellis encoding structure shown in Figure 3 to minimize a cost function such as the squared error distortion between source sample x and quantization value $\hat{x}$

$$d(x,\hat{x}) = (x - \hat{x})^2 \quad (1)$$

Each path maintains a metric which is the accumulated sum of $d(x,\hat{x})$. The least distortion path (minimum metric) is chosen and after a given decision depth a channel bit is released. The algorithm provides an optimal path but at a computational cost increasing exponentially with increasing K. Populating the codebook has been addressed by Stewart et. al. (1982) whereby the generalized Lloyd algorithm is used to iteratively train the codebook with a training database. Trellis coding systems approach the rate-distortion limits with increasing K and compare favourably with Vector Quantizers for multi-dimensional data compression.

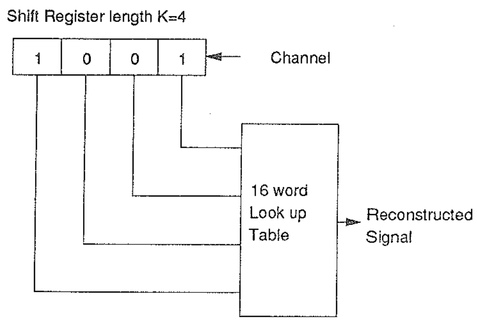


Figure 2. Trellis Decoder

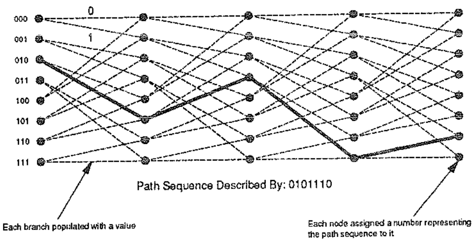


Figure 3. Trellis Encoder (K=4)

JOINT SOURCE AND CHANNEL CODING

Joint source and channel coding seeks to combine data compression with channel coding so as to minimize the loss of fidelity when subjected to channel noise. For the trellis encoder, this is accomplished by reducing the impact of an incorrect branch being taken due to channel errors. In effect, the path sequence is chosen so that if a channel error occurs, the resulting distortion is kept to a minimum.

Applied to a trellis coder, the difference between a source coder and the joint source and channel coder is the additive cost function used for weighting each path metric. The additive cost function for branch u_0 consists of the sum of the squared error measures between the source sample x and all possible branch values *B* that could occur should the path sequence be received in error, weighted by the probability of error. Thus the distortion measure (1) is modified to

$$d(x,y) = \sum_{u \in B} (x - f(u))^2 Pr(u|u_0) \tag{2}$$

where f(u) is the branch value associated with branch u, $y = f(u_0)$ and $Pr(u|u_0)$ is the conditional probability that u is received given that u_0 was sent. This probability distribution is dependent upon the expected BER of the channel.

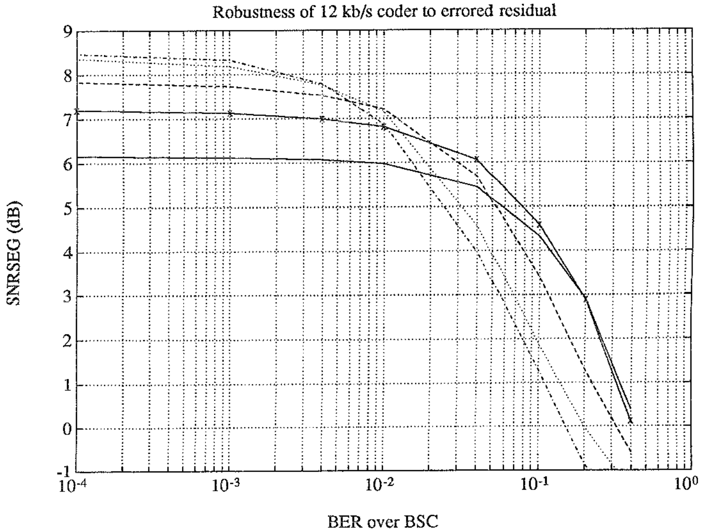
Applying the generalized Lloyd algorithm of Ayanoglu and Gray using the distortion measure of (2) to encode produces a channel-optimised (noisy) codebook that is robust to a given channel error distribution.

In practice with channels such as mobile radio and satellite links, the channel error probability is non-stationary. Deep fades can be construed as a high BER for a short period of time, while at other times the BER may be very low. Therefore, the question remains as to which channel probability of error to design for given the range of expected channel error rates.

RESIDUAL CODING

The less critical (perceptually) residual signal was chosen as a first candidate for the application of joint source and channel coding to a speech coder. Channel-optimised codebooks were designed for BERs of 0.01, 0.04, 0.1 and 0.2 for a Binary Symmetric Channel (BSC) using 8 seconds of speech. The complete coder was simulated with each of the codebooks incorporated into the residual trellis coder. The coded residual channel sequence was subjected to channel error rates of 0, 0.001, 0.004, 0.01, 0.04, 0.1, 0.2, 0.4 and the segmental SNR between original and reproduced speech was measured. Figure 4 demonstrates the effect of mismatching the codebook to the channel. While each codebook dominates other codebooks for its designed BER, it is evident that if the coder is to operate over a given range of BER (as with land mobile radio) there appears to be some codebooks that are more robust than others at covering the range. These preliminary results indicate that the 0.04 and 0.1 BER codebooks are the most robust to channel error probabilities in the range of 0.001 to 0.2, an appropriate range for speech coders operating in the mobile environment.

Comparing the noiseless codebook with the 0.1 BER noisy codebook, it may be noted that for a noiseless channel the difference in the resulting Segmental SNR between the two codebooks is only 1.5 dB while over a 0.1 BER channel the noisy codebook gains 3.3 dB. Subjectively, the slight loss in performance when operating over the noiseless channel is more than recovered when the bit-error rate increases.



Legend: -·-·-·- 0.00 BER; .... 0.01 BER; ------- 0.04 BER; —x—x— 0.10 BER; ——— 0.20 BER

Figure 4. Robustness of coder to errored residual

SPECTRAL CODING

It is necessary to define a measure by which a spectral coding scheme can be evaluated. One popular measure of spectral distortion is the mean unweighted SD, defined as

$$SD = \sqrt{\frac{1}{\pi}\int_0^{\pi}\left(10 \cdot \log\frac{S(w)}{\hat{S}(w)}\right)^2 dw} \tag{3}$$

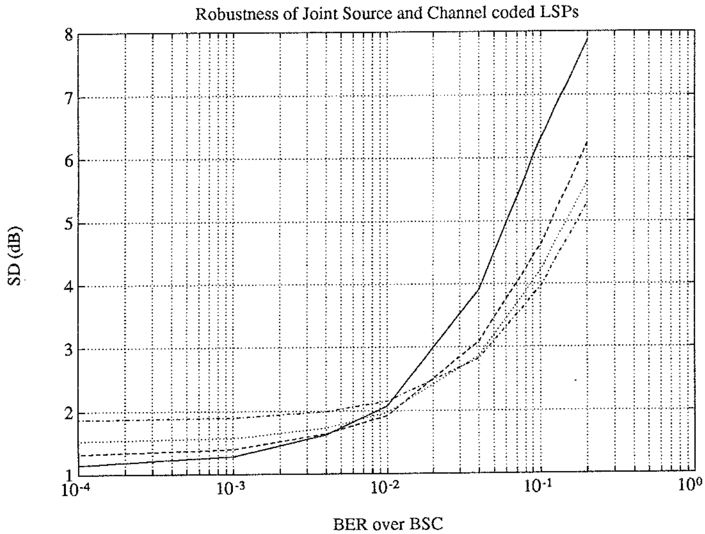
where $S(w)$ is the power spectrum evaluated from the unquantized LPC coefficients and $\hat{S}(w)$ is the power spectrum evaluated from the quantized LPC coefficients. The mean SD is evaluated across the database. It is generally accepted that a SD of 1 dB indicates neglible distortion has incurred in the quantization process.

The use of LSPs to code LPC coefficients is attractive due to their confined dynamic range and filter stability properties. A cascaded trellis structure is proposed whereby 10 (LPC order) trellises are connected with each trellis corresponding to a particular LSP. Thus the bit rate was 1 bit per LSP plus K-1 bits to specify a starting node. The sub-optimal squared error measure of (4) was used as the LSP quantization distortion. A check for strictly increasing LSPs is made at each pruning stage of the Viterbi algorithm.

$$d(w_i, \hat{w}_i) = (w_i - \hat{w}_i)^2 \tag{4}$$

Preliminary tests showed that the resulting spectral distortion was too high to be of practical use. Increasing the number of bits per LSP was accomplished by installing 2^L reconstruction values in a list on each branch. A value of L=2 was proposed resulting in 3 bits per LSP. Thus for 10th order analysis, 30 bits plus 3 bits (K-1) per frame were used for spectral information coding. This is in line with the number of bits used in current scalar LSP coding methods. Training of the structure was accomplished with four Harvard sentences (2 male + 2 female) totalling 1500 frames. The SD for a 250 frame male-spoken sentence outside the training sentences was 1.14 dB.

Channel-optimised codebooks were designed for BERs of 0.01, 0.04 and 0.1 using the same training sequence. Each codebook was used to quantize the spectral parameters of the same test sentence with each channel sequence subjected to BERs of 0, 0.001, 0.004, 0.01, 0.04 and 0.1 . Figure 5 demonstrates that each noisy codebook excels at its designed channel BER. Subjectively, using informal listening, the quality at high channel BERs for the noisy codebooks is significantly better than the noiseless codebook. This is a consequence of a large reduction in the number of unstable LPC filters resulting after transmission with the channel optimized codebooks. As with residual coding, there appears to be a desired BER (0.04 and 0.1) to design for if a range of channel error probabilities are expected.



Legend:
——— 0.00 BER
------- 0.01 BER
······· 0.04 BER
-·-·-·- 0.10 BER

Figure 5. Robustness of spectral coefficients

CONCLUSION

Incorporating joint source and channel trellis coding into narrow to medium-band speech coders offers effective robustness to high bit-error rates without added redundancy. Performance when mismatching the codebook to the channel is acceptable, making this technique attractive for use in coding systems operating with non-stationary channels.

ACKNOWLEDGEMENTS

Thanks is offered to the Wollongong University Postgraduate Awards Centre.

REFERENCES

Atal, Bishnu S. (1982) *Predictive Coding of Speech at Low Bit Rates*, IEEE Trans. Commun., vol. COM-30, pp 600-614.

Ayanoglu, E. & Gray R. (1987) *The design of Joint Source and Channel Trellis Waveform Coders*, IEEE Trans. Inform. Theory, vol. IT-33, pp 855-865.

Soong F.K. & Juang B.H. (1984) *Line Spectrum Pairs (LSP) and speech data compression*, Proc. ICASSP.

Stewart L. C. & Gray R. (1986) *The design of Trellis Waveform Coders*, IEEE Trans. Commun., vol. COM-34, pp 1073-1080.

Svendsen T. (1985) *Multi-Dimensional Quantization Applied to Predictive Coding of Speech*, PhD dissertation, Norwegian Institute of Technology