MAXIMUM A POSTERIORI DECODING FOR
SPEECH CODEC PARAMETERS

W.N. Farrell and W.G. Cowley

Institute for Telecommunications Research
Department of Electronic Engineering
University of South Australia

email : wade@spri.levels.unisa.edu.au

ABSTRACT – This paper looks at applying MAP decoding to low rate speech codec parameters as a means of protection at low channel SNRs. It has been shown that MAP decoding works well in protecting LSPs but the method has not been applied to other parameters. By using theoretical source data, results are obtained that compare MAP decoding with other more conventional techniques.

INTRODUCTION

When considering the use of a parametric low rate speech codec for a speech transmission system, the normal method of design involves using separate source and channel coding stages. The design goal is thus to compress the source information as much as possible and then apply a good channel codec to that information, given all the system and channel constraints. This philosophy is valid if all of the redundancy is removed from the source information. Unfortunately rarely does all redundancy be removed from the source information, especially when considering speech codec parameters.

An alternative method is to leave the residual redundancy present in the coded speech information and use the redundancy at the channel decoder. In this way all of the redundancy can be used for channel protection instead of trying to extract it by using complicated quantisation techniques.

The aim for such a channel decoder is to utilise all information about the signal, which involves both channel information and prior information about the source. This paper provides a formal analysis using Bayesian probabilities to describe a maximum a posteriori probability (MAP) decoder. It also explores how such a decoder can be used to protect low rate speech codec information. Two simple test cases are examined using Gaussian and Gauss–Markov source parameters.

MAXIMUM A–POSTERIORI DECODING

To construct a decoder, a model of a communication system is required for analysis. The model in Figure 1 is a simplified case, consisting of a speech codec providing speech parameter value X at time index n. This is channel coded and modulated to create symbol U_n. This symbol is passed through the channel, where noise is added to the symbol and then received as V_n ($V_n = U_n + Noise$). This paper assumes the use of an Additive White Gaussian Noise channel. After decoding, an estimate of the original value is obtained. This scalar case is used to simplify the analysis, however sequences are explored later.

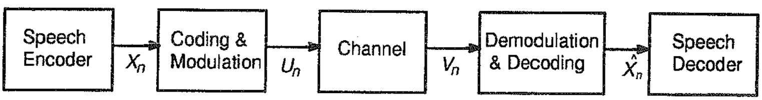

Figure 1 – Simple Speech Communication Model

Using the MAP model above, the estimator in equation (1), as described in [1], states that the source value X_n can be estimated as the expected value of X_n given the received value of V_n.

$$\hat{X}_n = E(X_n | V_n) \tag{1}$$

This estimator can be realised by performing a search over the maximum a posteriori (MAP) probability density function of X_n given V_n, for all possible values of X_n, as shown in equation (2).

$$\hat{X}_n = \arg\max_{X_n} p(X_n | V_n) \tag{2}$$

The estimator in equation (2) can be expanded using Bayesian probability shown in equation (3). This equation shows that the posterior pdf, $p(X_n|V_n)$ which is used as the decoder metric is made up of both the prior source pdf, $p(X_n)$ and also the pdf due to the channel noise $p(V_n|X_n)$.

$$p(X_n | V_n) = \frac{p(V_n|X_n)\ .\ p(X_n)}{p(V_n)} \tag{3}$$

Since we only require a metric for comparison over all values of X_n, $p(V_n)$ can be removed, since it is fixed over all values of X_n. If U_n is a direct mapping of X_n, with probabilities $P(X_n) = P(U_n)$, then the pdf $p(V_n|X_n) = p(V_n|U_n)$. Thus from the original estimator in equations (1) and (2) and by using the bayesian probability equation we obtain the MAP estimator, in equation (4).

$$\hat{X}_n = \arg\max_{X_n} p(V_n|U_n)\ p(X_n) \tag{4}$$

The MAP estimator shows us that the prior pdf can be just as important as the channel pdf. Their importance is scaled by how much information they carry about the source value X_n. If the prior pdf is uniform then it carries no information about the source. Conversely if the SNR of the channel is very low, then the signal carries little information about the source and the prior pdf should be used to help decode the source value. However it is the signal that is passed through the channel that always carries the information about the individual value of the source. The prior pdf usually provides only average information about each value. To implement the MAP estimator it is more efficient to take the natural log of the pdfs, as shown in equation (5).

$$\hat{X}_n = \arg\max_{X_n} (\ln p(V_n|U_n) + \ln p(X_n)) \tag{5}$$

An estimator which does not consider the prior pdf of the source, but still maximises the likelihood based on the channel information is called a Maximum Likelihood (ML) decoder, as shown in equation (6). This type of estimator assumes that the source pdf is close to uniform and thus cannot provide extra information that will help in estimation. There are many source coders which compress and transform the information enough to produce nearly uniform source pdfs. Vector quantisers are a good example of this. In such a situation the ML decoder will perform the same as the MAP decoder. However if redundancy remains in the source, which is usually the case for some speech parameters, then a MAP decoder should outperform the ML decoder.

$$\hat{X}_n = \arg\max_{X_n} p(V_n|U_n) \tag{6}$$

Sequence Estimation

The estimator in equation (4) is for a single value of X at time index n. Thus only instantaneous decoders can use this type of estimation. Unfortunately the instantaneous MAP estimator does not extend automatically to using a source vector, due to multiple integrations [1]. However a new metric (7) can be declared in which the posterior pdf for the vector parameter $\mathbf{X}_n$ of length L is maximised.

$$\hat{\mathbf{X}}_n = \arg\max_{\mathbf{X}_n} p(\mathbf{X}_n | \mathbf{V}_n) \tag{7}$$

Again the posterior pdf can be expanded to comprise of prior and channel pdf information, as provided in equation (8).

$$\hat{\mathbf{X}}_n = \arg\max_{\mathbf{X}_n} p(\mathbf{V}_n|\mathbf{U}_n) p(\mathbf{X}_n) \tag{8}$$

IMPLEMENTING MAP DECODING

Redundancy in a source parameter can be in either probability distribution and/or time correlation between parameters. If the distribution of a parameter is non–uniform, redundancy is present which can be utilised. Due to the frame nature of low rate speech codecs, time correlation can either consist of intraframe correlation (between parameters inside a frame) or interframe correlation (between parameter values of successive frames). A previous paper [2] utilised the intraframe correlation of line spectral pairs in constructing a TCM decoder. Alajaji et al [3] has implemented MAP decoding of LSPs using both intraframe and interframe correlation when using Reed–Solomon and convolutional codes. Secker [4] has also investigated optimising transmission of speech parameters by using channel–optimised trellis source coding. This paper focuses on methods of utilising interframe correlation for single value per frame parameters and provides examples using Gaussian and Gauss–Markov sources.

To implement MAP decoding, a pdf of the source parameter is required. This pdf may describe only the current sample i.e. $p(X_n)$. However dependencies can be included into the pdf to provide greater information, by either using previous values in time $p(X_n | X_{n-1} X_{n-2})$ or using other parameters $p(X_n | Y_n Z_n)$. Each dependency adds another dimension of the pdf, thus increasing the size of the pdf.

PDF Estimation

Consider a memoryless, non–uniformly distributed parameter X. We wish to find the pdf of X, $p(X)$, to satisfy the MAP estimator in (4). If the pdf is known or can be approximated, then this function can be used directly in the MAP metric for decoding.

If the pdf is unknown, which is much more common, it can be estimated by observing the parameter values over a large sample space, assuming the parameter is a stationary random variable. A histogram can be constructed by using training data to provide a sequence of parameter values of length L. In the case of speech parameters, a large speech training database is required, which provides a good variety of speakers. To obtain a histogram, the range of X is first divided into C number of cells, with each cell being Δx wide.

$$\Delta x = \frac{Xmax - Xmin}{C} \tag{9}$$

By traversing the values of X, the number of X values falling in each cell is counted, producing B_i, for each cell i, $i=1..C$. Dividing each B_i by the total number of X values N, provides the estimated probability of X lying in cell i, $P(x_i)$. However we do not want to use probabilities but work with pdfs. Thus by dividing this value by the width of the cell Δx, provides the estimated pdf value $\hat{p}(x_i)$.

$$\hat{p}(x_i) = \frac{B_i}{N \cdot \Delta x} \tag{10}$$

The result is an estimated pdf which consists of discrete components as an approximation of the true continuous probability density function, as shown in the example of Figure 2. Each estimated pdf value is the average value over the range of the cell. More memory is required in storing histogram information than using a pdf function, but using a histogram can provide more accurate information than a function match.

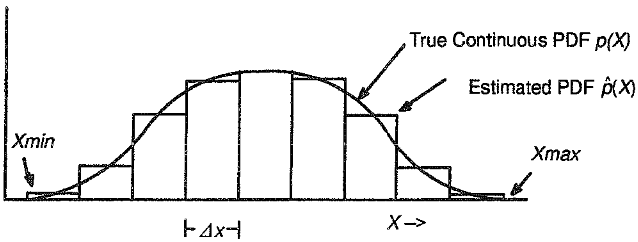

Figure 2 – Example of estimating a parameters pdf

If digital modulation is used, the metric must be calculated for each possible codeword, thus requiring a source pdf value $p(X_n)$ for each codeword. For optimum MAP decoding, each cell in the histogram must be matched to the digital modulation levels, i.e. M (number of modulation levels) = C (number of cells). If this is not possible due to certain constraints, it is possible to interpolate the histogram points to provide more cells.

When the source pdf has dependencies there are two approaches. The first is to compare every possible value in the pdf at the decoder. This is an exhaustive search which provides an optimum result, given the limits of the original pdf. If the pdf is very large, due to many dependencies such a decoder may not be achievable due to computational restraints. The other method is to cement some of the dependencies which will reduce the search size. Cementing is especially suitable when previous values have already been utilised and will not effect the system is a better value is found. However cementing provides a sub-optimum result as it assumes that the dependencies that have been chosen are correct, which may not be true.

EXAMPLE USING MEMORYLESS PARAMETERS

We wish to test the effectiveness of using a MAP decoder over a ML decoder when using a memoryless non–uniformly distributed parameter. Consider a memoryless, Gaussian distributed source parameter which is modulated using 2 symbols of QPSK modulation, providing 4 bits of quantisation. Either method of providing the pdf, i.e. using a histogram or pdf function, provides the same results. The decoder, in making its decision for each value, compares the 16 possible values of the source value, using the MAP metric.

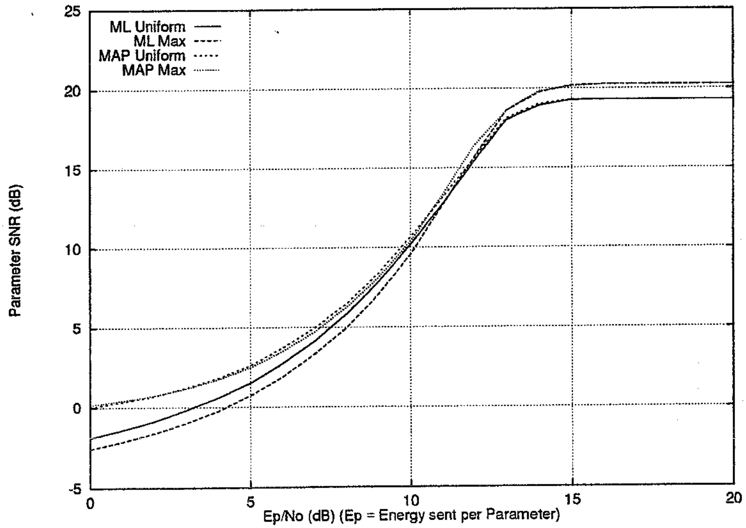

Figure 3 – Comparison between MAP and ML decoding for memoryless Gaussian Source

Results are shown in Figure 3 which include ML and MAP decoding using both uniform and Max quantisation [5] of the source. In applying uniform quantisation to the Gaussian source, different quantisation ranges were tested, with the best using a maximum truncation value of 2.5 σ_X. Max quantisation is a non–uniform scalar quantisation method which optimises the quantisation levels to maximise the resolution for a given source pdf. Parameter SNR is used as the performance measure against Ep/No = 2 x Es/No (where Es = Energy per modulation symbol and No is the noise density of the channel) since 2 symbols are transmitted per parameter value. The Max quantisation uses some of the redundancy in the distribution to gain lower quantisation error and achieve higher asymptotic performance. Due to the low amount of redundancy in the source parameter, MAP decoding only achieves a small gain in performance over the ML decoders. The MAP decoder with Max quantisation performs the best of all four schemes, implying that there is still redundancy remaining after Max quantisation.

EXAMPLE USING PARAMETERS WITH MEMORY

To test the ability of using time redundancy in a MAP decoder, the same setup was performed using a Gauss–Markov parameter with coefficient = 0.9, as in equation (11). Such a parameter has much greater redundancy than a memoryless parameter and thus more gain is available to the MAP decoder.

$$X_n = 0.9\, X_{n-1} + W_n \qquad \text{where } W_n \sim N(0, \sigma_w^2) \tag{11}$$

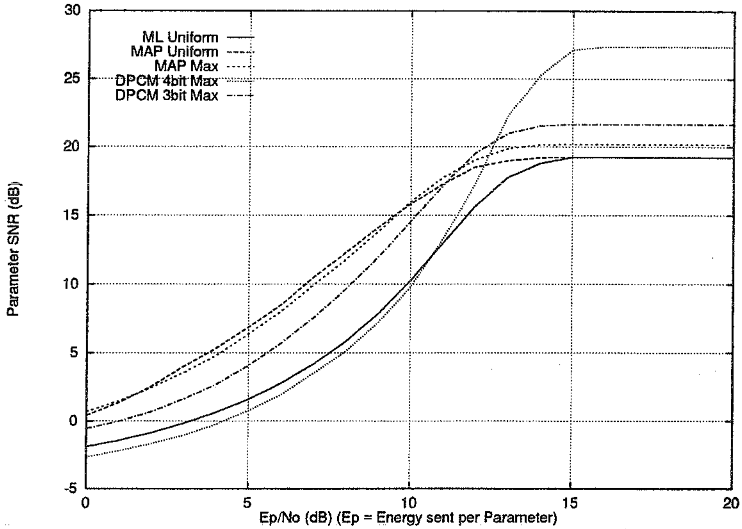

Figure 4 – Comparison between ML and MAP Decoding of Gauss–Markov Parameter

Figure 4 shows the performance of five different schemes. The ML decoder with uniform quantisation has the worst performance. When MAP decoding is applied, the asymptotic performance remains the same, however up to 3dB gain is achieved at lower SNR, showing the extra performance in using source redundancy at the decoder.

When Max quantisation is added to the MAP decoder, the asymptotic performance is improved but a slight reduction in performance at lower SNR due to the removal of some of the redundancy. This shows the trade off in extracting redundancy to obtain greater resolution compared with using the redundancy at the decoder. Applying a Max quantiser only removes redundancy in the magnitude, it does not consider redundancy in time. It also designs the quantisation levels on the premise of error free transmission.

Since a Gauss–Markov parameter has considerable correlation it is possible to use this at the source coder to provide better resolution. There are many techniques available which compress source information, such as vector quantisation and trellis codes. However only a simple scheme was chosen for comparison. Differential PCM (DPCM) [5] works by estimating the sample from previous samples and transmitting the difference D_n in equation (12).

$$D_n = X_n - \hat{X}_n \qquad \text{where } \hat{X}_n = 0.9\, X_{n-1} \tag{12}$$

Such a scheme improves asymptotic performance by providing greater resolution. However DPCM does not provide a good code structure for transmitting through a channel as there is considerable dependency between transmitted values, making it susceptible to errors. As a result, the number of bits for the source was reduced to 3 with the MSB sent as a BPSK symbol to provide greater protection. The result shows reduced asymptotic performance but a more robust scheme at lower SNR. Above 11.5 dB the 3 bit DPCM outperforms the MAP Max scheme. It depends on the distribution of channel SNR which scheme is better for a given transmission system.

USING SPEECH PARAMETERS

The Gauss–Markov example provides a good example for some of the speech codec parameters that retain large amounts of correlation between successive values. A good example of such a parameter is the gain parameter which is usually present codecs that use linear prediction, such as CELP and some sinusoidal codecs. The gain parameter, as it is highly proportional to the speech amplitude, is usually slow to change. It is also can be dependent on other parameters of the speech codec, making it have a time variant pdf. For instance, the gain exhibits a different pdf during silence than during speech. Also it tends to have difference characteristics during voiced speech compared with unvoiced speech. All of these dependencies make it difficult to apply MAP decoding to a speech transmission system. However with time spent on determining dependencies and developing coding and modulation schemes, high improvements in performance at lower channel SNRs are achievable.

REFERENCES

[1] S.M. Kay, (1993) *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice Hall International, Inc.

[2] W.N.Farrell, W.G.Cowley, (1994) *A Rate 3/4 TCM Decoder for Line Spectral Pairs Using MAP Information*, SST–94, Perth Dec 1994.

[3] F. Alajaji, N. Phamdo & T. Fuja, *Channel Codes That Exploit the Residual Redundancy in CELP–Encoded Speech*, Submitted to IEEE Transactions on Speech and Audio Processing.

[4] P. Secker, *Channel–Optimised Trellis Source Coding for the Memoryless Gaussian Channel and Applications to Spectral Parameter Coding*, Ph.D. Thesis

[5] N.S. Jayant & P Noll, (1984) *Digital Coding of Waveforms*, Prentice–Hall, Inc. New Jersey