A Robust Error Masking Hybrid Spectral Quantisation Scheme for Noisy Channels

J. Kostogiannis and A. Perkis*

Department of Electrical and Computer Engineering, University of Wollongong

*Norwegian Institute of Technology, Division of Communications

ABSRACT - In this paper a robust error masking hybrid spectral quantisation scheme, based on Line Spectral Pairs (LSPs), critical for perceptual speech quality in noisy channels is presented. Its performance, in comparison to conventional schemes, is evaluated considering error free conditions and in the presence of up to 10% random errors. In addition the inherent structures of the LSPs are utilised in designing non-redundant techniques for error detection and error masking incorporated in the quantisation scheme.

INTRODUCTION

An important class of voice coders, based on linear prediction, have been shown to be capable of producing high quality speech at bit rates as low as 4 kbit/s. These techniques have been used in vocoders, producing synthetic speech for message systems around 2.4 kbit/s. An integral feature of such coders is the representation of the short term spectral information of the speech signal. In the Code Excited Linear Predictive (CELP) (Atal, 1985) based coders this information is found by fitting an AR model to a short speech segment, resulting in a parametric representation of the speech signal in terms of a set of filter coefficients $\{a_k\}$, describing the all pole filter $1/A(z)$. This information has been shown to be far more sensitive to bit errors (Perkis & Ribbum, 1990), and great care should be taken in finding a robust representation of these parameters before transmission.

Several new applications have emerged, incorporating low bit rate voice coding, with mobile communications being one of the most important of these. All new proposed mobile communication schemes will be digital and operate in narrow bands, requiring some form of speech compression. Of the cellular services, the PAN European scheme employs a 13.1 kbit/s Residual Pulse Excited Linear Predictive (RPE-LTP) speech coder (GSM 06.10, 1989), while the digital cellular mobile communication scheme proposed for the US will be using a 7.95 kbit/s Vector Sum Excited Linear Predictive (VSELP) speech coder (EIA 2215, 1989). For satellite mobile communications, speech coder robustness is much more of a concern, due to occurrences of comparatively poorer channel conditions than for the land mobile systems. The design methodology, will however, in principle be the same for the two scenarios.

All these voice codecs are in some way dependent on modelling and representing the short term speech spectrum in an efficient and robust way. Line Spectrum Pairs are capable of reducing the bit rate by 25% (compared to other scalar encoding methods, eg. Log Area Ratios) for quantising the spectral information without any degradation to the speech (Soong & Juang, 1984). A perceptually important property of LSPs is that they introduce the possibility of checking for synthesis filter instability which has a great impact on the degradation in speech quality. Filter stability is conserved by ensuring the LSPs are sequentially ordered. A more desirable transformation of the LSP frequencies is its differences (LSPDs). It is less susceptible to speaker characteristics, as well as recording and analysis conditions. The problem associated with the LSPDs is that they are much more sensitive to random bit errors as compared to LSPs. High and low pitched resonances are predominant in coders using LSPDs over noisy channels (evidence of shifting in the speech spectrum). Noisy channels and to a lesser extent quantisation distort the spectral information, causing irritable bangs and squeaks in the synthesised speech (Kostogiannis & Perkis, 1990). The hybrid spectral quantisation scheme presented here incorporates features of both LSPs and LSPDs and is suitable candidate for quantisation of spectral information over noisy channels.

Section 2 presents the quantisation scheme compared to conventional LSP and LSPD schemes, while the implemented error detection and error masking capabilities are described in Section 3.

A HYBRID SPECTRAL QUANTISATION SCHEME

A robust hybrid spectral quantisation scheme combining both the LSPs and LSPDs is less sensitive to random errors (an inherent property of LSPs) and allows more efficient scalar encoding of the spectral information (a property of LSPDs). In conventional LSPD quantisation schemes the LSP frequency differences are transmitted as the spectral parameters (Kang & Fransen, 1984). If an error occurs in a LSP frequency difference, it is propagated throughout the spectrum. That is the speech spectrum once reconstructed at the decoder is either shifted to higher or lower frequencies. Whereas an error in a LSP frequency affects the prediction filter's spectrum near that frequency.

In the hybrid spectral quantisation scheme a decrease in the sensitivity to random errors is achieved by strategically replacing three LSPDs with LSPs in perceptibly important locations. Random errors are then limited to propagate only within these LSP intervals, thereby decreasing the degradation in the synthesised speech spectrum. The encoding strategy of the hybrid spectral quantisation scheme is presented below;

(i) quantise ω_i to $\hat{\omega}_i$ and set i=1.

(ii) form the difference $\Delta\omega_i = \omega_{i+1} - \hat{\omega}_i$

(iii) quantise $\Delta\omega_i$ to $\Delta\hat{\omega}_i$

(iv) reconstruct $\hat{\omega}_{i+1}$ where $\hat{\omega}_{i+1} = \hat{\omega}_i + \Delta\hat{\omega}_i$

(v) go to (ii) & increment i, if i = 4,7 & 10 go to (i) and then to (v), end after 10, the order of filter.

where $\hat{\omega}_i$ represents the quantised value of ω_i.

Figure 1 illustrates the performance of the hybrid scheme as compared to the conventional LSP and LSPD schemes. All the schemes are scalar encoded to 34 bits per frame of 30 msec speech. A speech database consisting of 24 seconds of Australian male and female utterances (2 speakers for each sex) is used to compare the performances of the schemes.

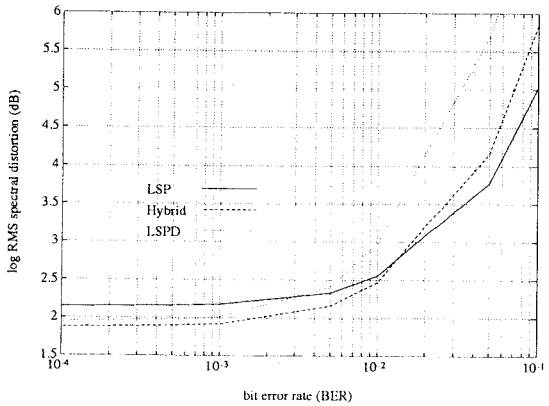

Figure 1 Objective measures of the quantisation schemes.

The hybrid spectral quantisation scheme and the conventional LSP and LSPD quantisation schemes perform acceptable in error free conditions. This is supported by log RMS spectral distortion measures (Schuyler, Barnwell & Clements, 1988) of less than 2 dB. The hybrid scheme out-performs both conventional schemes at bit error rates (BER) ranging from 0.01% to 1%. At BER above 1% the hybrid scheme is out-performed by the conventional LSP scheme but is infinitely better than the LSPD scheme. These results are reinforced subjectively by paired comparison tests conducted using ten listeners.

ERROR DETECTION AND ERROR MASKING

Two techniques for error detection and error masking have been developed, all based on the time correlation in LSP frequencies and the filter stability criteria. These techniques have all been implemented in conjunction with the hybrid quantisation schemes.

The first method is a substitution technique, whereby if an error is detected, the offending LSP frequency, ω_j, can be substituted with the previous frame's frequency. The error is detected by checking for the sequential ordering property of the LSP set. The ordering property, using the hybrid scheme, can only be violated at the strategically placed LSP locations, as between them lie LSPDs. The offending pair of LSP frequencies (one, a strategically placed LSP, ω_j, and the other a reconstructed LSP frequency, ω_{j-1}, from the previous two LSPDs and LSP frequencies) can be detected if $\omega_j - \omega_{j-1} < 0$. Now by measuring the deviation between these LSP frequencies, ω_{j-1} and ω_j, and their previous frame's frequencies, ω_{j-1}' and ω_j', one can safely predict which frequency has been corrupted and substitute accordingly. Let

$$d_1 = \omega_j - \omega_j' \text{ and } d_2 = \omega_{j-1} - \omega_{j-1}' \qquad (1)$$

If $d_1 > d_2$ then ω_j is replaced by ω_j' or alternatively if $d_1 < d_2$, then ω_{j-1} is replaced by a frequency determined by the following strategy. An assumption is made that there can only be one error between strategically placed LSPs. Therefore by measuring the deviations between the LSPD frequencies and the respective previous frame's LSPD frequencies, the offending LSPD frequency is determined and substituted with its previous frame's frequency. If the test fails, ω_j, ω_{j-1} and ω_{j-2} are replaced by their respective previous frame's frequencies.

The substitution method is non-optimum because not all errors can be detected by the simple ordering rule. In some cases, errors in the LSP frequencies (reconstructed from the hybrid scheme) will still preserve monotonicity, but cluster close to adjacent frequencies, creating unwanted spectral peaks. These spectral peaks cause large squeaks and bangs in the synthesised speech, found to be very annoying to the listener. This problem is addressed by introducing a "smart" filter stability correction algorithm involving adaptive bandwidth expansion on the LSP frequencies.

The "smart" filter stability correction algorithm sorts the LSP frequencies in ascending order and then checks for closeness. The close LSP frequencies are expanded by a certain bandwidth expansion (adaptive), depending on their location (first three by 40 Hz, the next five by 140 Hz and the rest by 200 Hz).

The techniques described above are all techniques which mask out any large bangs and squeaks in the synthesised speech during high BER (1-2%). Overall the synthesised speech using these techniques at abnormal high BER (>10%) is perceptibly pleasant and not painful to the ear as with conventional LSP filter stability correction algorithms (just sorting the LSP frequencies). However, errors are not corrected, as with redundant techniques like FEC, but are masked out (subjectively negating their effect) by utilising certain properties of the LSP frequencies. These techniques are implemented at a cost of introducing slight distortion, due to the violation of continuity introduced by information substitution. Therefore the techniques should only be enabled in noisy channel environments, and disabled during good conditions.

Objective results comparing the two methods are given in Figure 2.

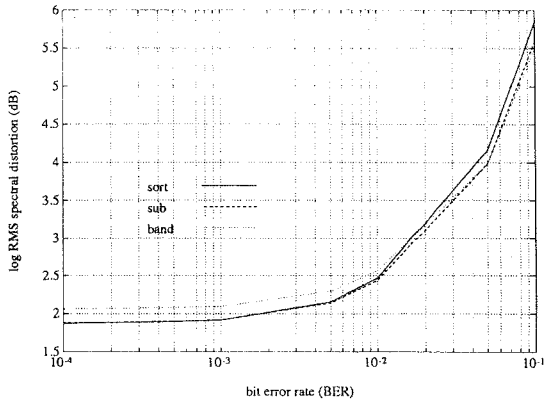

Figure 2 Objective evaluation of error masking techniques for the Hybrid scheme.

In Figure 2, the two error masking algorithms are compared to the standard sorting technique. The results show that for severe error conditions all the scheme out-perform the standard sorting technique, and that the two techniques have a similar performance. Subjectively though, there is a difference in performance. At a BER of 1%, large squeaks are apparent when only sorting is used. The substitution method masks out most of these, but still some errors are not detected. The adaptive bandwidth expansion masks out the remaining errors, replacing them by a more pleasant distortion.

The techniques have been evaluated in a formal subjective test confirming the informal listening results. Figure 2 also indicates the need for disabling the error masking techniques during good conditions, by noting that the adaptive bandwidth expansion technique gives a higher spectral distortion measure than the sorting technique for low BERs. Based on the figure, the substitution can be enabled at a BER of 0.5% and the bandwidth expansion technique at a BER of 1%.

CONCLUSION

In this paper a robust error masking hybrid spectral quantisation scheme incorporating properties of Line Spectrum Pairs (LSPs) and their differences (LSPDs), critical for perceptual speech quality in noisy channels is presented. It evaluates and compares its performance with the conventional LSP and LSPD schemes, considering error free conditions and in the presence of up to 10% random errors. At a BER of 0.1%, objective measures show an improvement of 0.3 dB and 0.1 dB, in logarithmic RMS spectral distortion, as compared to the conventional LSP and LSPD methods. At a BER of 1%, there is an improvement of 0.05 dB and 0.5 dB, compared to the conventional LSP and LSPD schemes.

Finally the spectral parameters represented by LSP frequencies are highly correlated with respect to time. In most cases each frequency differs minutely as compared to its respective frequency in the next frame. This correlation, in addition to the LSPs' inherent requirements for filter stability (observing LSP frequency monotonicity), are attributes that can be used to localise the effect of transmission errors in the frequencies, and provide scope for non-redundant techniques for error detection and error masking. The subjective evaluations, by paired comparison listening tests, show that speech processed with the non-redundant techniques enabled and transmitted over noisy channels, are always preferred over the speech with just the sorting algorithm.

REFERENCES

Kang, G.S., Fransen, L.J. (1984) *Application of line spectrum pairs to low bit rate speech encoders*, Navel Research Laboratory report.

Atal, B.S. (1985) *Code-Excited Linear Prediction (CELP): High Quality Speech at Very Low Bit Rates*, Proc. ICASSP.

Perkis, A., Ribbum, B. (1990) *Application of stochastic coding schemes in satellite communications: Advances in speech coding*, (Kluwer Academic publishers, New York).

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

Schuyler, R.Q., Barnwell, P.T., Clements, M.A. (1988) *Objective measures for speech quality*, (Prentice-Hall, New Jersey).

Kostogiannis, J., Perkis, A. (1990) *Evaluation of Linear Prediction Schemes for use in Mobile Satellite Communications*, SST-90, Melbourne.

GSM recommendation 06.10, (1989) *GSM full rate speech transcoding.*

EIA project number 2215, (1989) *Cellular system, Dual mode Subscriber equipment network, equipment compatibility specification*, IS 54.

AN 8 kb/s LD-CELP WITH IMPROVED EXCITATION MODELLING

R. Soheili, A.M. Kondoz, B.G. Evans

Centre for Satellite Engineering Research
University of Surrey, U.K.

Abstract

Backward prediction of the short term redundancies in speech has resulted in very low delay algorithms, with toll quality at 16 kb/s. At medium bit rates around 8 kb/s the modelling of the excitation signal by conventional CELP techniques can result in high complexity or poor output processed speech for services such as PSTN. In this paper we propose a low delay algorithm based on a vector quantised multi-tap adaptive codebook in producing a high quality speech signal operating at 8 kb/s. A report on the comparisons with other existing standards as well as simplification techniques in realising the algorithm are presented.

1 INTRODUCTION

Interests in low bit rate speech coding for Public Switch Telephone Networks (PSTN) has taken a sharp rise. This is mainly due to the increase in the number of users and limited bandwidth available. It is hoped that with the introduction of the new standards, namely ADPCM [1], and LD-CELP [2], running at 32 Kb/s and 16 Kb/s respectively, the congestion will be reduced considerably. However demand for the use of the PSTN is rising at a higher rate than the available channel capacity, which has forced CCITT to standardize an 8 Kb/s low delay speech coder as soon as possible. Therefore interest is now being centered around medium and low delay coders running at 8 Kb/s and below. Unlike forward adaptation schemes the LD-CELP is based on a buffering of 5 samples at 8 kHz sampling rate. Frequent update of the vector quantisation of the excitation signal and a high order backward LPC synthesis filter are the basis of the LD-CELP algorithm. In order to reduce the bit rate to 8 kb/s, predictions of long term as well as short term correlations present in the speech signals, plus vector quantisation of the speech parameters are necessary to reach the system requirements. It is widely believed that [3], the future CCITT standard will be a conventional CELP algorithm [4], with the exception of backward prediction of the short term correlation rather than conventional forward techniques. However such a modelling of the excitation signal will result in large number of parameters to be encoded, thus resulting in a high transmission of side parameters. Since certain parameters are not correlated then individual coding of these parameters would result in a high bit rate system or poor output processed quality due to inaccurate modelling with fewer bits. In this paper we propose a modelling which is much simpler but updated more frequently, hence resulting in a high quality processed speech. The algorithm is based on a multi tap long term predictor "LTP",