Masking of Errors in Transmission of VAPC-Coded Speech

Neil B. Cox and Edwin L. Froese
M.P.R. Teltech Ltd.
8999 Nelson Way
Burnaby, B.C., Canada, V5A 4B5
Phone: (604) 294-1471
FAX: (604) 293-5787

1. ABSTRACT

This study provides a subjective evaluation of the bit-error sensitivity of the message elements of a Vector Adaptive Predictive (VAPC) speech coder, along with an indication of the amenability of these elements to a popular error-masking strategy (cross-frame hold-over). As expected, a wide range of bit-error sensitivity was observed. The most sensitive message components were the short-term spectral information and the most significant bits of the pitch and gain indices. The cross-frame hold-over strategy was found to be useful for pitch and gain information, but it was *not* beneficial for the spectral information unless severe corruption had occurred.

2. INTRODUCTION

Application-specific information can often be exploited in the design of error-control methodologies for dedicated communication channels. While a concession is made to the generality of the system when such information is used, there are practical applications for which this concession is acceptable. One such application is speech transmission over mobile satellite channels. Here there are four sources of application-specific information: the channel characteristics, the speech coding format, predictable characteristics of the speech signal and the relative importance of signal components in speech perception. With the possible exception of channel characteristics, these options are not exploited if error control is based solely on general-purpose error correction codes.

The following factors should be considered for efficient control of transmission errors in VAPC-encoded speech:

1. The error-free delivery of all message bits is not required for meaningful speech communication, as human listeners are remarkably adept at inferring meaning from context. This implies that the goal of error control should be to reduce the *perceptual* effect of errors.
2. The short-term predictability of speech provides a variety of intuitive approaches to error compensation, such as adaptive smoothing or cross-frame hold-over of parameters.[1,9] While much of the effort in speech coding is devoted to the removal of this predictability, the coding algorithms generally update their parameters at a high enough rate to adequately represent the signal during its most transient conditions. Thus, residual predictability can be expected for a considerable proportion of the speech sequence.
3. The bits of a coded speech message have a widely-varying influence on the perceived speech quality. Ordered parameters are naturally comprised of bits with varying significance. Some parameters are interrelated or dependent on past samples, leading to a propagation of the errors within a frame and across frames. Certain parameters represent fundamental aspects of speech, whereas others only refine the quality.

Methods of accounting for the varying importance of message bits have been proposed in the literature. Numerous examples can be found where error detection and/or correction is applied to a subset of the message bits.[1,9] The parallel application of codes has been used to further concentrate the protection on the most important bits.[10] Rate-compatible punctured convolutional codes provide for selective allocation of code power without the need to switch between coders.[5] All of these approaches require a rank-ordering of message bits. Based in part on informal listening tests, it is common to leave residual information unprotected for linear prediction coders (LPC) and sub-band coders.[9, 10] It has been reported for the basic LPC-10 approach that the critical bits are the most significant bits of the first three or four prediction coefficients along with the most significant bits of the gain, pitch and voicing parameters.[1] The more complex LPC approaches are not directly comparable, as the encoding introduces dependence between parameters and between frames. Nonetheless, it is generally observed that residual information is less sensitive to bit errors than gain, pitch or spectral information.

The purpose of this study is to provide data on the bit-error sensitivity of the message elements in a Vector Adaptive Predictive (VAPC) speech coder. Existing information on this topic is sparse and has generally been acquired in an informal fashion. The sensitivity of each message element to random errors is addressed, along with relative merit of holding-over preceding message elements when errors are present. The evaluation was performed for a random error model and a 2-state Markov simulation of burst errors. The results provide useful guidance in the design of efficient error control techniques for VAPC-encoded speech.

3. VECTOR-ADAPTIVE PREDICTIVE CODING

The VAPC encoding algorithm is illustrated in Figure 1. Briefly, the speech waveform is passed to the encoder in 20 msec frames. The pitch-period is determined using a bounded search for the autocorrelation peak. A 3-tap linear pitch predictor is used to remove signal components that are related to pitch. The prediction points are separated from the predicted point by one pitch-period. This is followed by a 10-th order linear predictive inverse filter that models the spectral envelope. Gain information is derived from the output of the two filters. Finally, residual vectors are selected to minimize the difference between the input signal and a locally-synthesized output. This analysis-by-synthesis approach partially compensates for errors that result from quantization of the pitch, spectral and gain information. Further detail can be found elsewhere.[4, 6, 12]

The codec evaluated in this study has the following bit allocation. The pitch-period index (*idxp*) is a 7-bit linear quantization of the pitch-period. The pitch prediction vector index (*idxpp*) uses 6 bits to select a pitch prediction filter from a codebook of 64 candidates. The selected predictor provides the largest reduction of signal energy. The LSP error indices (*idxsp*) and the classification index (*idxcl*) are a complex relative representation of the short-term spectral information, where the *idxsp* are scalar quantizations of the difference between computed Line Spectrum Pair (LSP) coefficients and values predicted from the previous frame through first-order linear vector prediction, and *idxcl* chooses one of four sets of coefficients for the vector predictor. This consumes 29 bits. The residual gain index (*idxg*) is a 6-bit logarithmic quantization of the residual energy. Finally, the sixteen 7-bit residual vector indices (*idxr*) are a multi-stage vector quantization of the excitation signal that minimizes the analysis-by-synthesis error.

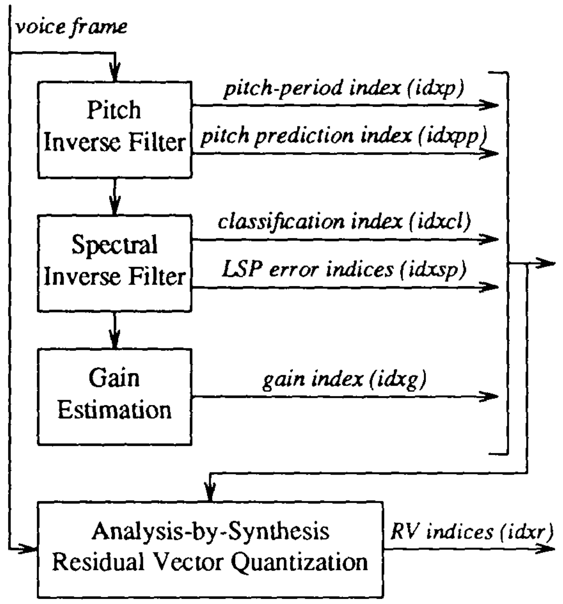
voice frame
Pitch Inverse Filter
pitch-period index (idxp)
pitch prediction index (idxpp)
Spectral Inverse Filter
classification index (idxcl)
LSP error indices (idxsp)
Gain Estimation
gain index (idxg)
Analysis-by-Synthesis Residual Vector Quantization
RV indices (idxr)

Figure 1: Structure of a Generic VAPC Encoder

4. EXPERIMENTAL SETUP

The codec described above has been implemented for real-time operation on a single 20 mHz Motorola DSP56001 signal processor. Sufficient time is available to also perform adaptive echo cancellation if the 27 mHz version of the chip is used. In fact, a single-chip real-time 2-channel implementation has been developed by excluding the echo cancellation and post filtering, and shortening the residual vector codebooks.

A locally-developed test bed for codec evaluation was used for this study. This test bed provides flexible synchronization, communication and data input/output among general-purpose DSP56001-based processing cards. A separate program can be downloaded to each card and interactively modified when necessary. Three cards were used for the present study. Two of the cards received the VAPC encoder/decoder program and the third card received an error imposition program.

The error imposition program is capable of imposing random errors on specified groups of bits within a frame, or optionally holding over preceding indices instead of imposing the errors. The bit-error probability is adjusted by a thumbwheel switch while the program is running. A special "decoder hold-over mode" was implemented for the short-term spectral information. Here the LPC parameters derived within the decoder are held over whenever an error is imposed on the classification index or the LSP error indices. By using a derived parameter set, the effects of a bit error are spread to a number of the parameters, thus making it advisable to hold-over the entire LPC parameter set. Due to their relative encoding format, the transmitted spectral indices are not themselves suitable for cross-frame hold-over.

Two models for the bit errors were implemented; a random error model and a 2-state Markov model. The first model simply applies random errors to the data at the rate specified by the thumbwheel switch. The two states for the Markov model are random error models. At the start of each frame, one of the states is selected based on pre-specified state transition probabilities. The test configuration was set up so that an error-free state was chosen 90% of the time. That is, both rows of the state transition matrix were (0.9 0.1). The error rate for the "bad" state is determined by the thumbwheel switch. This simulates a channel with bursty errors, and loosely conforms to data described in an AUSSAT codec test procedure.[11]

5. ASSESSMENT METHODOLOGY

The subjective assessment methodology is similar to the Degradation Category Rating (DCR) procedure.[2] This is a pair-wise comparison procedure where the listeners are asked to judge the degradation of the second sample of each pair relative to the first. The following 5-point degradation scale was used:

5 = no degradation
4 = slightly annoying
3 = annoying
2 = very annoying
1 = extremely annoying

Eight listeners were seated in a quiet room and provided with written instructions about the format of the test. The listeners were not allowed to discuss or compare judgements. The test samples were presented over a high quality audio system. Twelve "practice" samples were presented at the start of the test in order to familiarize the listeners with the task and to expose them to the range of degradations that they will encounter. These judgements were excluded from subsequent analyses.

The test stimuli were recorded in random order on a test tape as a series of A-B pairs, where A is a voice sample that is passed through the codec without imposed errors, and B is the same sample with some form of imposed degradation. Two such A-B pairs were included for each test configuration. The voice sample for one of the pairs was a female reciting "The navy attacked the big task force, see the cat glaring at the scared mouse". The voice sample for the other pair was a male reciting "March the soldiers past the next hill, a cup of sugar makes sweet fudge". These sentences are from the phonetically-balanced Harvard sentence collection. Finally, six "null pairs" (A-A) were included to test the anchoring of the listeners' assessments.

6. TEST CONFIGURATIONS

The bit-error probabilities for the test configurations are summarized in Table 1 and Table 2. These error levels were derived in an informal preliminary trial to produce degradation ranging from slight to severe. The implied frame-error probability (i.e. the probability of one or more bit errors within a frame) is included in brackets. In order to keep the test trials at a manageable length, single-bit evaluations were only performed for the ordered indices (*idxp* and *idxg*), and each *idxr* and *idxsp* index was not individually tested.

Both the random error mode and the random hold-over mode were tested for all but the single-bit configurations. In addition, the "full dropout" condition (i.e. BER=0.5) was evaluated for the Markov model. A total of 158 test pairs were tested, comprised of 60 random error pairs, 16 Markov error pairs and 3 anchoring (null degradation) pairs for each voice sample.

Table 1: Bit-error probabilities for random error test configurations. The degradation was judged in an informal preliminary trial. "n" = P_{error} for each bit, and "(n)" = the implied P_{error} for each frame. Bit 1 is the least significant.

Errored	Degradation		
Index	low	mid	high
idxg bit 1	.1	.2	.5
idxg bit 3	.05	.1	.2
idxg bit 5	.02	.05	.1
idxp bit 1	.1	.2	.5
idxp bit 3	.05	.1	.2
idxp bit 5	.02	.05	.1
idxr	.02 (.997)	.05 (≈1)	.1 (≈1)
idxg	.02 (.11)	.05 (.26)	.1 (.47)
idxp	.02 (.13)	.05 (.30)	.1 (.52)
idxpp	.02 (.26)	.05 (.47)	.1 (.74)
idxp+idxpp	.02 (.23)	.05 (.49)	.1 (.75)
idxcl+idxsp	.001 (.03)	.005 (.13)	.01 (.25)
all	.001 (.15)	.005 (.55)	.01 (.80)

Table 2: Bit-error probabilities for the "bad" state in Markov error test configurations. Here an average of 90% of the frames are error-free, and the remaining frames have the random bit-error probability designated below. The degradation was judged in an informal preliminary trial. "n" = P_{error} for each bit in a "bad" frame, and "(n)" = the implied P_{error} for each "bad" frame.

Errored	Degradation		
Index	low	mid	high
idxcl+idxsp	.01 (.25)	.02 (.44)	.05 (.77)
all	.01 (.80)	.02 (.96)	.05 (≈1)

7. RESULTS

The means of the degradation scores for all test configurations are summarized in Tables 3 through 7. Each mean was derived from 16 judgements (i.e. 8 listeners and 2 samples per listener). The average variance was approximately 0.5 for these judgements. Based on a one-tailed Students' *t*, this implies that differences of greater than 0.6 are significant at the 1% level, and differences of greater than 0.4 are significant at the 5% level. The mean of the degradation for the null pairs was 4.9, indicating that the judgements were well anchored.

Table 3: Degradation MOS for random errors in bits of the gain index *idxg*. Bit 1 is the least significant.

Errored	Bit-error probability				
bit	.02	.05	.1	.2	.5
1	---	---	4.9	5.0	4.8
3	---	4.9	4.8	4.6	---
5	3.3	2.2	1.7	---	---

Table 4: Degradation MOS for random errors in bits of the pitch-period index *idxp*. Bit 1 is the least significant.

Errored	Bit-error probability				
bit	.02	.05	.1	.2	.5
1	---	---	4.5	4.9	4.3
3	---	4.2	3.4	3.1	---
5	4.8	3.5	2.5	---	---

The single-bit conditions summarized in Table 3 and Table 4 demonstrate the expected relationship between bit-error sensitivity and bit significance for *idxg* and *idxp*. For *idxg* there was a sudden onset of severe degradation; corruption of bit 1 or bit 3 had little effect, but corruption of bit 5 caused severe degradation. This is partially explained by the logarithmic quantization of this index. The onset of degradation was more gradual for *idxp*.

The following observations can be drawn from the single-index conditions summarized in Table 5:

— Corruption of the residual vector indices (*idxr*) caused a moderate level of degradation for the tested bit-error rates. Cross-index hold-over provided a statistically significant reduction of the bit-error sensitivity, but notable degradation was still present.

— Corruption of the gain index (*idxg*) caused severe degradation at all tested error levels. The cross-frame hold-over strategy provided a large improvement, with only moderate degradation produced by the worst error rate (BER=0.1).

— The pitch-period index (*idxp*) was relatively sensitive to bit errors. Fortunately, as with the gain index, cross-frame hold-over provided a significant improvement.

— The pitch prediction index (*idxpp*) was relatively insensitive to bit errors, and no significant improvement was obtained from cross-frame hold-over. Furthermore, there appears to be no significant interaction between *idxp* and *idxpp* in terms of the bit-error sensitivity, as corruption of both indices has approximately the same effect as corruption of *idxp* alone.

Table 5: Degradation MOS for random errors in the pitch, gain and residual indices. Data are for random bit-errors, and random index hold-over in response to such errors.

	Random Errors			Random Hold-Over		
Errored	Bit-error probability			Bit-error probability		
Index	.02	.05	.1	.02	.05	.1
idxr	3.4	2.7	2.2	4.0	3.7	2.5
idxg	2.3	1.7	1.5	4.6	3.4	3.3
idxp	2.9	2.6	1.6	4.0	3.3	2.6
idxpp	4.4	3.6	2.8	4.6	3.8	3.1
[*p+pp*]	2.9	2.6	1.9	4.3	2.8	2.4

Table 6: Degradation MOS for random errors in spectral indices (*idxcl*+*idxsp*) and all indices. Data are for random bit-errors, and random index hold-over in response to such errors. The "decoder hold-over mode" was used for the spectral indices.

	Random Errors			Random Hold-Over		
Errored	Bit-error probability			Bit-error probability		
Index	.001	.005	.01	.001	.005	.01
[*sp*+*cl*]	4.1	2.9	2.6	3.9	3.0	1.9
all	3.6	2.2	1.6	4.4	2.3	1.9

Table 7: Degradation MOS for bursty errors in spectral indices (*idxcl*+*idxsp*) and all indices. A Markov error model is used, where an average of 90% of the frames are error-free, and the remaining frames have the random bit-error probability designated below. Data are for random bit-errors, and random index hold-over in response to such errors. The "decoder hold-over mode" was used for the spectral indices.

	Markov Errors				Markov Hold-Over			
Errored	Bit-error probability				Bit-error probability			
Index	.01	.02	.05	.5	.01	.02	.05	.5
[*sp*+*cl*]	4.0	3.1	3.7	1.3	3.9	3.2	3.2	3.3
all	4.3	3.7	2.8	1.0	3.6	3.7	3.5	2.8

When one considers that the bit error rates in Table 6 and Table 7 are 10 times less than those in Table 5, it is clear that the spectral indices (*idxcl* and *idxsp*) are by far the most sensitive to bit errors. Except for the "full dropout" condition (BER = 0.5) in the Markov error simulation, the cross-frame hold-over strategy did not improve the situation, and produced a significant *degradation* at a random bit-error rate of 0.01. Thus, the hold-over strategy should only be counted on when data transmission is severely compromised. This view is supported by the "all indices" data in these Tables, as a significant improvement was only provided when severe corruption was present. The one exception (random errors at a bit-error rate of 0.001) may be due to the shortness of the speech samples, as few bit-error combinations are encountered at low error rates.

There was a wide diversity in the quality of the perceived error effects. Corruption of *idxr* caused "garbling" of the speech but did not produce an alarming disturbance. Errors in the gain index, on the other hand, tended to impose intermittent and extremely loud bursts. The spectral errors caused intermittent alarming "whoops" and "squawks", that is, the disturbances were very loud and irritating, and appeared to have an entirely inappropriate frequency content. Finally, corruption of the pitch indices had the expected effect of introducing a hoarse quality to the speech, with intermittent abnormal jumps in pitch.

8. DISCUSSION

A general conclusion of this study is that most of the effort in error control should be devoted to protection of the short-term spectral information (*idxcl* and *idxsp*), with attention also given to the most significant bits of the gain index (*idxg*) and the pitch-period index (*idxp*). The spectral parameters were followed in importance by the gain index (*idxg*), the pitch-period index (*idxp*), the residual vector indices (*idxr*) and the pitch prediction index (*idxpp*). Errors in the three least significant bits of the pitch and gain indices (*idxp* and *idxg*) had little perceived effect. Also, there is little reason to protect the pitch prediction index if the residual vector indices are left unprotected, as the degradation caused by corruption of *idxr* is relatively severe before corruption of *idxpp* becomes noticeable.

If a moderate degradation is acceptable at bit-error rates of 0.05 or more, then the practice of leaving residual vector indices unprotected is justified. A comprehensive error correction protocol requires excessive redundancy, as the RV indices comprise the majority of the bits of the message. The lack of a natural ordering for the residual vectors makes it difficult to rank order the message bits, thus ruling out bit-selective strategies. While this study indicates that some improvement can be obtained by using a cross-index hold-over strategy, this requires a coding method with sufficient power to localize the error(s) to specific indices.

It is recommended that global application of cross-frame hold-over should only be relied upon in burst error conditions, at least for the short-term spectral information. Here the main advantage is in preventing extreme and highly irritating signal disturbances. However, the hold-over strategy was beneficial for the gain index (*idxg*), the pitch-period index (*idxp*) and to a lesser extent the residual vector indices (*idxr*).

Other methods of error masking may be beneficial to augment or replace the hold-over strategy. For example, progressive muting of the output during bursts has been recommended.[1,9] Both linear and non-linear approaches can be used to derive estimates of corrupted parameters based on past history. Cross-frame hold-over is a special case of this. Other examples are linear extrapolation and median filtering.

Running estimates of the probability distribution or other statistics of parameters would be useful in accounting for context-dependent effects. The parameters used in such an analysis can be taken from any stage of the decoder. The use of "sped-up speech" in combination with automatic repeat request (ARQ) protocols has been proposed for bursty channels.[8] The bursty speech that results may be less annoying than the disturbance associated with the other strategies.

Index assignment optimization methods have been proposed for error masking.[3,7] Here a measure of the effect of an error is assumed, and the indices are assigned such that the most probable error patterns produce the smallest effects. Such strategies are attractive in that they are simply implemented and do not require added redundancy or added run-time computation. Unfortunately, a number of factors argue against their success. For example, mathematical measures of error effects have not demonstrated a good correlation with the actual perceived effect. Even if the measure is accurate, most parameters of speech are highly nonstationary, so an optimized index allocation based on a fixed statistical model may well be inferior in many conditions. Nonetheless, this approach may be beneficial in situations where other strategies are not practical, such as for protecting residual vector indices.

The small size of this study limits the general applicability of the results. We have limited ourselves to random errors imposed on short, albeit phonetically-balanced, samples of speech passed through a single VAPC codec. It is recognized that the length of the sample is undoubtedly insufficient for thorough testing of all speech contexts, particularly at low error rates. Limiting the experiment to two English-language speakers neglects numerous external factors, such as age, health, linguistic background, habitual pitch, etc.. The effects of changing the codec configuration or the input amplitude were not tested. Finally, the diversity of perceived effects caused by various bit-errors makes it potentially misleading to use a single opinion score as a basis for comparison.

It is recognized that the sound reproduction and listening environment were of a higher quality than can reasonably be expected in most applications. This method of test presentation facilitates the detection of subtle degradations and makes it easier to concentrate throughout the test. An informal verification of the presentation format was performed, where one listener repeated the test on a different day using a standard telephone handset. As expected, there was a reduced ability to detect subtle degradations over the handset, and the severely degraded samples were not as alarming. The variance of the difference between the two sets of judgements from this listener was approximately the same as the average variance of the audio-speaker-based assessments across listeners. A complete assessment of this issue requires simulation of the range of receiving apparatus and noise environments.

REFERENCES

[1] **Bryden, B., Seguin, G.E., Conan, J., Bhargava, V.L., and Brind'Amour, A.**, 1989. "Error Correction/Masking for Digital Voice Transmission Over the Land Mobile Satellite System." *IEEE Trans. on Commun..* pp. 309-314.

[2] **CCITT,,** 1988. "Subjective Performance Assessment of Digital Processes using the Modulated Noise Reverence Unit (MNRU)." *CCITT Blue Book, Vol. V, Supplement 14.* pp. 341-360.

[3] **Chen, J.H., Davidson, G., Gersho, A., and Zeger, K.**, 1987. "Speech Coding for the Mobile Satellite Experiment." *IEEE Int. Conf. on Commun..* pp. 756-763.

[4] **Chen, J.H. and Gersho, A.**, 1987. "Real-Time Vector APC Speech Coding at 4800 bps with Adaptive Postfiltering." *IEEE Int. Conf. on ASSP.* pp. 2185-2188.

[5] **Cox, R.V., Hagenauer, J., Seshadri, N., and Sundberg, C.E.**, 1988. "A Sub-Band Coder Designed for Combined Source and Channel Coding." *IEEE Int. Conf. on ASSP.* pp. 235-238.

[6] **Davidson, G. and Gersho, A.**, 1988. "Multiple-Stage Vector Excitation Coding of Speech Waveforms." *IEEE Int. Conf. on ASSP.* pp. 2185-2188.

[7] **DeMarca, J.R.B. and Jayant, N.S.**, 1987. "An Algorithm for Assigning Binary Indices to the Codevectors of a Multi-dimensional Quantizer." *IEEE Int. Conf. on Commun..* pp. 1128-1132.

[8] **Lynch, J.T.**, 1987. "Evaluation of the Comprehension of Noncontinuous Sped-up Vocoded speech: A Strategy for Coiping with Fading HF Channels." *IEEE J. on Selected Areas in Commun. 5.* pp. 308-311.

[9] **McLaughlin, M.J. and Rasky, P.D.**, 1988. "Speech and Channel Coding for Digital Land-Mobile Radio." *IEEE J. on Selected Areas in Commun. 6.* pp. 332-345.

[10] **Suda, H. and Miki, T.**, 1988. "An Error Protected 16 kbit/s Voice Transmission for Land Mobile Radio Channel." *IEEE J. on Selected Areas in Commun. 6.* pp. 346-352.

[11] **Wilkinson, M.H.**, 1990. *Voice Codec Test and Evaluation Procedure VCTEP/2.* Telecom Research Laboratories; Australian Telecommunication Corporation

[12] **Yong, M., Davidson, G., and Gersho, A.**, 1988. "Encoding of LPC Spectral Parameters using Switched-Adaptive Interframe Vector Prediction." *IEEE Int. Conf. on ASSP.* pp. 402-405.