Performance of a Low Data Rate Speech Codec for Land-Mobile Satellite Communications

Allen Gersho
Voicecraft, Inc.
815 Volante Place
Goleta, California 93117 USA
Phone: (805) 683-2800
FAX: (805) 964-5893

Thomas C. Jedrey
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California 91109 USA
Phone: (818) 354-5187
FAX: (818) 354-6825

ABSTRACT

In an effort to foster the development of new technologies for the emerging land-mobile satellite communications services, JPL funded tow development contracts in 1984: one to the University of California, Santa Barbara (UCSB), and the second to the Georgia Institute of Technology, to develop algorithms and real-time hardware for near-toll quality speech compression at 4800 bits per second. Both universities have developed and delivered speech codecs to JPL, and the UCSB codec has been extensively tested by JPL in a variety of experimental setups. The basic UCSB speech codec algorithms and the test results of the various experiments performed with this codec are presented in this paper.

INTRODUCTION

Over the past several years, a significant amount of research and development in the area of low bit rate (4800 bits per second) speech coding has taken place. As a result of this research, the emerging land-mobile satellite communications services will in all likelihood use these codecs to provide voice communications. In an effort to accelerate the development of these codecs, JPL funded two development contracts in 1983 with the University of California, Santa Barbara (UCSB), and the Georgia Institute of Technology to develop the necessary algorithms and real-time hardware for near toll quality speech codecs at 4800 bits per second.

As a result of these contracts, several speech codecs were developed and delivered to JPL for use in the NASA Mobile Satellite Experiment (MSAT-X) Program. These codecs have been integrated into the MSAT-X land-mobile satellite communication terminal, and the UCSB codec has been extensively tested in environments ranging from a simulated satellite (a 1000 foot tower), to a full scale land-mobile satellite channel. In addition to these tests, the UCSB codec has been independently tested by the US Department of Defense [1].

The UCSB speech codec algorithms and test results from the various experiments performed are presented in this paper. Techniques employed in the codec to mitigate the effects of channel errors will be stressed, including frame synchronization and frame repeat strategies. Results from both the aeronautical and land-mobile experiments will be presented.

USCB SPEECH CODEC

Three candidate algorithms were identified at UCSB for the MSAT-X application. Of these three, two algorithms Vector Adaptive Predictive Coding (VAPC) and Pulse Vector Excitation Coding (PVXC) were chosen for hardware implementation. The final algorithm selected for use in the MSAT-X testing was the VAPC algorithm, and all test results and further discussions in this paper are restricted to this algorithm [2].

The VAPC algorithm encodes and decodes telephony bandwidth speech sampled at 8 kHz. The resulting speech at a cumulative data rate of 64 kHz is analyzed without frame overlap at 22.5 ms intervals. As discussed below, the VAPC algorithm is based extensively on the use of vector quantization, a powerful generic technique for efficient coding of sets of parameters that characterize attributes of speech. With vector quantization, a relatively short binary word is often sufficient for accurately specifying the amplitude of a large number of parameter values, or waveform samples needed for reproducing speech sounds at the receiver.

In speech coding below 16 kb/s, one of the most successful scalar coding schemes is Adaptive Predictive Coding (APC) developed by Atal and Schroeder [3]. It is the combined power of vector quantization and APC that led to the development of VAPC.

The basic idea of APC is to first remove the redundancy in speech waveforms using adaptive linear predictors, and then to quantize the prediction residual using a scalar quantizer. In VAPC, the scalar quantizer is replaced with a vector quantizer. The motivation for using the vector quantizer was two-fold. First, although linear dependency between adjacent speech samples is essentially removed by linear prediction, adjacent samples may still have a nonlinear dependency which can be exploited by vector quantization. Secondly, vector quantization can operate at rates below one bit per sample. This is not achievable with scalar quantization, but is essential for speech coding at low bit rates.

VAPC Structure

The basic structure of an early version of VAPC, shown in Figure 1, is quite similar to that of conventional APC. In the transmitter, the redundancy due to pitch quasi-periodicity is first removed by a long delay predictor, or "pitch predictor". A short delay predictor is then used to remove the short term redundancy remaining in the pitch-prediction residual, and the final residual is quantized by a gain-adaptive vector quantizer. In the receiver, the speech waveform is reconstructed by exciting two cascaded synthesis filters with the quantized prediction residual.

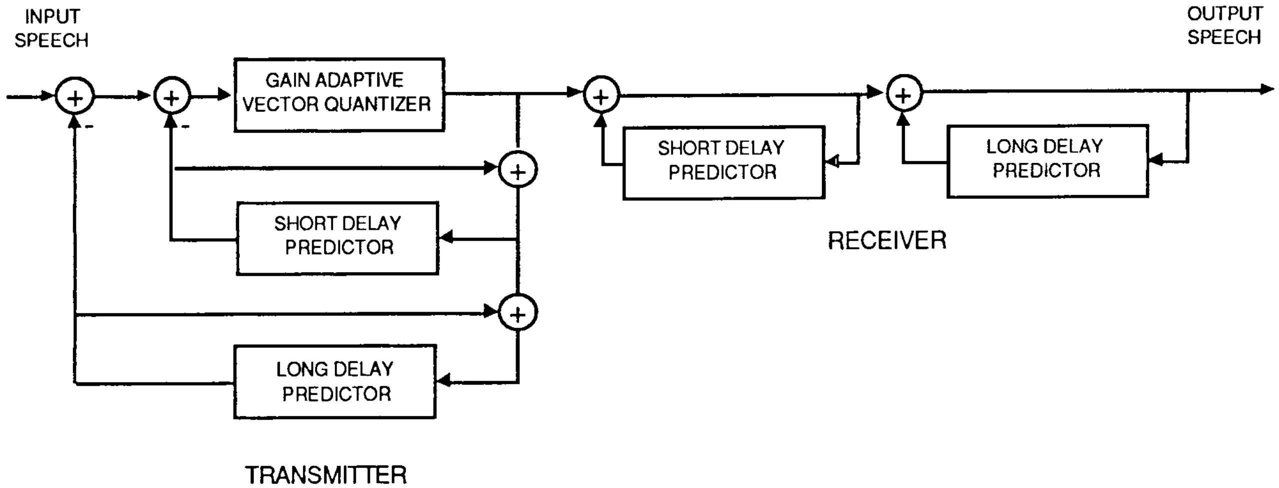

Figure 1 Basic Structure of VAPC

The structure shown in Figure 1, was modified to produce the efficient structure shown in Figure 2. To encode each vector of speech samples, the pitch prediction residual vector is generated, passed through a perceptual weighting filter, and the zero input response vectors are subtracted from it. The resulting vector is then compared with the N stored zero-state response vectors. The index of the nearest neighbor is then used to extract the corresponding vectors in the vector quantization codebook. This codevector is then used to excite the LPC synthesis filter to generate data for use in pitch prediction of the subsequent speech vectors.

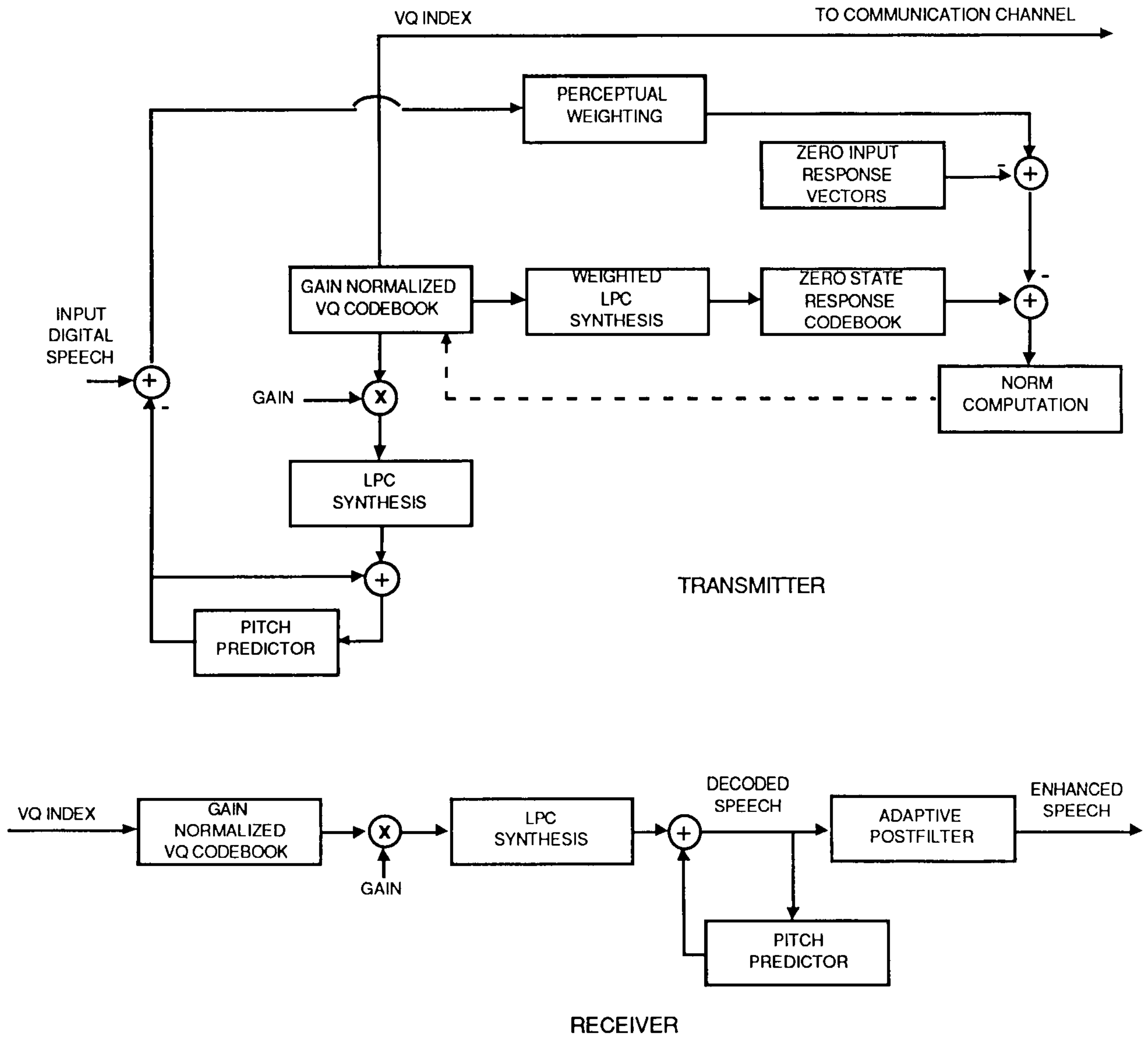

Figure 2 VAPC Transmitter and Receiver

The vector quantization codebook is designed as the gain-normalized codebook of a forward gain-adaptive vector quantizer. The normalized vector quantization codebook is fixed, while the zero-response codebook changes from speech frame to speech frame.

To further improve the perceptual quality of the coded speech, a novel adaptive post-filtering technique was developed that greatly reduces the perceived level of coding noise without introducing significant distortion in the filtered speech [2].

VAPC Channel Optimization

Once the basic algorithm was fixed, optimization of the VAPC codec for the communication channel was considered. The specified channel was a bursty channel with an average bit error rate of 10^{-3}. Several techniques for combatting the channel effects were considered and implemented, including frame synchronization, pseudo-Gray coding, error detection, and frame repeat strategies. However, prior to implementing any error detection/mitigation strategies, the VAPC algorithm was tested in the presence of bit errors from a simulated satellite channel. The results indicated that the basic algorithms were relatively insensitive to isolated errors and even to moderated bursts of errors, depending on the locations of the errors.

As mentioned above the basic VAPC algorithm frame length is 22.5 ms. This corresponds to a 108 bit frame. Of these 108 bits four bits were allocated for frame synchronization and error detection (more bits could have been allocated, however this would have reduced the quality of the coded speech). This translates into an overhead rate of 200 bits per second for link maintenance. Based on the low number of bits allocated per frame for this purpose, it was decided to minimize the number of bits used for frame synchronization (based on the constraint that the received data is initially synchronized) and to restrict the remaining bits (three) to error detection.

In the case of frame synchronization, there were several issues to be considered, including proper detection of an out-of-synch frame, and proper re-synchronization of a frame once the out-of-synch condition has been detected. In addition to these issues, there is the requirement that the reframing time be kept to a minimum. Based on a tradeoff between acceptable reframing time (one second), detection of the out-of-synch condition, proper resynchronization once the out of synch condition has been detected (versus false detection), the desire to keep the link maintenance overhead at a minimum, and computational complexity, an alternating pattern of ones and zeroes was chosen for the synchronization pattern.

The out-of-synch condition is declared by the codec when the received synchronization pattern over an eight frame history differs from an alternating pattern by more the a single error or two consecutive errors. When the out-of-synch condition is declared, the speech decoder produces silence until the in-synch condition is declared.

Once the codec enters the resynchronization state, a pattern matching algorithm is implemented to detect the alternating synch pattern, and this algorithm operates until a sufficient number of synchronization bits (7 out of 8 bits) are correctly received.

Based on the above algorithms and the channel statistics, it has been computed that the minimum time to detect the out-of-synch condition is three frames, and the probability of non-detection of the out-of synch condition after eight frames is approximately 6%. The average resynch time is estimated to be approximately 8 frames.

In the case of error control, as mentioned above, three bits per frame are allocated for error detection. Given the limited number of bits per frame allocated for this purpose (driven by the speech quality constraint), only the most critical channel errors are addressed: that of burst errors (a mitigation strategy for isolated errors - pseudo-Gray coding is discussed below). To that end, the three bits in each frame are designated as parity bits jointly covering the speech frame. A 108 bit frame is divided into three words as follows. The first word is formed by concatenating the first frame bit with every third bit after it. The second and third words are formed in a similar fashion. The parity bits are then chosen to force the three words formed to have even parity. The probability that an error burst goes undetected is then approximately the probability that an even number of errors occurs in each parity word (i.e., approximately 13%). Although this probability is relatively high, the experimental results on the robustness of the VAPC algorithm in the presence of errors in the absence of any error protection indicate that it is quite robust to isolated errors. When the parity bits do not check, the previous speech frame is repeated if the number of consecutive repetitions is below two, otherwise silence is played until the error burst ends. Experimental results have shown that this frame repeat strategy significantly reduces the perceptual impact of error burst that last two frames or less.

Finally, a technique to mitigate the presence of isolated errors that involves no coding overhead, called pseudo-Gray coding has been studied and implemented. This technique involves assigning the binary indices to code vectors and codebook design so that isolated channel errors produce minimal perceptual errors (very similar to Gray coded QPSK). Simulation results with PCM on the binary symmetric channel with bit error rates between .01 and 10% have indicated a substantial gain of 2-4 dB in SNR, roughly uniform over the error probability range.

Combining the speech coding/decoding and the channel overhead, the overall complexity of the VAPC algorithm is approximately 4 million multiply/adds per second, and the algorithm requires approximately 8 kwords of RAM for fixed and variable data, and program storage. This algorithm is implemented using two DSP32's for the MSAT-X program. It has also been implemented on a single Motorola 56000 DSP chip at Voicecraft, and at Microtel Pacific Research (with appropriate support chips in both cases).

CODEC TESTING

The speech codec testing consisted of laboratory tests at JPL and UCSB, quantitative tests by the US Department of Defense, field tests by JPL in various environments, and quantitative tests by the Australian TELECOM. The qualitative test results from the JPL field tests and the quantitative results from the US Department of Defense and Australian TELECOM tests are discussed below.

US Department of Defense Testing

The final version of the VAPC algorithm was evaluated by the US Department of Defense [1] in 1988 as part of a very extensive and thorough study of 4800 bit per second speech codecs. The testing program consisted of subjective evaluations of quality under a variety of operating conditions. Subjective ratings were made using the Diagnostic Rhyme Test (DRT) and the Diagnostic Acceptability Measure (DAM). The DRT test measures the ability to distinguish between pairs of rhyming words, and is a measure of the intelligibility of the speech. The DAM test uses complete sentences and listeners judge various quality attributes that lead to an overall measure of speech quality. Clearly, in terms of user acceptability, the DAM scores are the most important, while in cases where intelligibility is of prime concern (e.g., air traffic control) the DRT scores are the most important.

As a result of these tests, the VAPC algorithm was found to have the highest DAM scores (of the seven different codecs that underwent detailed testing) for quiet speech (no background noise), office speech (typical office background noise), speech through a carbon microphone, and a noisy aircraft environment. Under the quiet background noise environment, the VAPC algorithm received a DAM score of 65.5. In comparison, the LPC-10 2400 bit per second standard has a DAM score of 48 under the same conditions. VAPC ranked poorly in the helicopter noise, tandeming, and 1% bit error rate environments. A point to note is that the VAPC algorithm was not designed to operate in the latter environment. For the DRT tests, the VAPC algorithm tended to score somewhat lower, averaging around fifth out the seven codecs tested.

JPL Field Tests

The VAPC codec has been tested in three separate field tests by JPL. These tests range from a simulated satellite environment using a 1000 foot tower as the satellite simulator, to an aeronautical mobile experiment using the INMARSAT Marecs B2 satellite, to land-mobile satellite experiments in Australia in conjunction with AUSSAT, using the Japanese ETS-V satellite.

In all three field trials, the VAPC codec performed well, providing an intelligible, good quality voice link, through which the experimenters communicated between the mobile unit and the fixed ground station. All users of the speech link were impressed with the quality of the speech and the ability to identify to far end speaker (as compared to LPC-10).

During the aeronautical mobile experiment [4], the full-duplex voice link was established often and used as the main (in fact the only available) method for direct communication between the experimenters on the aircraft (an FAA Boeing 727 flying along the East Coast of the United States) and in the fixed ground station. These links were run routinely at the same signal to noise ratio that resulted in 10^{-3} BER. There was no perceptible difference in speech quality or intelligibility between in-flight and operation on the ground. Jet noise had no significant effect on the communications. A formal part of the experiment was the demonstration of the voice link for air traffic control applications. During one of the flights, an FAA engineer on-board the aircraft read a variety of air traffic control-type messages into the VAPC codec. The voice received at the ground station was assessed by FAA personnel and recorded. Live conversations were also recorded. The intelligibility and quality of the speech, and the robustness of the link, were deemed acceptable by the FAA staff. Remarkably, the audio output of the codec at the ground station, which was available on a headphone speaker, was acoustically (not electrically) patched to a telephone headset and through a long-distance line to an FAA listener attending a meeting in Montreal, Canada. The listener found the voice to be intelligible and its quality to be acceptable.

The last field test that the VAPC codec was tested in was the full scale land-mobile experiment conducted in Australia [5]. During this experiment, the satellite based speech link was used as the primary means of communicating between the mobile terminal and the fixed terminal (an available HF link provided at best, poor quality communications). The experimental performance of the codec was similar to that obtained in the previous two field tests and was dictated by the overall bit error rate performance of the mobile and fixed terminals. During the tests, several speech links were established and maintained over periods of two hours while the mobile terminal travelled the Australian countryside. This link was maintained even in the presence of heavy blockage. During these tests as well as the previous two tests a considerable number of voice tapes ranging from DRT and DAM tapes through live conversations were recorded. All users of the system were impressed by the quality of the speech. Indeed, interested parties at AUSSAT (the Australian national satellite systems providers) were very impressed by the performance of the codec, and rated the overall performance of the codec and terminal superior to the other analog and digital systems they were currently reviewing. As a result of these tests, the Australian land-mobile satellite system specification has been changed from an approach based to analog speech (ACSSB) to digital speech at approximately 5000 bits per second, such as that provided by the VAPC codec.

Australian TELECOM Testing

As part of the Australian experiment mentioned above, the MSAT-X modem and the VAPC codec were installed in the Australian TELECOM laboratories where a variety of channel tests were performed. These tests ranged from the codec/modem performance in the heavily shadowed Rician fading environment to a Rician fading environment (K=20). Of significant interest were the results of the codec/modem pair when compared to the performance of one of the best ACSSB modems available over the Rician fading environment. The overall performance was rated based on the Mean Opinion Score (MOS), a subjective measure of the overall quality of the received speech. The basic results were that the codec/modem pair had an average MOS of slightly better than 3.0 (on a 5.0 scale, with toll-quality speech rated at 4.2) for C/N0 values ranging from 45 dB-Hz to 56 dB-Hz. This MOS value fluctuated slightly over this range due to the sample sizes used in the experiment, but was approximately 3.1 at 45 dB-Hz, and at 56 dB-Hz. In comparison, the ACSSB modem achieved a MOS score of approximately 1.8 at 45 dB-Hz and 3.5 at 56 dB-Hz.

CONCLUSIONS

The development program for 4800 bit per second speech codecs under the MSAT-X program in that several different codecs have been developed that provide good quality specch at this data rate. Of particular note is the performance of the VAPC codec as described in this paper. This codec provides good quality speech at 4800 bits per second, and ranks well when compared to other codecs at the same data rate. A very important distinction between this codec and many of the other 4800 bit per second codecs is the required number of computations per unit time [1]. When compared with other codecs with the same level of computational complexity, the VAPC codec appears to be distinctly superior. In particular, the VAPC algorithm has less than half the complexity of the CELP algorithms tested by the US Department of Defense and appears to be the only one implemented with a single fixed point DSP chip.

Modifications of the VAPC algorithm have led to very high quality codecs at 8 and 16 kilobits/second and commercial liscenses of the algorithm have already been issued. In particular, Compression Labs, Inc. uses VAPC at 8 kbits/s for the audio signal in its low bit rate video codecs.

ACKNOWLEDGEMENTS

This work was performed at the University of California, Santa Barbara, and the Jet Propulsion Laboratory, California Institute of Technology, under contract to the National Aeronautics and Space Administration.

REFERENCES

[1] D. P. Kemp, R. A. Sueda, T. E. Tremain, "An Evaluation of 4800 BPS Voice Coders", Proceedings Intl. Conf. on Acoustics, Speech, and Signal Processing, pp. 200-203, 1989.

[2] J. H. Chen, A. Gersho, "Real-Time Vector APC Speech Coding at 4800 BPS Using Adaptive Postfiltering", Proceedings Intl. Conf. on Acoustics, Speech, and Signal Processing, pp. 2185-2188, 1987.

[3] B. S. Atal, M. R. Schroeder, "Predictive Coding of Speech Signals and Subjective Error Criteria", IEEE Trans. Acoustics, Speech, and Signal Processing, Vol. ASSP-27, June, 1979.

[4] T. C. Jedrey, K. I. Dessouky, N. E. Lay, "An Aeronautical Mobile Satellite Experiment", JPL Report in preparation.

[5] T. C. Jedrey, W. Rafferty, "The MSAT-X/AUSSAT Land-Mobile Satellite Experiment: An Overview", MSAT-X Quarterly No. 22, JPL 410-13-22, January 1990.