A SUB-BAND CODEC WITH EMBEDDED REED-SOLOMON CODING FOR MOBILE RADIO SPEECH COMMUNICATION

K.H.J. WONG, L. HANZO(*), R. STEELE

UNIVERSITY OF SOUTHAMPTON, ENGLAND, SO9 5NH

(*) also TELECOMM. RES. INSTITUTE,1025.BUDAPEST POB 15,HUNGARY

Abstract

The performance of a jointly optimised sub-band codec (SBC), channel codec, and post processing speech enhancement system for binary phase shift keying (BPSK) transmissions over Rayleigh fading channels is presented. Forced up-dates of the SBC quantizer step sizes, and step-size leakage algorithms were examined and it was found that for a channel BER of 10^{-2} the segmental SNR increased by approximately 10dB compared to the basic SBC. At high BERs forced up-dating was preferable. Systematic and non-systematic Reed Solomon (RS) coding and convolutional coding were employed. The RS codec was also embedded into the SBC, and speech enhancement based on template matching deployed. By using an embedded RS codec with speech enhancement the gain in segmental SNR over the robust SBC for BERs in excess of 10^{-2} was 7 dB.

1.Introduction

The sub-band codec (SBC) produces near toll quality speech at 16 kbit/s and can be implemented using a DSP chip. It is therefore an appropriate codec for the transmission of speech signals in mobile radio environments, with the proviso that suitable channel coding and speech enhancement methods are employed.

2.Sub-band Codec

In the particular SBC codec [1] we considered,the speech bandwidth was divided into eight sub-bands by a bank of FIR quadrature mirror filters (QMF) . The order of the FIR filters in the first, second and third stages of the QMF bank were 32, 16 and 12, respectively. The signals at the output of each sub-band were encoded using Jayant's one-word memory quantizer [2]. By observing the signals at the outputs of the QMF stages, i.e., by observing where the spectral energy resides, the SBC envoked one of four bit allocation strategies depending on whether voiced speech, unvoiced speech, voiced/unvoiced transitions, or voice band data were deemed to be present. The bit allocation deployed assigned a particular number of bits to the sub-band quantizers. For any assignment the bits allocated to a particular quantizer were selected according to perceptual criteria. The number of bits (B) allocated was either 0, 2, 3 or 4.

The bit allocation for the quantizer was reviewed every 6 ms, and the decision was transmitted to the receiving decoder. The 2-bit word defining the semi-adaptive bit allocation was repeated twice for error protection to yield a side information rate of 6 bits per 6 ms or 1 kb/s. Each sub-band was 500 Hz wide and was sampled at 1 kHz. Only six of the sub-bands spanning the range 0 to 3 kHz were utilised. For each input Nyquist speech sample, the six quantizers produced 15 bits and consequently the sub-band coded data was multiplexed at 15 kb/s. The total transmission rate was therefore 16 kb/s. As the bit allocation interval was 6 ms the frame or packet contained (6 x 15) + 6 = 96 bits.

Figure 1 shows the variation of segmental SNR, (SEG-SNR), as a function of relative input power and sub-band centre frequency for female speech. The lower sub-bands had a SNR of over 20 dB, whereas the higher sub-bands were recovered with an SEG-SNR of typically 3 dB. Notice that the SEG-SNR profiles were correlated to the typical variations found in the spectral energy distribution of speech, i.e., the formant structure.

In order to mitigate the effects of transmission errors in Jayant's adaptive quantiser [2] the previous step-size was modified by a leakage factor b, viz:

$$s(i) = G\ M\ s^{b}(i-1) \qquad (1)$$

where s(i) and s(i-1) were the step-sizes at the i-th and (i-1)-th sampling instants, M was the multiplicative factor

that depended upon the previous quantisation level number, and G was a gain factor. Typical values of b were 63/64, 31/32 and 15/16 for G = 1.08, 1.18 and 1.5, respectively. The values of M were 0.85 and 1.90 for B = 2; 0.85, 1.00, 1.00 and 1.50 for B = 3; and 0.90, 0.90, 0.90, 0.90, 1.20, 1.60, 2.00 and 2.40 for B =4. Experiments revealed that b should be $\geq 31/32$.

An alternative approach adopted to mitigate the misalignment between the step-sizes at the encoder and decoder due to transmission errors, was to periodically force the step-sizes to specific values. In order to find the appropriate periodicity of enforcement, and the values of the forced step-sizes, we conducted simulations using both male and female speech. We arranged to force the step-sizes of each of the six SBC quantizers at the commencement of every Nf-th 96-bit packet, where Nf was allowed to have values of 1, 3, 5 and 7, and the forcing occurred over the complete range of the step-sizes.

The forced step-size was given by

$$s_F = Fz\,(s_{max} - s_{min}) + s_{min} \qquad (2)$$

where s_{max} and s_{min} were the maximum and minimum step-sizes and Fz , was the step-size scaling factor. The highest SEG-SNR was obtained, as expected, when no step-size forcing was used, as by its very nature arbitarily changing the step-size introduced impairments. However, in the presence of transmission errors these relatively small impairments constituted a small price to pay for the significant gains that accrued in overcoming the digital noise due to transmission errors. In the sub-bands from 0 to 0.5 kHz and 2.5 to 3.0 kHz the optimum Fz was 0.25, whereas in the other sub-bands it was generally zero, particularly as it was advantageous to force update over several frames to minimise the loss of SEG-SNR in the absence of transmission errors. Increasing Nf beyond 7 yielded negligible gains in SEG-SNR. The step-size forcing algorithm achieved a lower SEG-SNR for the values of Nf considered when compared to the leakage algorithm.

3. Channel Coding

Reed Solomon (RS) codes have maximal minimum distance properties [3], and are able to correct both random and burst transmission errors. Therefore they were chosen for our application of transmitting SBC speech over mobile radio channels. We used the block length of 432 bits to code three 96-bit SBC frames by a 2/3 rate RS code.

This RS coding can be embedded into SBC in order to achieve robustness against channel errors, an acceptable speech quality even under the most severe channel conditions, and a near optimum exploitation of the error correcting capability of the RS code. Our approach was to provide more channel coding protection to those bits that were more influential on the perceived speech quality. Experiments were performed where transmitted data corresponding to a specific sub-band coded signal were systematically subjected to a range of transmission errors. The degradation in the recovered speech quality was noted. These experiments were repeated for each of the six sub-bands and the same range of BERs. The greatest perceived distortion occurred when errors were introduced into the 500 to 1000 and 1000 to 1500 Hz sub-bands. The sub-bands at the edges of the overall codec band were less vulnerable to transmission errors. The overall performance of the RS embedded-SBC codec was found to be controllable by allocating RS protection to different bits in each SBC frame, taking into consideration the effect of the different bits on the subjective quality.

We investigated a number of embedded scenarios which had different attributes. The one we preferred was arrived at from the following considerations. Figure 2 shows the situations for the four types of assignments, namely, voiced, intermediate, unvoiced, and signalling/voiced band data signals, in terms of the number of bits from the encoder associated with each sub-band. Thus for voiced speech the bands 0 to 0.5, 0.5 to 1.0, 1.0 to 1.5, 1.5 to 2.0, 2.0 to 2.5 had encoders having word lengths 4, 4, 3, 2, 2, respectively. The bits in each word were labelled sequentially in Figure 2, with the most significant bits (MSBs) encircled, there being 15 bits generated per Nyquist sample. When these 15 bits were loaded into a frame they are packed together (see the last row for the voiced speech assignment). As the bits allocated to each encoder depended on the signal assignment, the positions of the MSBs in the frame were often different when they were placed into the frame. For example, for the fourth position in the frame an MSB was present for voiced, intermediate and signalling conditions. As errors in the MSBs caused significant degradation in the recovered speech, it was most appropriate that the fourth bit in the frame was protected. Similar comments can be made regarding bit numbers 8, 11 and 13. Bit number 15 was the MSB

irrespective of the bit assignment. From our earlier perceptual experiments we knew that both the MSBs and the next MSBs were the most important in controlling speech quality. Accordingly we arranged to send to the buffer, B , all the MSBs, the second MSBs corresponding to bit numbers 7, 10 and 14. Also conveyed to B were bits 2 and 12. This procedure not only reflected the significance of the bits, but also the probability of occurrence of the bit assignment strategy. The buffer B had 6+6x10 = 66 bits, and after three 96 bit frames had been generated there were 198 bits in B . Using an RS code (57,33) over GF(64), i.e., 6 bits/symbol, the 198 bits were coded into 342 bits. As 30 bits per frame were not RS coded, the code frame consisted of 3x30+342 = 432 bits. The RS coding continued for every contiguous three SBC frames.

4. Speech Post-Enhancement Techniques Using RS Decoders

Even after careful RS code design there were occasions when more than t symbol errors occurred in a block. The effect on a non-systematic RS code was as if the error burst that caused the error correction to fail was exaggerated to the entire block. The extended error burst had disastrous consequences because the error extension effects in the quantizers of the SBC due to incorrect step-sizes resulted in long periods of perceptually unacceptable speech.In order to combat this effect we employed post-enhancement methods. As three SBC blocks lasting for 18 ms were RS encoded, we removed the 18 ms segment of erroneous speech whenever the RS codec was overloaded. The missing speech was replaced using the approach of Goodman et.al. [4] for packet switched speech. Specifically, the speech of 6ms duration equal to one SBC block immediately preceding the missing frame was used as a template and was slid back in time along the search window comprising the previously recovered two 18ms long speech frames. The cross-correlation between the template (6 ms) and the momentarily 'covered' speech in the search window (36 ms) was computed, and that part of the speech in the window associated with the highest cross-correlation was noted. The rejected frame of speech was now replaced by three blocks of SBC speech, commencing with the block identified in the search process. In the computation, the normalised cross-correlation of the template T(i) and the search window W(i) employed was

$$R(i) = \frac{\sum_{m=1}^{M} T(m) \cdot W(i+m)}{\sum_{m=1}^{M} |T(m)| \cdot W(i+m)} \qquad (3)$$

where M was the number of samples in the template, and i characterised the position of the template along the search window. Another pattern matching method [4] that was easier to implement was

$$S(i) = \sum_{m=1}^{M} \mathrm{sgn}[T(m)].\mathrm{sgn}[W(m+i)] \qquad (4)$$

although both R(i) and S(i) yielded recovered speech of similar quality. A refinement was to smooth the substituted frame in the recovered speech signal at its edges by employing a raised cosine weighting function. The merging interval included eight samples of the substituted speech and eight of the error free speech at each of the two frame boundaries.

In our experiments whenever the non-systematic RS decoder was overloaded it flagged this information to the SBC decoder that the decoded speech was unacceptable. The last frame of regenerated bits was then applied to the SBC decoder in order to partially correct the SBC quantiser step-sizes. We then substituted for the rejected block of speech a block that was estimated by the speech enhancement technique based on Equation (4) to be similar to the original speech.

An advantage of systematic over non-systematic RS coding was that when code overload occurred, the corrupted parity symbols were separated from the corrupted information data. The latter was not completely erroneous, and was substituted for the deleted frame of data. The enhancement procedure for systematic RS coding was similar to that for non-systematic RS coding. However, when a frame was overloaded, the information data was applied directly to the SBC decoder in its raw state in order to provide appropriate step-size information. The waveform substitution process was as described above.

5. Overall Performance and Discussion

Concatenated male and female speech was bandlimited from 200 Hz to 3,200 Hz, sampled at 8 kHz and applied to the SBC using the step-size leakage algorithm with b = 63/64. Convolutional or Reed Solomon coding ensued and the data stream was transmitted via BPSK over either a Gaussian or a Rayleigh fading channel. After demodulation and bit regeneration at the receiver, channel decoding and SBC decoding were performed, followed by speech enhancement to give the recovered speech signal. Parameters measured were the channel BER, effective BER and the

segmental-SNR of the speech signal. Informal listening experiences were obtained for the speech quality.

Applying the following nomenclature: R, C, P, $\bar{P}$, E, $\bar{E}$, S and $\bar{S}$ representing Reed-Solomon coding, convolutional coding, post processing (i.e., enhancement) of the speech, non-post processing of the speech, embedded coding, non-embedded coding, systematic RS coding and non-systematic RS coding, respectively. In addition, RHG stands for robust Hagelbarger convolutional coding. Figure 3 shows the BER performance of the various schemes for the Rayleigh fading channels. For a BER below 10^{-2} all the schemes, save the RHG, enhanced the BER performance measured in the absence of channel coding. When the BER exceeded 10^{-2}, the systematic RS coding with no post processing ($S\bar{R}P$) was superior, particularly the embedded version $ES\bar{R}P$. For the Rayleigh fading channel, the non-systematic RS coding yielded high effective BERs for high channel BERs, while the convolutional code had a poor performance over the entire range of BERs considered. Of particular interest in mobile radio communications is the channel BER from 10^{-2} to 10^{-1}, and the scheme with the best BER performance over this range was ESRP, being asymptotic to the effective BER when no channel coding was used. The ESRP is not as good as ESRP at low channel BERs, as the unprotected least significant bits (LSBs) now became a significant factor in the effective BER.

The variation of segmental SNR of the recovered SBC speech as a function of the channel BER for a Rayleigh fading channel in the absence of channel coding is displayed in Figure 4 for the two robust step-size algorithms used in the SBC. The step-size leakage algorithm employed leakage factors of either 63/64, 31/32 or 15/16, while the forced step-size algorithm used forced updates Nf at either 1, 3, 5 or 7 SBC frames. As a bench marker we display the curves of the basic SBC codec whose performance deteriorates significantly for BER values in excess of 10^{-4} resulted in more than an order improvement in the BER performance, and were effective for transmissions over both types of channels. We noted that for BERs of approximately 5×10^{-3}, the step-size leakage algorithm was better than the forced step-size algorithm and vice versa for BERs above 5×10^{-3}. Suitable values of b and Nf were 63/64 and 7, respectively.

When the SBC codec employing b = 63/64 was protected by channel coding the improvement in segmental-SNR as a function of BER was substantial as is evident by comparing Figures 4 and 5.Two bench markers were used in Figure 5, neither of which employed channel coding. One was the basic SBC, and the other was the SBC with a step-size leakage b of 63/64. As a segmental-SNR of 13 dB is closely associated with toll quality speech, while communications quality speech is obtained when the segmental-SNR is between 8 and 10 dB, the protected codec was able to convey either communication or toll quality speech for BERs < 2×10^{-2} over both Gaussian and Rayleigh channels.

Whenever the RS decoder was overloaded, the non-systematic version generated more errors than the systematic RS codec. As a consequence the systematic RS codec yielded a lower segmental-SNR than the systematic version (see Figure 5). For example, the $ES\bar{R}P$ curve was significantly better than the $E\bar{S}\bar{R}P$ curve. By comparing $E\bar{S}\bar{R}P$ with $E\bar{S}RP$ or $E\bar{S}RP$ with $ES\bar{R}P$, we can gauge the combination of non-systematic RS coding with post speech enhancement versus systematic RS coding with no post enhancement. The former was superior, although post processing significantly increased the complexity.

The embedded systematic and non-systematic RS coding provided a significantly higher segmental-SNR over high BER channels and a marginally inferior performance over low BER channels compared to the non-embedded arrangement. The best solution with a minimum of complexity was the embedded systematic RS coding without post speech enhancement. By using the post processing procedures further gains in segmental-SNR could be achieved but at the expense of hardware complexity.

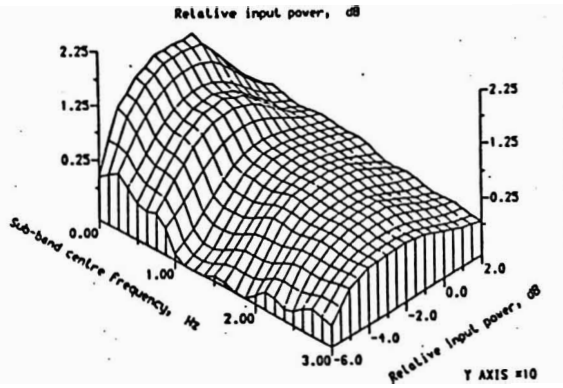
References

[1] R.B. Hanes, "A 16 Kbit/s speech codec for four-channel 64 Kbit/s transmission", Br. Telecom. Technol. J., Vol. 3, No. 1, January 1985, pp. 5-13.

[2] N.S. Jayant, "Adaptive quantization with a one-word memory", BSTJ, Vol. 52, No. 7, September 1973, pp. 1119-1144.

[3] R.E. Blahut, "Theory and practice of error control codes", Addison- Wesley Publishing Company, 1983.

[4] D.J. Goodman, G.B. Lockhart, O.J. Wasem, W-C Wong, "Waveform substitution techniques for recovering missing speech segments in packet voice communications", IEEE - ASSP, Vol. 34, No. 6, December 1986, pp. 1440-1448.



		0.0 TO 0.5 kHz	0.5 TO 1.0 kHz	1.0 TO 1.5 kHz	1.5 TO 2.0 kHz	2.0 TO 2.5 kHz	2.5 TO 3.0 kHz
VOICED	ASSIGNMENT	4	4	3	2	2	0
	ENCODED	1 2 3 ④	5 6 7 ⑧	/ 9 10 ⑪	/ / 12 ⑬	/ / 14 ⑮	
	FRAME	1 2 3 ④	5 6 7 ⑧	9 10 ⑪ 12	⑬ 14 ⑮		
INTERMEDIATE	ASSIGNMENT	4	3	2	2	2	2
	ENCODED	1 2 3 ④	/ 5 6 ⑦	/ / 8 ⑨	/ / 10 ⑪	/ / 12 ⑬	/ / 14 ⑮
	FRAME	1 2 3 ④	5 6 ⑦ 8	⑨ 10 ⑪ 12	⑬ 14 ⑮		
UNVOICED	ASSIGNMENT	2	3	3	3	2	2
	ENCODED	/ / 1 ②	/. 3 4 ⑤	/ 6 7 ⑧	/ 9 10 ⑪	/ / 12 ⑬	/ / 14 ⑮
	FRAME	1 ② 3 4	⑤ 6 7 ⑧	9 10 ⑪ 12	⑬ 14 ⑮		
SIGNALLING/VOD	ASSIGNMENT	0	0	4	4	4	3
	ENCODED	/ / / /	/ / / /	1 2 3 ④	5 6 7 ⑧	9 10 11 ⑫	/ 13 14 ⑮
	FRAME	1 2 3 ④	5 6 7 ⑧	9 10 11 ⑫	13 14 ⑮		

Figure 1. Sub-band codec performance, 16 kb/s, female speech: 3-dimensional plot of the sub-band SEG-SNR as a function of relative input power and sub-band centre frequency;

Figure 2. SBC quantizer bit assignments, showing the 15 bits,labelled 1, 2, ..., 15, for each Nyquist sample. Each column is four bits wide, and the '/' is used when the sub-band quantizer has less than four bits. The circle around a digit indicates the MSB of a quantizer. The "frame" row represents the scenario when the bits are grouped into a frame for transmission.

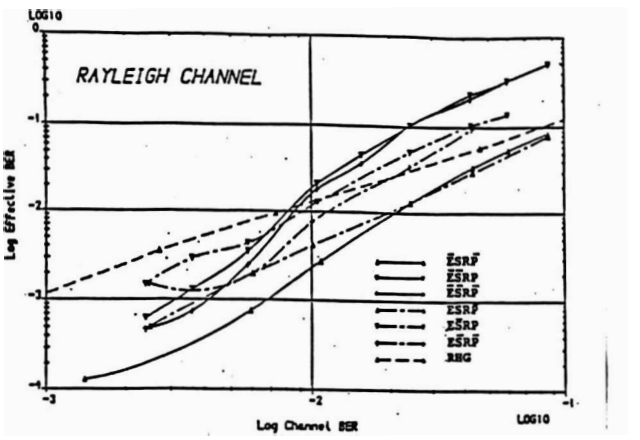


Figure 3. Effective BER (after channel decoding) versus channel BER for a Rayleigh fading channel and for various system configurations.

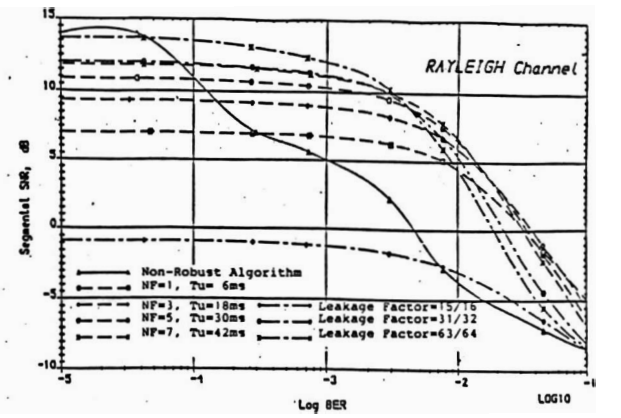


Figure 4. Performance of SBC codec. Segmental SNR versus channel BER for a Rayleigh fading channel and for various system configurations.

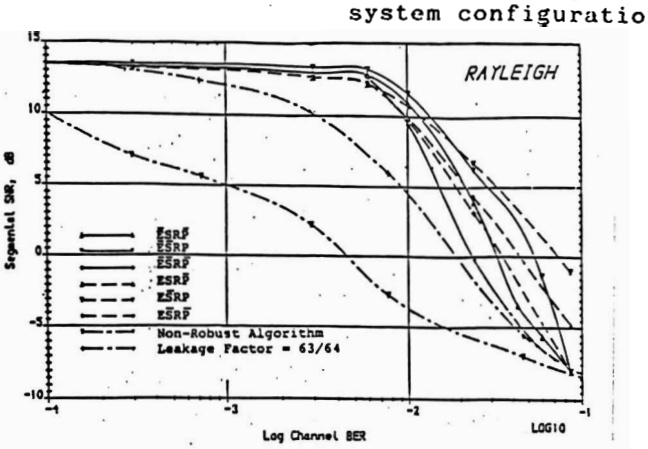


Figure 5. The overall system performance. Segmental SNR versus channel BER for a Rayleigh fading channel and for various system configurations.