Lossless Wideband Speech Coding

C. H. Ritz, J. Parsons
TITR, University of Wollongong, NSW, Australia
chritz@elec.uow.edu.au

Abstract

This paper investigates lossless coding of wideband speech by adding a lossless enhancement layer to the lossy baselayer produced by a standardised wideband speech coder. Both the ITU-T G.722 and G.722.2 speech coders are examined. Entropy results show that potential compression rates are dependent on the type and bit rate of the baselayer coder as well as the symbol size used by the lossless coder. Higher compression rates were obtained by adding a decorrelation stage prior to lossless encoding. The resulting lossless speech coder operates at a bit rate that is approximately 58% of the bit rate of original digitised wideband speech signal.

1. Introduction

Wideband speech refers to speech sampled at 16 kHz and offers superior quality to narrowband speech traditionally used in telephony applications (Mermelstein, 1988). Existing research into wideband speech coding has mainly focused on lossy coding and has resulted in the standardisation of numerous speech coding algorithms. Lossy speech coders aim to reproduce a "perceptually lossless" version of the speech signal at bit rates much lower than the original digitised version. In contrast, lossless speech coding aims to allow full reproduction of the original digitised speech signal and is the focus of this work. Lossless speech coding finds applications where only perfect reconstruction is tolerated. Examples include speech storage for future editing in the recording or movie industries and archiving of legal proceedings or historical events.

While there have been numerous proposals for lossless coding of audio signals (Hans and Schafer, 2001), there is only limited research into lossless coding specifically for speech signals. Many of these existing approaches rely on firstly removing correlation from the speech signal (using for example Linear Prediction (LP)), before encoding the resulting signal with a lossless encoder. Lossless speech coders combining LP with dynamic Huffman coding (Ramsey and Gribble, 1987; Garafolo, Robinson and Fiscus, 1994), arithmetic coding (Stearns, 1995), and Golomb-Rice encoding (Giurcaneanu, Tabus and Astola, 2000) have been proposed. The lowest compression rate (defined as the ratio of lossless bit rate to original bit rate) achieved by these proposals was 43-45% when operating of 16 kHz sampled speech (Garafolo, Robinson and Fiscus, 1994; Giurcaneanu, Tabus and Astola, 2000)). Results for lossless audio coders applied to speech signals have achieved compression rates down to 33% (Li, 2003), although these results were for speech sampled at 44.1 kHz; hence overall bit rates are higher than those reported for wideband speech.

The focus of this research is to investigate lossless speech coding by providing a lossless enhancement layer for an existing standardised wideband speech coder. The motivation for this approach is to allow for both a lossy and lossless version to be available in the one scheme. The two wideband speech coders investigated in this paper are; the ITU-T G.722 speech coder (Mermelstein, 88) and the Adaptive Multi Rate - Wideband (AMR-WB) speech coder (Bessette, *et. al*, 2002), standardised as ITU-T G.722.2.

Section 2 of this paper describes the overall lossless coding scheme proposed. Section 3 examines the residual characteristics of the chosen wideband speech coders in the context of lossless coding. Section 4 presents results of methods for reducing the bit rate required for lossless residual compression while Section 5 presents results for a practical lossless wideband speech coder and compares results with existing lossless coders. Conclusions are presented in Section 6.

2. Lossless coder structure

The structure of the coder proposed here is illustrated in Figure 1. In Figure 1, the speech signal is first encoded with the lossy speech encoder producing a lossy bitstream denoted b_{ly}. The difference between the original and reconstructed speech signal from the lossy coder, referred to as the residual, is then encoded using the lossless encoding stage which produces a second bitstream denoted b_{ls}. In the decoder, both the lossy speech and residual bitstreams are decoded using the lossy and lossless decoders respectively. The synthesized speech signal is added to the recovered residual signal to form the original speech signal.

Proceedings of the 10th Australian International Conference on Speech Science & Technology
Macquarie University, Sydney, December 8 to 10, 2004.

Accepted after full paper review

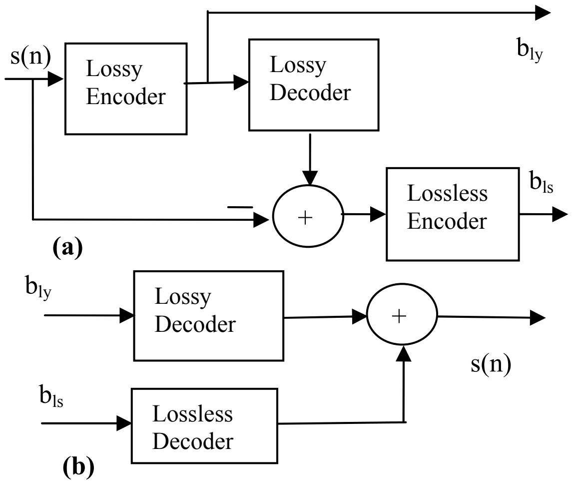

Figure 1. Lossless coding scheme. (a) Encoder, (b) Decoder

	Entropies (bits/sample)		
Symbol Size (bits)	**48**	**56**	**64**
8	9.5	9.1	8.7155
16	7.3	6.8	6.44575
32	6.2	5.9	5.646

Table 1. Entropy results for variation of symbol sizes for different G.722 bit rates (kbps).

	Entropies (bits/sample)			
Symbol Size (bits)	**6.6**	**14.25**	**19.85**	**23.85**
8	12.3	12.2	12.2	12.1
16	10.5	10.3	10.3	10.3
32	7.0	6.9	6.9	7.0

Table 2. Entropy results for variation of symbol sizes for different G.722.2 bit rates (kbps).

3. Residual characteristics of G.722 and G.722.2

Both wideband speech coders selected can operate at multiple bit rates: 48 kbps, 56 kbps and 64 kbps for G.722; and nine different bit rates ranging from 6.6 kbps to 23.85 kbps for G.722.2. To determine which coder would result in the lowest potential lossless rate, the entropy was measured for residuals obtained using different bit rates and coders for the baseline lossy coder. All speech used in this work was taken from the ANDOSL database (ANDOSL, 1998), bandlimited to the range 50 Hz to 7 kHz and resampled to 16 kHz. The entropy was measured in bits per sample using the equation:

$$H = \sum_{i=1}^{N} p_i \log_2 p_i \tag{1}$$

In expression (1), N is the alphabet size and p_i is the probability of each symbol. Note that the residual signals were assumed to be generated by an independent and identically distributed (iid) source and so probabilities were approximated as the frequency of occurrence of each symbol. Hence, this is an upper limit for the entropy.

Initial entropy calculations assumed 16 bit symbol sizes corresponding to the 16 bit residual samples. An investigation was performed into the benefits of using smaller or larger symbol sizes. Specifically, 16 bit residual samples were broken up into 8 bit symbols and also combined to form 32 bit symbols. Tables 1 and 2 present results for entropy of the residual signals in bits per 16 kHz sample using different symbols sizes and averaged over four speech files for the G.722 and G.722.2 coders, respectively.

Table 1 shows that the entropy in bits per sample decreases as the bit rate of the lossy coder increases for all symbol sizes. This result is reasonable as G.722 is an ADPCM coding scheme and hence residuals will have decreased dynamic range and hence entropy as the bit rate and hence coding accuracy increases. Table 1 also shows that the entropy decreases as the symbol size increases. The decreased entropy for 16 bit symbols as opposed to 8 bit symbols indicates correlation may be present between the Most Significant Bytes (MSBs) and Least Significan Bytes (LSBs) of each sample. The further decrease in entropy for 32 bit symbols could indicate correlation between adjacent residual samples.

In contrast to Table 1, Table 2 shows the entropy in bits per sample for a given symbols size is relatively similar for all bit rates of the lossy coder. This could be explained by the fact that G.722.2 uses a Linear Prediction (LP) model and an analysis-by-synthesis coding scheme. Hence, the residual signal will follow a similar shape to the original speech signal resulting in a similar dynamic range regardless of bit rate. However, similar to the results of Table 1, Table 2 shows that the entropy decreases as the symbol size increases. This result could be explained using similar reasoning as used to explain the corresponding results of Table 1.

To compare total overall lossless bit rates when adding the additional lossy base layer, Figure 2 plots potential bit rates for each coder and for each symbol size. Here, potential bit rates are estimated as the sum of the bit rates of the lossy and lossless stages of Figure1. Figure 2 shows that the lowest bit rate results when using 32 bit symbol sizes and the G.722.2 coder operating at 6.6 kbps. However if 16 or 8 bit symbol sizes are selected then the G.722 coder operating at 48 kbps gives the best performance. For both coders tested, the highest lossless compression rate is achieved when operating the respective coder at the lowest bit rate.

Accepted after full paper review

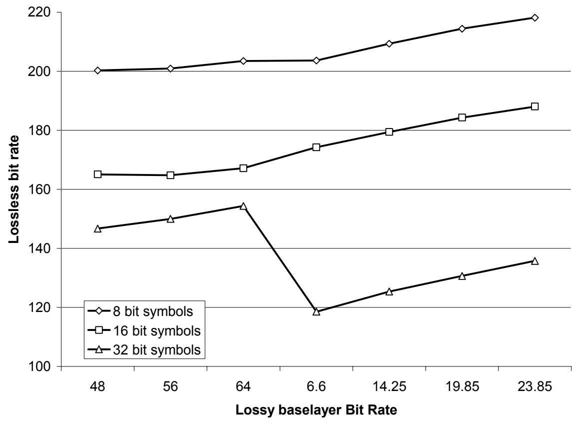

Figure 2. Potential bitrates for different symbols sizes and lossy coders. 48-64 kbps: G.722; 6.6-23.85 kbps: G.722.2.

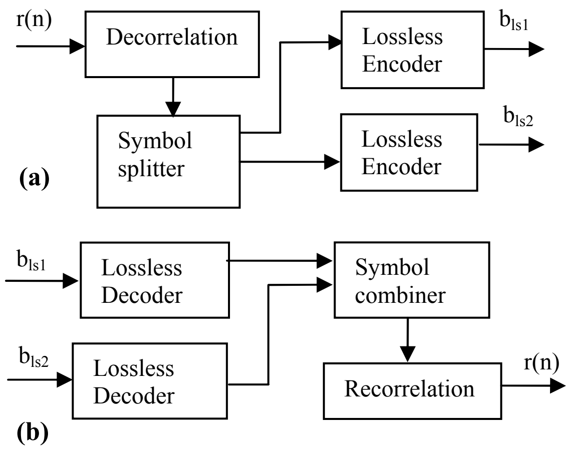

Figure 3. Enhanced lossless coder using symbol splitting and decorrelation. (a) Encoder (b) Decoder.

4. Symbol Splitting and Correlation Reduction

Results of Section 3 indicate that using larger symbol sizes should lead to a coding gain. However, increasing the symbol size leads to an increase in alphabet size. For a practical lossless coder, e.g. a Huffman coder, this requires formation and transmission of a larger Huffman table, hence increasing processing time and lossless rate. Here, a compromise using symbol splitting is investigated.

Results from Section 3 also showed that removing correlation should also lead to a coding gain. Here, decorrelation of the residual is investigated using first order linear prediction via Differential Pulse Code Modulation (DPCM).

4.1. Symbol Splitting

The technique of symbol splitting involves separating larger symbols into smaller symbols and using a separate lossless coder for each new symbol.

Two methods of splitting the 16 bit samples and combining to form new symbols were investigated. The first method split the 16 bit samples into two 8 bit symbols formed from the LSB and MSB. This aimed to exploit the observation that the residual samples were mostly of small magnitude, hence the LSBs and MSBs should have different lossless coding requirements. The second method combined the adjacent 16 bit samples to form two new 16 bit symbols. Each new symbol was formed by combining the adjacent MSBs and LSBs, respectively in an attempt to exploit sample-to-sample correlation.

4.2. Decorrelation via DPCM

In this approach, the residual signal is predicted using DPCM. A new residual is obtained as the difference between the DPCM synthesised residual and the original residual; this new signal is then losslessly coded. This approach aims to exploit correlation between adjacent signal samples as indicated from results of Section 3.

4.3. Combined approach

To take advantage of both symbol splitting decorrelation using DPCM, a combined approach was also investigated. The overall structure of the lossless coder for the residual using this combined approach is shown in Figure 3.

In Figure 3, the residual, r(n) is passed through a decorrelation stage, in this case DPCM. The resulting decorrelated signal samples are then split into two new 8 bit symbols using the first splitting technique investigated in Section 4.1. Each of these symbols are encoded with separate lossless encoders to produce two lossless bit streams b_{ls1} and b_{ls2}.

4.4. Results

Figure 4 shows the resulting overall minimum bit rates when measuring the entropy of these new methods for the same speech files as presented in Figure 2. For comparison purposes, results are re-plotted when using the original 16 bit and 32 bit symbol sizes.

Figure 4 shows that the split method with 16 bit symbols results in a lower bit rate than both the split 8 bit symbols and the original 16 bit symbols. Comparing with Figure 2, results for split 8 bit symbols are also superior to those for the original 8 bit symbols.

Figure 4 also shows that adding the DPCM stage results in a significant reduction in the overall lossless bit rate for the G.722.2 coder but increases the bit rate for the G.722 coder. Results for DPCM and 8-bit split symbols are similar to results obtained without DPCM but using 16 bit split symbols.

While results for 32 bit symbols are still superior, to avoid the practical problems of creation and storage of large Huffman tables, the method utilising a DPCM stage and 8 bit symbol sizes was chosen for the remainder of the work presented here. Based on the results of Figure 4, the G.722.2 coder operating at 6.6 kbps was chosen as the baseline lossy coder.

Proceedings of the 10th Australian International Conference on Speech Science & Technology
Macquarie University, Sydney, December 8 to 10, 2004.

Accepted after full paper review

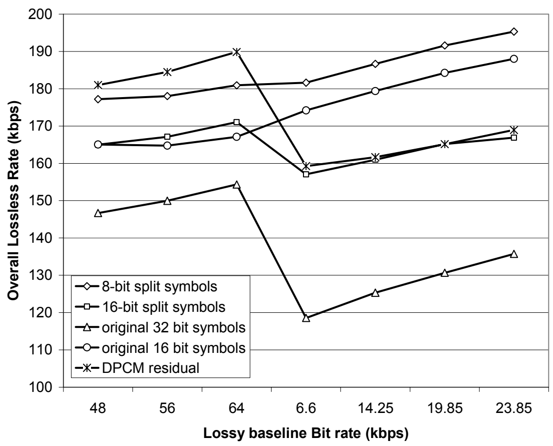

Figure 4. Potential bit rates resulting from using different symbol sizes and decorrelation. 48-64 kbps: G.722; 6.6-23.85 kbps: G.722.2.

5. A practical lossless speech coder

Three well known lossless coding methods for the residual were investigated: Huffman coding, Arithmetic coding and Lempel-Ziv (LZ) coding (Sayood, 2000). For the LZ coding method both LZ77 and LZ78 was utilised. Results were obtained for 14 speech files consisting of male and female sentences encoded using the combined approach described in Section 4.3 and the 6.6 kbps G.722.2 coder.

A further technique was also investigated whereby the DPCM stage described in Section 4.2 was replaced by an Adaptive DPCM (ADPCM) stage similar to that utilised in the G.722 speech encoder. To compare with an existing state of the art lossless audio coder, results were also obtained when coding the residual signals using Monkey's audio (Ashland, 2002). Monkey's audio was chosen as it has been shown to provide superior performance over many other lossless audio coders (Li, 2003). Entropy based compression rates for overall compression rates using each lossless coding technique are shown in Table 3.

Coding Method	Lossless Bit Rate (kbps)
Huffman	160.9
Arithmetic	159.3
LZ77	173.3
LZ78	170.6
Entropy	158.7
Huffman with ADPCM	149
Monkey's audio	123

Table 3. Resulting bit rates for a practical lossless coder.

The results of Table 3 show that both Arithmetic and Huffman coding achieve similar results that are superior to the LZ results. The inferior results of LZ techniques agree with other authors (Giurcaneanu, Tabus and Astola, 2000; Garafolo, Robinson and Fiscus, 1994).

The results of Table 3 also show that the ADPCM technique is able to reduce the bit rate by approximately 10 kbps below the estimated entropy however Monkey's audio results in the lowest overall bit rate. The superior performance could be explained by the more sophisticated decorrelation stage of Monkey's audio compared with the ADPCM technique.

6. Conclusions

This paper has described lossless coding of speech using a lossy coder as the baselayer and Huffman coding applied to a decorrelated residual signal. Results show that approximately 58% compression of the original speech signal could be achieved. This result is still inferior to the rate achievable by a state of the art lossless audio coder.

Future work should focus on two aspects. Firstly, the implementation of a lossless coder capable of encoding 32 bit symbols and avoiding the construction of large symbol dictionaries, using, for example, adaptive Huffman coding (Sayood, 2000). Secondly, the investigation of more sophisticated techniques for removing correlation from the residual signal such as non-linear predictive techniques as used in Monkey's audio. In particular, a more thorough investigation into the most appropriate correlation reduction technique for residual signals obtained from lossy speech coders should be performed.

7. References

ANDOSL: Australian National Database of Spoken Language CD ROM, 1998.

Ashland, M. T., "Monkeys Audio, A fast and powerful lossless audio compressor", www.monkeysaudio.com, Version 3.97, 2002.

Bessette, B., *et. al*, "The adaptive multirate wideband speech codec (AMR-WB)", *IEEE Trans. Speech and Audio Proc.*, Vol.: 10, Issue: 8 , pp. 620 -636, Nov 2002.

Garafolo, J. S., Robinson, T. and Fiscus, J. G., "The development of file formats for very large speech copora: SPHERE and SHORTEN", Proc. ICASSP'94, Vol. I, pp. 113-116, 1994.

Giurcaneanu, C. D., Tabus , I. and Astola, J., "Adaptive context-based sequential prediction for lossless audio compression", *Sig. Proc.*, Vol. 80, pp. 2283-2294, 2000.

Hans, M. and Schafer , R. W., "Lossless compression of digital audio", *IEEE SP Magazine*, pp. 21-32, Jul., 2001.

Li, J., "A progressive to lossless audio coder (PLEAC) with reversible modulated lapped transform", *Proc. ICASSP2003*, Vol. V, pp. 413-416.

Mermelstein, P., 'G.722, A New CCITT Coding Standard for Digital Transmission of Wideband Audio Signals', *IEEE Comm. Magazine*, Vol. 26, No. 1, pp. 8-15, Jan. 1988.

Ramsey, L. T. and Gribble, D., "Information-theoretic compressibility of speech data", *Proc. ICASSP '87*, Vol. 12, p. 17-20, Apr 1987.

Sayood, K., "Introduction to Data Compression", 2nd Ed., Academic Press, 2000.

Stearns, S. D., 'Arithmetic Coding in Lossless Waveform Compression', *IEEE Trans. on Signal Processing*, Vol. 3, No. 8, pp. 1874-1879, August 1995.