A New Approach to Variable Rate Time Domain Encoding of Speech

Ali A. Obaidi and Gordon B. Lockhart
Dept. of Electronic and Electrical Engineering
Leeds University, Leeds LS2 9JT, UK

Abstract

A new variable rate Time Encoded Speech (TES) system is proposed. Shapes between zero crossings and associated gain values are coded using vector quantisation techniques. Shape and gain codebook entries are generated using the K-means clustering algorithm and statistical frequency analysis is used to design a variable index for zero crossing distances in terms of Nyquist samples. The TES decoder is straightforward. It receives a variable index for zero crossing distances and matches it with the shape and gain indices to reconstruct the signal.

Introduction

Time Encoded speech[1] (TES) systems encode and transmit zero crossing distances, shape descriptors between zero crossings and magnitude information related to the shape descriptors. Each transmitted speech segment specifies the quantised distance and the shape between successive zero crossings. Signal processing for TES schemes can be straightforward in comparison with other coding methods but previous studies have indicated relatively high distortion levels.

Previous TES variable-to-constant rate channel transmission produced transmission delays, sometimes badly affecting communication and requiring an increase in TES symbol rate or complex channel coding. A variable transmission scheme using Huffman coding[3] is therefore employed in the proposed scheme. TES transmitted distance indices are Huffman coded with short length codewords assigned to distances with a low number of Nyquist samples as these represent a high proportion of all distances. Shapes between zero crossings are inherently variable as they are proportional to zero crossing distance lengths. This procedure was found to be simple and very effective in lowering the bit transmission rate of the proposed TES system. Further reduction in transmission rate for shapes with more than 13 Nyquist samples was achieved by sub-dividing shapes. Separate codes for each sub-shape were eliminated in favour of one code for all sub-shapes.

TES System

The proposed TES system is outlined in Fig. 1. The system has three main parts comprising encoder, channel transmission and decoder. The encoder incorporates real zero analysis (RZA), shape and gain codebooks, and shape-gain VQ[4]. The real zero analysis block counts the number of Nyquist samples associated with each zero crossing, divides the signal into segments between zero crossings and splits the speech segments into two main groups. The shape and gain codebooks require the use of VQ clustering algorithms for their design. The third part is the implementation of shape-gain vector quantisation using data generated from the other parts of the encoder. The decoder is straightforward. It receives information on zero crossing distances, the shape between zero crossings and the gain and reconstructs the original speech accordingly.

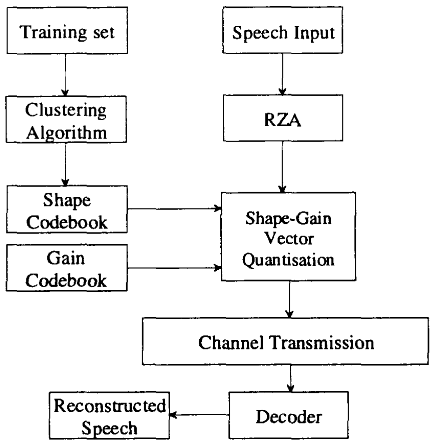

Fig. 1 Proposed TES System.

Statistical analysis of several test speech signals was performed to find the distribution of the number of Nyquist samples between zero crossing (8kHz sampling rate). The analysis showed that the number of Nyquist samples between zero crossing lies between 1 and 39 samples and that zero crossings of more that 13 Nyquist samples have very little chance of occurring.

The number of Nyquist samples between zero crossings were divided into two main groups. Group 1 consists of zero crossings from 1-13 Nyquist samples Whereas group 2 contains zero crossings of 14 Nyquist samples and above. The number of Nyquist samples in group 2 is divided into 2 or 3 sub-groups such that each sub-group represents Nyquist samples in the same range as group 1. This is performed by allocating every 2nd, 3rd sample etc. to different sub-groupings. Sub-grouping economises on shapes for high sample members and creates shapes similar to those already produced by smaller sample members. After finding the number of Nyquist samples associated with each zero crossing, an index representing the number of Nyquist samples is encoded forming one part of the information transmitted to the decoder.

The shape codebook was constructed using the K-means algorithm[5] and depends on the above statistical analysis. The shape codebook is divided into 13 levels. Each level represents M (M= 1,... 13) Nyquist samples and contains N entries depending on the level. After fixing the entries in each level a partition of N cells describing the best shapes is performed using LBG Algorithm. Optimisation of the shape codebook is performed separately for each level and was based on the complexity of the shapes between zero crossings which is related to the number of complex zeros associated with each shape entry.

Gain codebook entries ware optimised using a training set of speech segment resulting in a Max Quantiser[6] where the optimum decision levels are half-way between neighboring reconstruction levels.

The decoder receives three binary indices. The first is the number of Nyquist samples between zero crossings which is variable. Zero crossing indices determine the indices for shapes and are also variable. The third index is the gain. the decoder has 2 sets of codebooks (shape and gain) and when the distance index is received, the decoder checks the distance group (Group 1 or Group 2) and subsequently all the subgroups. The decoder then selects the specified shape and gain entries from the codebooks.

TES Transmission

The generation of TES symbols do not occur at regular time intervals. This implies that the use of constant transmission rate is wasteful given that the zero crossings with a low number of Nyquist samples have a high probability of occurrence. On the other hand the codebooks representing the shape information have different numbers of indices, being high for zero crossings with a large number of Nyquist samples and vice versa. For these reasons, attention was focused on the use of variable transmission rate algorithms.

The Huffman coding algorithm was applied to the proposed TES system as illustrated in Table 1 and 2. The first column of Table 1 represents the distance between zero crossings which is limited to 13 Nyquist samples. Zero crossings with more than 13 Nyquist samples are divided into sub-segments where each sub-segment contains shapes of at least 7 Nyquist samples and not more than 13. The second column gives the probability of occurrence for each distance in descending order. Starting from the end of the table, the 2 least probabilities are combined to form the new probability and proceeding upwards with the other distances until a probability of 1.0 is reached. When two pairs of probabilities grouped together into a subtree, the subtree is given a O value for the upper branch and a value of 1 for the lower branch. The code for each distance is determined by following the path from the top of the table towards the specified distance and allocating either 1 or 0 to form the complete binary code. Table 2 gives the detailed and complete code for distances and shapes following Table 1. The shape codes are represented using the symbol “x” (which is either O or 1) indicating the number of bits allocated for the shapes. The symbol "y" indicates if the signal represents a continuation or not.

Some sub-segments contain equal numbers of Nyquist samples when divided while others contain non-equal sub-segments. The procedure for allocating bits for each sub-segment is wasteful but a more efficient representation of repeated distances is possible. For zero crossings of less or equal to 13 (Table 1) an extra bit is needed to differentiate zero crossings of 7-13 samples from zero crossings of more than 14 Nyquist samples. This extra bit is unavoidable. For zero crossings of 14 to 26 Nyquist samples an extra bit is needed to represent a repetitive distance. Further extra bits are needed to indicate if the two sub-segments for a given distance are similar or not. Finally, an extra bit is needed to indicate that these indices (14-26) are of 2 sub-segments only to differentiate from distances of 3 sub-segments. A maximum of 5 bits can be saved for distances of 14 to 26 Nyquist samples. For distances of 27 to 39 the same procedure is applied with the addition of an extra bit used to check if the last 2 sub-segments (from a total of 3 sub-segments) are of similar number of distances. A maximum of 12 bits can be saved for distances of 27 to 39 Nyquist samples.

Further bit reductions can be achieved if indexes are not need to follow a related sequence. This results in a maximum of 61.91% of total bits be saved relative to normal Huffman procedure.

Discussion

Signal to Noise ratio (SNR) performance (Fig. 2) comparisons of TES with waveform coders are afforded by the guide of figure 3[7]. Performance in Fig. 3 is based on a 24 kbps transmission rate. In terms of signal quality, the proposed TES system provided SNR values higher than PCM at a lower transmission rate and a quality comparable with a DPCM with a first or higher order predictor. DPCM-APF produces a speech quality higher than the proposed TES scheme only when a high order predictor is used with entropy coding. The quality produced by TES is also comparable with DM-AQB and PCM-AQF even when a high order predictor is used and even with a high order predictor with entropy coding. DPCM-AQF can exceed the TES by a small margin but at a rate at least 33 percent greater than that of TES. DPCM-AQF-APF is superior when used in connection with a high order predictor and entropy coding but this is far more complex than the proposed TES scheme. Intelligibility tests on the TES system indicate a very intelligible speech quality and an informal listening test was conducted to compare the intelligibility of the 5 test signals with a 32 kbps ADPCM coder with high order predictor. Some speech test signals such as "Apple" and "coaches" compared favorably with the ADPCM coder at bit transmission rates more than twice the rate of the proposed TES system.

The complexity of the TES coder was estimated according to the number of multiplication and comparisons needed per sample. It was found that the TES system matches the complexity of most simple DPCM and ADPCM (low order prediction) coders.

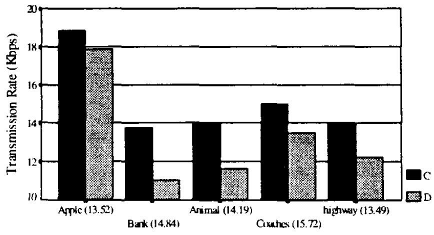

Fig. 2 Transmission Rate Vs SNR for 5 Test Speech Signals.
C: Normal Huffman Coding; D: Optimised Huffman coding

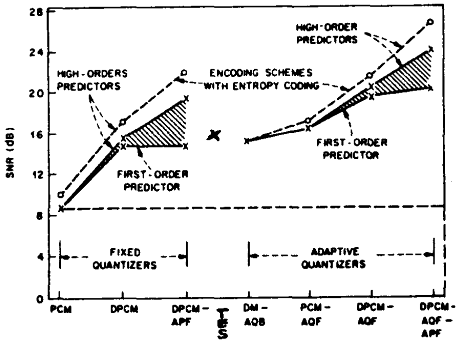

Fig.3 Performance of several 24 kbps Waveform Coding Schemes

References

1. King R.A. etl., "Time Encoded Speech", IEE conf. Pub., 180, pp. 140-143, 1979.
2. Mason D. etl., "Relationship Between System Delay and Transmission Rate in TES", Elec. Letters, 16, pp. 128-130, 1980.
3. Huffman D. A., "A Method For The Construction Of Minimum Redundancy Codes", Proc. of the IRE, 40, pp. 1098-1101, 1952.
4. Sabin M. J. etl., "Product Code Vector Quantisers For Speech Waveform Coding", Conf. Globelcom '82, pp. 1087-1091, 1982.
5. Linde Y. etl, "An Algorithm For Vector Quantizer Design", IEEE trans. on Comm. Com-28, 1, 1980.
6. Max J., "Quantizing For Minimum Distortion", IRE trans. on IT, pp. 7-12, 1960.
7. Nell P., "A Comparative Study of Various Quantization Schemes For Speech Encoding", Bell System Tech. Jorn., pp. 1597-1614, 1975.

Table 1 Example of Huffman Binary Coding

		Auxiliary Message Ensembles											
NS	M.Pr	1	2	3	4	5	6	7	8	9	10	11	12
													1.
												.57	
											.43	.43	
										.32	.32		
									.25	.25	.25		
1	.22	.22	.22	.22	.22	.22	.22	.22	.22	.22			
								.21	.21	.21			
							.17	.17	.17				
						.15	.15	.15	.15				
2	.13	.13	.13	.13	.13	.13	.13	.13					
3	.12	.12	.12	.12	.12	.12	.12	.12					
					.11	.11	.11						
				.10	.10	.10	.10						
4	.09	.09	.09	.09	.09	.09							
5	.08	.08	.08	.08	.08	.08							
			.08	.08	.08								
6	.07	.07	.07	.07	.07								
7	.06	.06	.06	.06									
8	.05	.05	.05	.05									
9	.05	.05	.05										
10	.05	.05	.05										
		.05											
11	.03	.03											
12	.03												
13	.02												

Table 2 Huffman Binary Coding Results.

i	P(i)	$L_D(i)$	$P(i)L_D(i)$	code	$L_S(i)$	$L_D(i)+L_S(i)$	$p(i)\{L_D(i)+L_S(i)\}$	Final code
1	.22	2	.44	10	0	2	.44	10
2	.13	3	.39	010	2	5	.65	010xx
3	.12	3	.36	011	3	6	.72	011xxx
4	.09	4	.36	0000	3	7	.63	0000xxx
5	.08	4	.32	0001	4	8	.64	0001xxxx
6	.07	4	.28	0011	4	8	.56	0011xxxx
7	.06	4	.24	1100	5	9	.54	1100yxxxxx
8	.05	4	.20	1101	5	9	.45	1101yxxxxx
9	.05	4	.20	1110	5	9	.45	1110yxxxxx
10	.05	4	.20	1111	5	9	.45	1111yxxxxx
11	.03	5	.15	00101	6	11	.33	00101yxxxxxx
12	.03	6	.18	001000	6	12	.36	001000yxxxxxx
13	.02	6	.12	001001	6	12	.24	001001yxxxxxx
			$L_{D(av)}$= 3.44				$L_{S(av)}$= 6.46	

i: Distance between zero crossings. P(i): Distance probability. $L_D(i)$: The distance code length in bits. $L_D(av)$: The average distance code length in bits; $\Sigma P(i)L_D(i)$ Ls(i): The shape code length in bits. $L_{S(av)}$: The average distance and shape code length in bits; $\Sigma P(i)\{L_D(i) + L_S(i)\}$