European Conference on
Speech Technology
Edinburgh, Scotland, UK
September 1987

Coding Schemes For Time Encoded Speech (TES) Voice Messages.

S Longshaw[1], J Holbeche[1], R D Hughes[1] and R A King[1]

ABSTRACT

This paper reports on an investigation into the use of Time-Encoded Speech (TES) [1] for the economical storage of digital voice messages in the tactical military arena. Initial results indicate that bit rate reductions of between 20% and 55% may be available using simple coding schemes.

INTRODUCTION

The need for an economical digital description of the human speech waveform for applications in the voice message arena is well documented. Time-Encoding [1] appears to commend itself in this role. For simple TES schemes, high intelligibility and military quality may be obtained at bit-rates of 12-16 kb/s [2] with time delays of 1-2 seconds [3]. These delays impose severe limitations for digital transmission but for voice messaging systems, would appear to be insignificant. Further, since a variable (noisy) channel is absent, efficient coding schemes may be employed to reduce the number of bits per TES message [4], without the vulnerabilities usually associated with such codes. Inspection of the TES symbol stream reveals a clustering of symbols associated with the acoustic event from which they were derived. This suggests that an efficient re-coding of the appropriate segment of the TES symbol stream may permit significant reductions to be achieved. For reconstruction and playback the coding process would be reversed to reproduce the original symbol stream. A detailed description of the TES coding format is reported in reference [5].

CODING SCHEMES

In the present investigations three simple *segmentation* protocols and four *inter-frame coding* options have been exercised in non real-time on a 15.36 second sentence (see appendix). The procedures involved are described below:

SEGMENTATION

Fixed-length Segmentation : The TES symbol stream is partitioned into short (order of 10-20ms) fixed length time-frames.

Variable-length Segmentation : The TES symbol stream is partitioned into variable-length time-frames, where the boundaries of the time-frames are defined by an arbitrarily assigned maximum number N_{max} of permitted elements within the frame. Typically, $7 < N_{max} < 15$.

Whole-message Segmentation : The TES symbol stream for the whole voice message is stored.

Within each time-frame, formed by any of the above schemes, there are N_Z non-zero elements of the First Order Distribution (FOD), from an alphabet of N_A possible elements. For fixed-length and whole-message time-frames N_Z may approach N_A, dependent upon the acoustic material in the utterance. For variable-length time-frames N_Z may be arbitrarily restricted to an "aperture" N_{max} to permit an efficient coding of the frame. The data frames consist of a total of N_T symbols. N_T will be dependent upon the acoustic material and the frame length.

The variable-length time-frame segmentation deployed in this investigation utilizes a "threshold" level, L_T, to signal the point of segmentation. When L_T equals zero and the $(N_{max}+1)$th element of the FOD is non-zero, segmentation is complete and the next frame commences with the last symbol processed. If L_T is non-zero then segmentation does not occur until the frequency of

[1] Military Communications Research Group, School of Electrical Engineering and Science, RMCS (Cranfield), Shrivenham, Swindon. SN6 8LA.

10.21437/ECST.1987-176

occurrence of (N_{max}+1) elements of the FOD is greater than L_T. In its simplest implementation only the N_{max} most frequent elements of the FOD are retained for coding and reconstruction.

INTER-FRAME CODING

For fixed and variable-length segmentation, the FOD of the symbols within each frame is observed. An efficient variable- or fixed-length code, selected from a small number of candidates, is allocated. The TES symbol stream is then efficiently re-coded and stored with housekeeping information to indicate :- the identity of the symbols within the frame, their distribution and the start (or stop) boundary of the frame. Economies achieved are a function of the efficiency of the code versus the housekeeping overheads, per frame.

Three inter-frame coding techniques have been investigated for each of the TES segmentation routines described above. These coding techniques utilised Linear, Bounded Entropic and Huffman coding [4] and have been compared with the Huffman code formed from the complete voice message, as a bench-mark for evaluation. A more detailed description follows:

Linear Coding : For each segment the N_Z elements may be efficiently re-coded using a linear code of 'M' bits per symbol, where $2^M > N_Z$. The information required uniquely to re-code and store each frame is:
- a 'B' bit code to represent the number of non-zero elements N_Z, where $N_A \leq 2^B$. This is required for fixed-length time-frames only.
- N_Z codes of B bits to define the elements which are non-zero within the FOD.
- N_T codes of 'M' bits, where $2^M > N_Z$.
- a B bit code to mark the End of Frame (EOF).

Bounded Entropic Coding : Using a variable-length bounded entropic code (figure 1) the data required uniquely to re-code and store each frame is:
- a 'B' bit code to represent the number of non-zero elements N_Z, where $N_A \leq 2^B$. This is required for fixed-length time-frames only.
- N_Z codes of B bits, where $N_A \leq 2^B$, to define the elements which are non-zero within the FOD. These are stored in decreasing order of probability.
- N_T entropic codes of E_i bits, where $E_i = 1,2,...,N_Z$.
- a N_Z+1 bit entropic code to mark the EOF.

Huffman Coding (Short-Segment) : Using the FOD, the huffman codes are calculated for that distribution and employed to re-code the data frame. The data required uniquely to re-code and store the data frame is:
- a 'B' bit code to represent the number of non-zero elements N_Z, where $N_A \leq 2^B$. This is required for fixed-length time-frames only.
- N_Z codes of B bits, where $N_A \leq 2^B$, to define the elements which are non-zero within the FOD. These are stored in decreasing order of probability.
- N_Z+1 codes to represent bit length changes of the new huffman codes, relative to those of the previous frame. A 1 bit code to indicate whether a change has occured and a 4 bit code represents the difference, if any.
- N_T huffman codes for re-coding the symbol stream.
- (N_Z+1)th huffman code to mark the EOF.

Huffman Coding (Whole-message) : In this measure the whole message is stored and the FOD calculated. Using the FOD the Huffman code is derived and employed for an "optimum" re-coding of the complete message.

RESULTS

TES symbol stream generated from the input message (see appendix for details) required 119,495 bits at 5 bits per epoch (7780 bits/sec). Huffman (whole-message) coding yielded a 27% data reduction.

Data reductions achieved for fixed-length segmentation are presented in Table 1. For 20ms segments Bounded entropic inter-frame coding yielded a data reduction of 25.5% - close to the huffman (whole-message) option.

Data reductions achieved using variable-length time-frames for various aperture and threshold values are presented in figures 2, 3 and 4. In general, bounded entropic coding proved to be more efficient than Huffman coding due to the greater housekeeping overheads per frame required by the subject Huffman protocols. For example, in Table 2 for $L_T = 0$, the short-segment Huffman coding protocol generated more housekeeping than data information resulting in a negative data reduction (data increase) in both fixed- and variable-length segmentation.

CONCLUSIONS

A 15 second phrase has been time-encoded and subjected to a number of simple coding schemes, as candidates for economical digital voice messaging. Initial results indicate data storage reductions of between 20% and 55% may be achieved. Further work is required to confirm these results with a representative population of speakers and phrases and to optimise the relatively immature housekeeping protocols exercised. Code selection on a frame by frame basis is also a candidate for further investigation.

ACKNOWLEDGEMENTS

The partial funding of R D Hughes by the Science and Engineering Research Council under contract GR/D43051 is gratefully acknowledged.

APPENDIX

Sentence used for the results presented was as follows:
"John, this is Charles Bottleneck. The meeting at Leeds on Thursday has been postponed. Richard Hughes will contact you later today with the new dates. See you tomorrow at Portsmouth. Cheers".

Male Speaker, Bandlimited 300 - 3,400 Hz., Duration : 15.36 seconds.

REFERENCES

[1] King R A and Gosling W : **Time Encoded Speech.** IEE Conf. Publ. 180, pp140-143, 1979.

[2] Ching P C : **A Study of time-encoded Speech.** The Radio and Electronic Engineer, Vol 53, No. 9, pp 321-324, Sept 1983.

[3] Mason D C and Balston D M : **Relationship between system delay and transmission rate in time-encoded speech.** Electron. Lett, 16, (4), pp 128-130, 1980.

[4] King R A and Holbeche J : **Impact of entropic coding on the time delays associated with time-encoded speech (TES) systems.** Electron. Lett., 17, (12), pp 394-396, 1981.

[5] Holbeche J, Hughes R D and King R A : **Time Encoded Speech (TES) Descriptors as a Symbol Feature Set for Voice Recognition Systems.** Proc. of IEE Int. Conf. on Speech Input/Output; Techniques and Applications. pp 310-315, 1986.

TES Symbol	Codeword
1	1
2	10
3	100
•	•
•	•
•	•
8	10000000

Figure 1 : Bounded Entropic Codes.

Duration (ms)	Inter-Frame Coding			Whole-message Huffman
	Linear	Bounded Entropic	Huffman	
10	14.2	18.6	-13.4	27
15	18.4	23.1	-2.5	
20	20.8	25.5	4.6	

Table 1 : Percentage Data Reduction for Fixed-length Time Frames.

Thres-hold	Linear	Bounded Entropic	Huffman
0	77.5 / 22.5	70.6 / 29.4	49.1 / 50.9
1	88 / 12	84.1 / 15.9	67.5 / 32.5
2	91.3 / 8.7	88.4 / 11.6	74.7 / 25.3
4	94.2 / 5.8	92.4 / 7.6	82 / 18
6	95.4 / 4.6	94 / 6	85.8 / 14.2
8	96.2 / 3.8	95.1 / 4.9	88.2 / 11.8

D/H D : Data H : Housekeeping

Table 2 : Typical Percentage Trade-Offs Between Data and Housekeeping Iinformation. Aperture = 9.

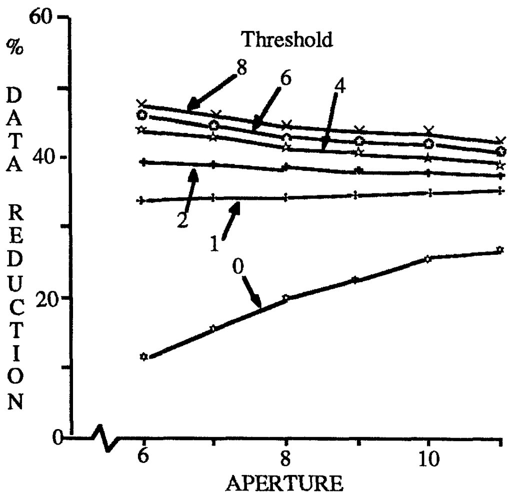

Figure 2 : Percentage Data Reduction Achieved using Variable-length Time Frame and Linear Inter-Frame Coding.

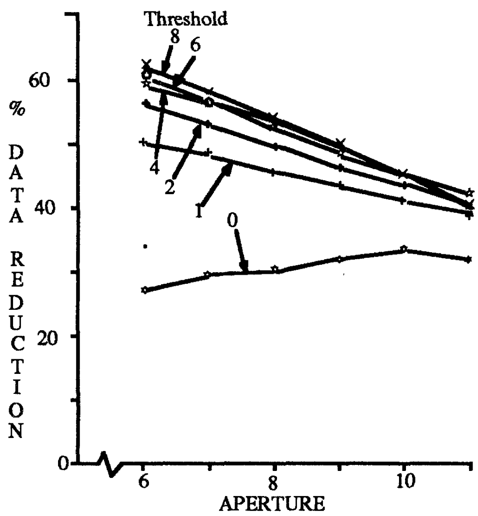

Figure 3 : Percentage Data Reduction Achieved using Variable-length Time-Frame and Bounded Entropic Inter-Frame Coding.

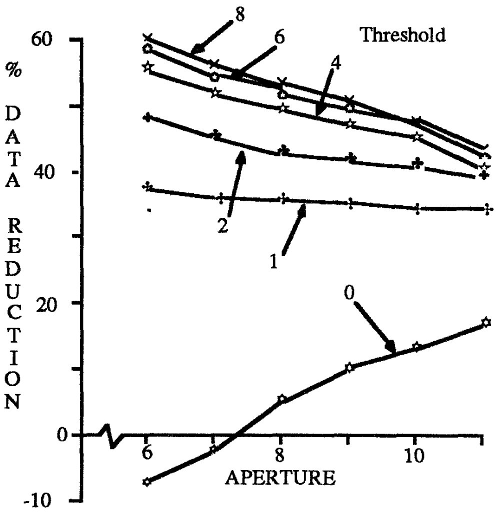

Figure 4 : Percentage Data Reduction Achieved using Variable-length Time-Frame and Short-Segment Huffman Coding.