IEEE International Conference on Communication Information and Computing Technology (ICCICT), June 25-27, 2021, Mumbai, India.

2021 International Conference on Communication information and Computing Technology (ICCICT) | 978-1-6654-0430-3/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICCICT50803.2021.9510116

Speech Compression with Wavelet Transform and Huffman Coding

Ranjushree Pal, Department of Electronics and Telecommunication Engineering,
Dwarkadas J Sanghvi College of Engineering, Mumbai: 400056, India.

Abstract— **The speech signals as it is generated carry lot of information which is redundant and has to be discarded before efficient storage and transmission. Thus, an efficient technique of speech compression without loss of much information giving less distortion in the reconstructed waveform and also retaining the perceptual qualities of speech is very much needed. This paper deals with the recording of female speech signals in Praat software. Wavelet Transform is carried out on these signals which removes lot of redundant information followed by Huffman coding which compresses the signals further. The average compression ratio obtained is three times higher than that obtained with only Wavelet Transform. The reconstructed signals have very less distortion (10^{-4}) while retaining the maximum percentage of signal energy and perceptual qualities of speech. The detailed results are summarized in the concluding section**.

Index Terms— **Compression Ratio, Huffman Coding, Signal distortion, Signal energy, Speech Compression, Wavelet Transform.**

I. Introduction

The compression of speech is required for efficient storage and transmission of speech signals. When speech is generated, it consists of lots of repetitive data which can be discarded without loss of much information and also retaining the perceptual qualities of speech [3]. As speech is compressed, it requires less memory space for storage and also can be quickly stored and retrieved. The compressed speech needs lesser number of bits to be transmitted per second which reduces transmission bandwidth and Intersymbol Interference. Wavelet transform decomposes a signal into highpass and lowpass subbands known as detailed level and approximation level respectively. In the second level of decomposition the lowpass subband is again divided into highpass and lowpass subbands. The coefficient length at each stage is half of the coefficient length of the previous stage. Larger DWT coefficients represent the information part while smaller DWT coefficients represent the noise part of the signal. The coefficients having very small values can be removed by selecting proper threshold. After wavelet transform, the coefficients having higher values than threshold are again compressed with Huffman source coding technique. This coding technique optimizes the average information per binary digit. This is a variable length coding where each message is coded depending on the probability of occurrence of the message. The lesser the probability, higher the code word length.[2]

II. Implementation

A. Flow of Implementation

Praat software is used to record eight female voice signals at a sampling frequency of 8000 Hz and with mono channel. The speech compression and reconstruction flowchart using wavelets and Huffman coding is shown in Fig. 1.

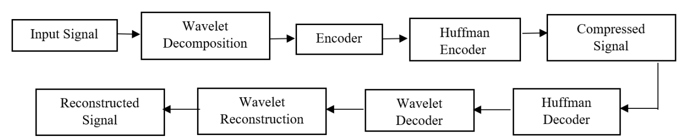

Fig. 1. Flow chart of Signal compression with Wavelets and Huffman Coding.

The full paper is implemented in MATLAB [6]. The input signals are first decomposed by Wavelet transform. A threshold value is calculated and the DWT coefficients below the threshold are neglected. Thus, the information is now in few DWT coefficients. This is how compression is achieved after wavelet transform. Though there is distortion but it is very less. The coefficients are then encoded in the Encoder [1]. The non zero coefficients are coded with Huffman source coding technique and the resultant is a much compressed signal. In the decoding part, the compressed signal is decoded with Huffman decoder and then with wavelet decoder [1]. The reconstructed signal is stored as a wav file.

B. Compression with Wavelets

The wavelet chosen for this experiment is db6.This is selected as it contains the maximum energy after level-one of decimation and gives maximum percentage zero count after level-five of decimation.[1] Wavelet decomposition up to 6^{th} level is done on all recorded signals by 'db6' wavelet. Each level of decomposition consists of two parts: Approximation part and Detailed part [5][4]. While the Detailed part is kept intact, the Approximation part is decomposed again into Approximation part and Detailed part. This decomposition is continued up to six levels after which it is found that most of the coefficient values are negligible and can be removed. So, a threshold is calculated in MATLAB [6] and after thresholding operation, the coefficient values which are more than the threshold forms the elements of row matrix which gives the compressed signal [1]. The other matrix in the encoder stores the index of the nonzero elements which is required during signal reconstruction.

C. Compression with Huffman Coding

The non-zero elements of the wavelet encoder are rounded up to the next higher integer. These form the symbols of Huffman encoder. The probability of each symbol is found using histogram. The Huffman dictionary and Huffman encoder are implemented in MATLAB [6]. The output of the Huffman encoder consists of strings of ones and zeros. The length of the output signal gives the total number of bits required to be transmitted for each signal. The compression ratio and distortion of the original signal is calculated with only wavelet transform[1] and wavelet transform with Huffman coding and the results are compared.

$$Compression\ Ratio = \frac{length\ of\ original\ signal\ in\ bits}{length\ of\ compressed\ signal\ in\ bits} \quad (1)$$

D. Reconstruction of original signal

The Huffman decoder and wavelet decoder [1] are implemented in MATLAB [6]. The reconstructed waveform is written as a wav file which can be audio-played in Praat software. The original input signal and reconstructed output signals are plotted in MATLAB [6]. The distortion (Mean Square Error) and retained signal energy of the reconstructed signal are calculated as follows:

$$Retained\ Energy = \frac{Energy\ in\ reconstructed\ signal}{Energy\ in\ original\ signal} \quad (2)$$

$$Mean\ Square\ Error = \frac{1}{N}\sum_{k=0}^{N}(RSignal(k) - OSignal(k))^2 \quad (3)$$

N = total number of Samples.
RSignal= Reconstructed Signal
OSignal= Original Signal

III. RESULTS

The eight signals are recorded using ‘Praat’ in female voice having a sampling frequency of 8000 Hz and with mono channel:

1. This is a shirt.
2. This is a Test signal.
3. Should we chase.
4. A dtsp assignment.
5. My name is Ranjushree.
6. Joy and Woe are woven fine.
7. This is a book.
8. The quick brown fox jumped over the lazy dog.

These eight signals are used to calculate the Compression Ratio from (1), Retained Signal energy from (2) and distortion from (3). Table I shows the comparison of these eight signals w.r.t. Compression Ratio, Distortion of the original signal when compressed with Wavelet Transform and Wavelet Transform with Huffman Coding and Retained signal energy with Wavelet Transform and Huffman Coding. The average compression ratio with only wavelet transform is found to be 20:1, whereas with wavelet transform and Huffman coding average compression ratio is 62:1. The distortion of the original signal measured with mean square error is of the order of 10^{-4} for both the types of compression. There is a slight increase in distortion with the latter type of compression. The retained signal energy with wavelet transform and Huffman coding is also calculated. The data shows that most of the signal energy is retained after compression. The audio files of the reconstructed waveform are observed in Praat software. Fig.2a and Fig. 2b shows the comparison of spectrogram of some of the above original signals with their reconstructed waveform. The reconstructed signal is also very close to the original signal when played in audio player. Fig.3 shows the histogram plot of the symbols and their frequency of occurrence. The probabilities of the symbols were then calculated which is required for Huffman coding.

TABLE I

Comparison of original input signal and reconstructed output signal with different parameters.

Signal	Compression Ratio with Wavelet Transform	Compression Ratio with Wavelet Transform and Huffman Coding	Distortion with Wavelets	Distortion with Wavelets and Huffman coding	Retained Signal energy with Wavelet Transform and Huffman Coding
1	19:1	57:1	0.0000882	0.0001128	99.06%
2	33:1	106:1	0.0000608	0.0000745	98.74%
3	30:1	77:1	0.000122	0.0001391	98.90%
4	12:1	38:1	0.000136	0.0001700	98.52%
5	17:1	60:1	0.000160	0.0001870	94.82%
6	12:1	40:1	0.000091	0.0001254	99.32%
7	30:1	85:1	0.000039	0.00005511	100%
8	10:1	34:1	0.000067	0.0000844	100%
Average	20:1	62:1	0.0000955	0.0001185	98.67%

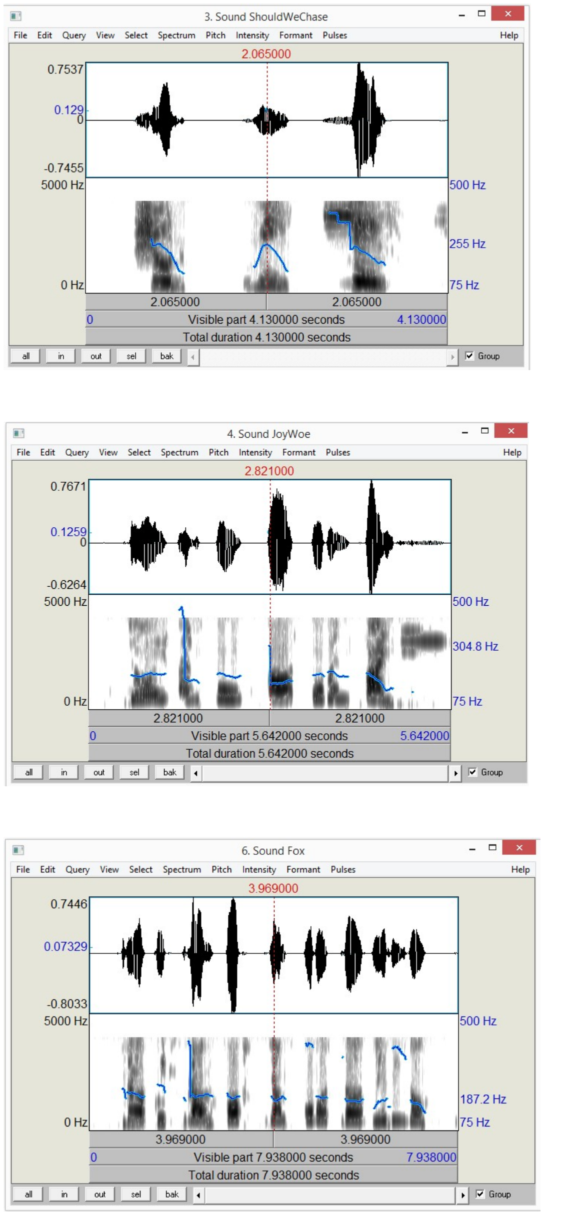

Fig. 2a. Spectrogram of the audio files of the original input signal.

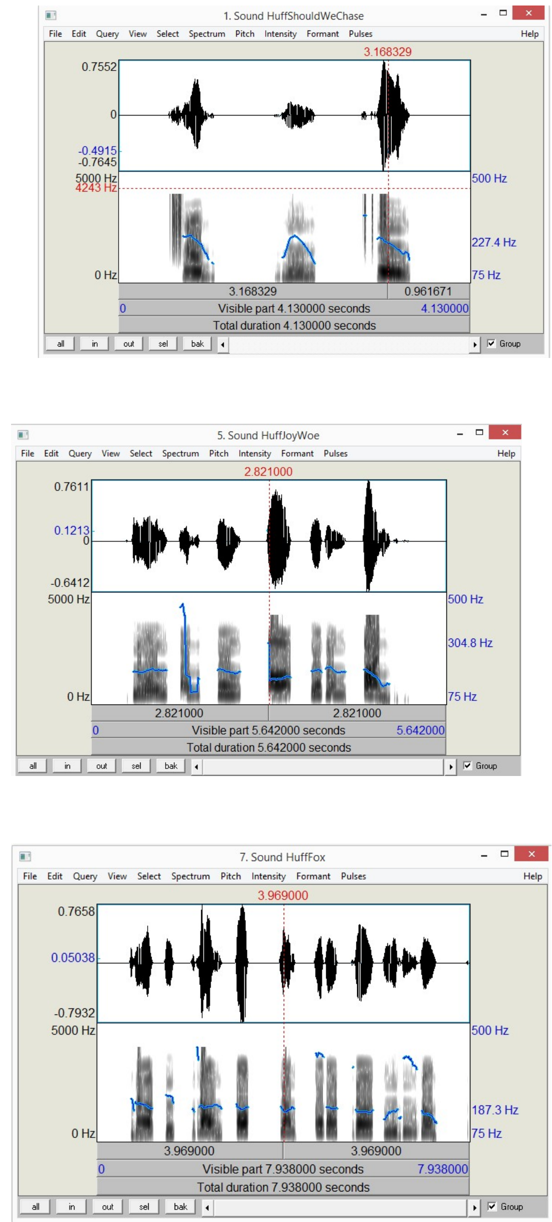

Fig. 2b. Spectrogram of the audio files of the corresponding reconstructed output signal.

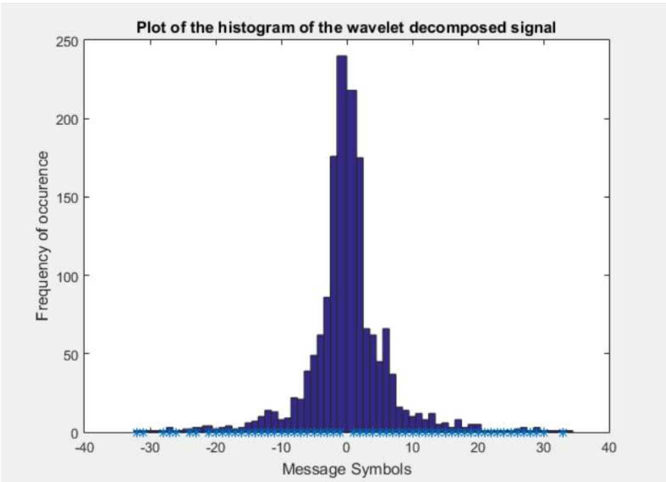

Fig. 3. Histogram of the symbols with their probability of occurrence.

IV. Conclusion

Eight female speech signals are recorded using 'Praat'. These signals undergo two levels of compression - first by Wavelet Transform and second by Huffman coding. The results of compression with Wavelet Transform only and with both Wavelet Transform and Huffman coding are compared. From Table 1 it can be seen that the average Compression ratio with wavelet Transform and Huffman coding is found to be three times more than that with only wavelet transform. The mean square error for both the compression techniques is of the order of 10^{-4}. The distortion with wavelet transform and Huffman coding is slightly more though it can be neglected as the order of the error is very low. The perceptual qualities of original speech are also well restored when compressed with wavelet transform followed by Huffman coding. Thus, wavelet Transform along with Huffman coding is more efficient in Speech Compression rather than only Wavelet Transform.

References

[1] Pal R. (2020) Speech Compression with Wavelets and µ-Law for Wireless Communication. In: Vasudevan H., Gajic Z., Deshmukh A. (eds) Proceedings of International Conference on Wireless Communication. Lecture Notes on Data Engineering and Communications Technologies, vol 36. Springer, Singapore. Available: https://doi.org/10.1007/978-981-15-1002-1_33

[2] Simon Haykins '*Digital Communications*' (2006). India: Wiley-India.

[3] Lawrence R Rabiner '*Digital Processing of Speech Signals'* (2nd Ed), Pearson Education, 2005 ISBN 81-297-0272-X.

[4] Mallat Stephen, "*A Wavelet Tour of Signal Processing*" San Dieago: Academic Press, 1999, ISBN 0-12- 466606.

[5] https://in.mathworks.com/videos/understanding-wavelets

[6] https://www.mathworks.in/help/matlab/