Yanbu Journal of Engineering and Science

Volume 5 | Issue 1 Article 3

2012

SPEECH DATA COMPRESSION BASED ON A SPEEDUP ALGORITHM USING VECTOR RUNLENGTH CODING

Ahmed A. Abdelwahab
Qassim University,KSA, abdelwahab@qec.edu.sa

Follow this and additional works at: https://yjes.researchcommons.org/yjes
The journal in which this article appears is hosted on Digital Commons, an Elsevier platform.

Recommended Citation
Abdelwahab, Ahmed A. (2012), SPEECH DATA COMPRESSION BASED ON A SPEEDUP ALGORITHM USING VECTOR RUNLENGTH CODING, *Yanbu Journal of Engineering and Science*: Vol. 5: Iss. 1, 19-34.
DOI: https://doi.org/10.53370/001c.24063
Available at: https://yjes.researchcommons.org/yjes/vol5/iss1/3

This Research Article is brought to you for free and open access by Yanbu Journal of Engineering and Science. It has been accepted for inclusion in Yanbu Journal of Engineering and Science by an authorized editor of Yanbu Journal of Engineering and Science.

Yanbu Journal
of
Engineering
and Science
ISSN: 1658-5321

Vol. 5, October 2012 (1433H)
www.yjes.org.sa

SPEECH DATA COMPRESSION BASED ON A SPEEDUP ALGORITHM USING VECTOR RUNLENGTH CODING

Ahmed A. Abdelwahab

Electrical Engineering Department, College of Engineering, Qassim University, Buraidah, KSA
abdelwahab@qec.edu.sa

ABSTRACT

This paper proposes a new speech data compression technique based on a new speedup algorithm using Vector Runlength Coding (VRLC). VRLC is proposed and software implemented for speech speedup based on some speech specific properties both in the temporal and transform domains. The successful implementation of VRLC for real-time secure speech transmission system for multimedia applications is described. In this paper, there are two main objectives which are speech data compression and secure transmission for the compressed (speedup) speech by making the transmitted signal completely imperceptible using a lower complexity algorithm compared to other similar methods in the speech coding literature to compare with. Therefore, the reconstructed received speech signal would be as much as possible of a good perceptual quality. The algorithm is applied in both temporal domain and one dimensional discrete cosine transform (1D_DCT) domain. The simulation results show the superiority of applying the new algorithm in transform domain over temporal domain, in terms of the achieved total imperceptible compressed speech transmitting signal and the intelligibility of the reconstructed speech at the receiver.

Keywords: *Vector Runlength Coding, Speech Speedup, Speech Secure Transmission, and Discrete Cosine Transform*

1. INTRODUCTION

Runlength coding is first introduced for coding of two-tone images and binary text and documents in the field of facsimile transmission. It is considered as the simplest method of encoding binary data where clusters of 0's and 1's occur frequently in bit plane coding. Runlength coding is also the standard coding technique for block transform-based image/video compression. A block of quantized transform coefficients is first represented as a sequence of RUN/LEVEL pairs that are then entropy coded where RUN is the number of consecutive zeros and LEVEL is the value of the following nonzero coefficient[1-2]. There are two different main temporal compression techniques that aim to reduce the amount of time it takes to listen to a speech recording while retaining all of its important information: excision where unimportant portions of the recording are removed and speed-up where the playback rate is altered while keeping speaker pitch constant. Excision reduces the recording length by removing automatically selected portions of audio data, effectively compressing it. One simple excision technique removes the between-word silences in the recording. Another acoustically motivated excision

Received: Feb. 26, 2012 Accepted: May 5, 2012

technique which attempts to remove parts of the recording which are acoustically similar to silence regardless of whether these parts contain speech. Speed-up constructs a new recording in which the speaking rate has been artificially altered. Techniques for altering speech rate range from simple sample-frequency alteration (which changes both the playback rate and speaker pitch), to more complex frequency domain procedures [3-4].

For the purpose of comparison, the Faslo player, shown in Fig. 1, is used. Files can be played by selecting the file from the dialog box that appears after clicking Open Files or Open URL button, and pressing the Play button.

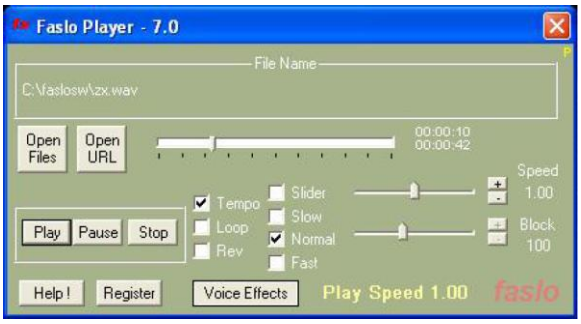

Fig. 1. Faslo Player

Speed of playback can be altered by using the speed slider and clicking the Slider button. Speed can be set anywhere between 6.00 (6 times the normal speed) to 0.25 (1/4 the normal speed) but best quality is attainable in the range from 2.00 to 0.6. There are several factors that can affect the speedup speech quality such as how fast or slow the speaker is speaking. Quality can be fine tuned by selecting the Block (vector) Length slider normally from 22 to 220 [5-6]. For pitch changes (like male/female voice alteration) use the Tempo. A higher (lower) pitch can be obtained by unselecting Tempo and changing to higher (lower) speed. For example, unselecting Tempo and to get a female voice from male, set the speed to 1.15 and click on Play while to get a male voice from female, set the speed to 0.9 and click on Play. Using Faslo player for comparison purposes in this research, only Tempo and Slider will be selected, the block length will be set to the vector lengths used in this paper and the speed will be adjusted according to each case.

In this paper, the vector runlength coding is first introduced for speech speedup in section 2. Although, the speedup speech can range from good understandable quality with low speedup (compression) percentage (which reduces the recording length) to completely no understandable speech with high speedup (compression) percentage for secure transmission purposes, a reconstructed speech of original length can still be reconstructed with faire to good quality at the receiver. Matlab based Computer simulation results in both temporal domain and one dimensional discrete cosine transform (1D_DCT) domain are shown in section 3. Comparisons between some results of the proposed algorithms with Faslo player for speedup speech files are also included in section 3. Section 4 concludes the paper results.

2. THE PROPOSED VECTOR RUNLENGTH CODING FOR SPEECH SPEEDUP

2.1 Reference Book Formation

In the proposed Vector Runlength Coding (VRLC) Speech algorithm, the speech data file is divided into relatively short non-overlapping vectors (blocks) of length k samples each resulting in the original speech data array. The reconstituted speech data file is represented by a reference book which can be considered as an online local speech message dependent codebook. The reference book is extracted from the original speech data array where each vector in the original speech data array contributes in forming a

reference vector in the reference book. The absolute average error (AAE), which is defined in equation (1), are calculated between the first vector and the following consecutive vectors in the original speech data array and compared with a chosen threshold value.

$$AAE = \frac{1}{k}\sum_{i=1}^{k} | x_i - y_i | \qquad (1)$$

Where k is the vector size and x_i & y_i are the elements of the first vector and the following consecutive vector. The number of the following consecutive vectors whose AAE is less than the chosen threshold value plus one is the runlength of this reference vector and is stored in the first element of a vector called the **num vector**. The first reference vector in the reference book is the average vector of the first vector and the following consecutive vectors in the original speech data array whose AAE is less than the chosen threshold. The second reference vector in the reference book is the average vector of the following consecutive vector in the original speech data array and its following consecutive vectors in the original speech data array whose AAE is less than the chosen threshold. The reference book resulted after processing all speech vectors in the original speech data array represents the speedup speech data array where excision is achieved by removing the redundant parts of the speech message and speedup is also achieved by reducing the original length of the speech message. The speedup speech could be, according to the value of the threshold, range from understandable (perceptible) to completely unrecognizable (no understandable or high imperceptible) for secure transmission. The resulted num vector holds the runlength of each corresponding reference vector. The minimum runlength of a reference vector is equal to one when it represents only itself. Therefore, sum of all elements of num vector must equal to the total number of the original speech vectors in the original speech data array. The VRLC speedup percentage for a given threshold value is calculated as shown in equation (2)

$$\text{VRLC speedup \%} = [(n_o - n_r) / n_o] * 100\% \qquad (2)$$

Where n_o = # of vectors in the original speech array = sum of the num vector elements and n_r = # of the reference vectors in the reference book = size (# of elements) of the num vector. The achieved speed by the speedup algorithm is calculated as the ratio between n_o / n_r.

At the receiver, an acceptable (high to good perceptible) version of the original speech message can be reconstructed by repeating consecutively every reference vector in the reference book number of times equal to its corresponding runlength in its associated element in the num vector. The reference vector will not be repeated if its runlength value equals to one.

The larger the AAE threshold value and/or the chosen vector length, the larger the speedup percentage and the less intelligibility (understandability) of the speedup (compressed) speech and the less perceived understanding for the reconstructed speech that can be achieved at the receiver.

2.2 Secure Transmission for the Speedup Speech

In order to increase security for the speedup speech file (the reference book) during transmission and allow using a relatively small AAE threshold value for acceptable intelligibility quality for both the speedup speech and the reconstructed speech file at the receiver, both the num vector and the reference vectors (rows) of the reference book array are permutated by one secret permutation since the number of rows of the reference book equals to the size of the num

vector. This secret permutation represents the key which should be transmitted to the receiver over a secure channel while both the permutated reference book and the permutated num vector can be transmitted over an open channel.

A completely unrecognizable (no understandable) reconstructed speech is resulted if the permutated reference book and its associated permutated num vector are intercepted and used to reconstruct the original speech message without the secret permutation key.

At the receiver, a recognizable version of the speedup speech can be reconstructed using the received secret key which is used to inverse permutate the received permutated reference book. The received secret key is also used to inverse permutate the permutated received num vector and a recognizable version of the original speech message can then be reconstructed.

Average vector signal-to-noise ratio (AVSNR) is calculated for the reconstructed speech for evaluation purposes. Vector SNR (VSNR), which is similar to the segmented SNR (a speech vector is considered as a speech segment) for each speech vector is defined in equation (3).

$$VSNR = 10 * \log_{10}\{\frac{\sum_{i=1}^{i=k} x_i^2}{\sum_{i=1}^{i=k}(x_i - r_i)^2}\} \qquad (3)$$

Where x_i and r_i are the original samples and the reconstructed samples in the speech vector of size k. Accordingly, AVSNR can be defined as AVSNR = [sum of all VSNR of the speech file vectors / total number of speech file vectors]. Moreover, a subjective evaluation for both the speedup speech and the reconstructed speech are carried out by calculating the mean opinion score (MOS) for a group of five people who listens to both the speedup speech and the reconstructed speech processed using this proposed algorithm with different values of vector length and different AAE threshold values. Each one in the group is asked to rank the speech quality from 1 (no understandable and high musical noise) to 5 for excellent quality. The mean opinion score (MOS) out of 5 is then calculated.

3. COMPUTER SIMULATION RESULTS

3.1 ORIGINAL SPEECH FILES GENERATION

In order to provide the speech data for simulation purposes, Matlab 7 function WAVRECORD which records sound using Windows audio input device is used.

WAVRECORD (N, FS, CH) records N audio samples at FS Hertz from CH number of input channels from the Windows WAVE audio device. Standard audio rates are 8000, 11025, 22050, and 44100 Hz. CH can be 1 or 2 (mono or stereo). Samples are returned in a matrix of size N x CH. If not specified, FS=11025 Hz, and CH=1.

WAVRECORD (..., DTYPE) records and returns data using the specified data type DTYPE. Supported data types and the corresponding number of bits per sample recorded in each format are as follows [7]:

DTYPE	bits/sample	Range
'double'	16	signed real from – 1 to 1 (Double precision)
'single'	16	signed real from – 1 to 1 (Single precision)
'int16'	16	signed integers from – 32768 to 32767
'uint8'	8	unsigned integers from 0 to 255

In this research, the sampling frequency is set to 8 KHz and the data type is chosen to be uint8 so that the speech data samples have values from 0 to 255 and the bit rate equals to 8 bits/sample.

The 10_second Arabic speech clip #1 is recorded by 16-year old girl, while the 20_second Arabic/English speech clip #2 is

recorded by 55-year old man and the 10_second Arabic speech clip #3 is recorded by 56-year old man. Figures 2, 3 and 4 show the original speech time signals for the three speech clips used in this research.

The speech message (clip) is divided into relatively short non-overlapping vectors (blocks). Based on comprehensive statistical study, while keeping the speaker pitch unaltered, three different vector lengths and six different AAE threshold values are used. Namely three vector lengths 30, 40 and 50 samples whose time durations are 3.75, 5 and 6.25 msec respectively and six AAE threshold values [4, 6, 8, 10, 12 & 14] are considered for each vector length and this results in 18 different cases for each speech clip.

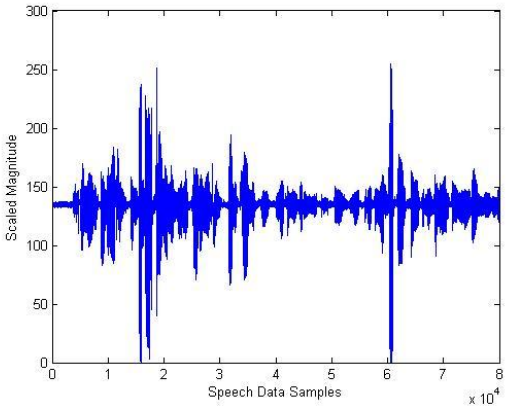

Fig. 2. Clip #1 Original Speech Time Data Signal

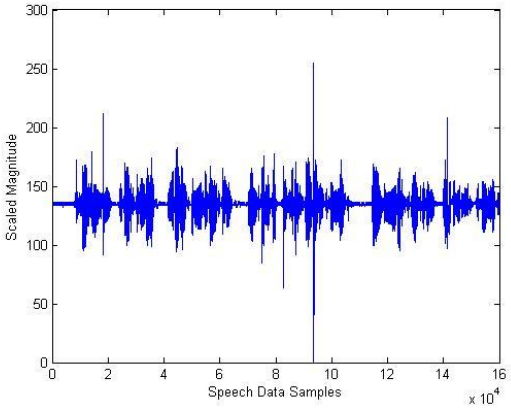

Fig. 3. Clip #2 Original Speech Time Data Signal

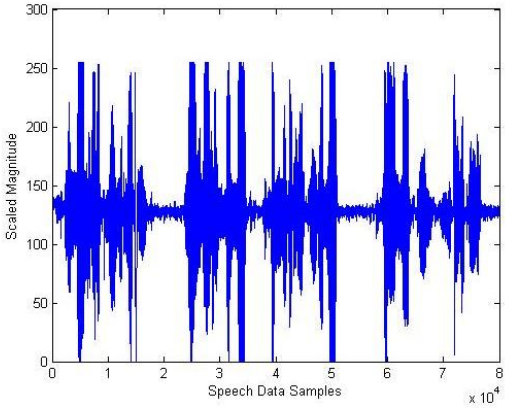

Fig. 4. Clip #3 Original Speech Time Data Signal

3.2 SPEECH PROCESSING USING VRLC ALGORITHM IN TEMPORAL DOMAIN

VRLC algorithm is implemented in temporal domain using Matlab 7. Tables 1, 2 and 3 show the simulation results for speech files clip #1, 2 and 3 respectively.

It can be seen that for a given vector length, increasing the threshold value from 4 to 14 (three and half times) will increase the speedup percentage by about 153% on the average but on the expense of decreasing the reconstructed speech quality by about 8 dB on the average. The best MOS values are obtained for both the speedup speech and the original-length reconstructed speech at the smallest threshold value of four while the maximum speedup percentage ratios are obtained at the largest threshold value

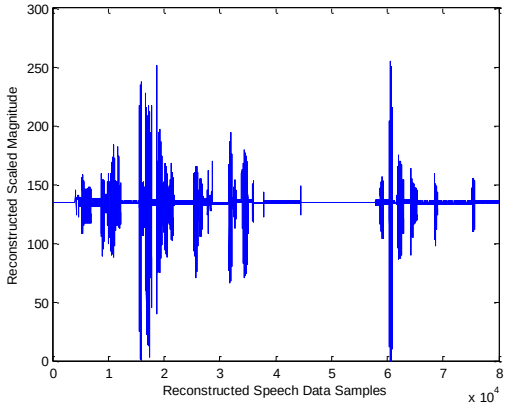

Fig. 5. Clip #1 Reconstructed Speech Time Data Signal (MOS=1.37) Vector size=40, Threshold=14 and Speedup Percentage = 84.35

TABLE 1: TEMPORAL DOMAIN SIMULATION RESULTS FOR 10-SECOND SPEECH CLIP #1 (80 000 SAMPLES)

Original Speech Size in Samples	Threshold Value (Temporal Reference Book Size)	VRLC Speedup Percentage %	MOS For		AVSNR in dB For Reconstructed Speech
			Reconstructed Speech	Speedup Speech	
2666x30	4 (1882 x30)	29.4074	4.64	4.1	38.8818
2666x30	6 (1534 x30)	42.4606	4.35	3.9	36.2526
2666x30	8 (1255x30)	52.9257	3.78	2.5	34.4407
2666x30	10 (1012x30)	62.0405	3.67	2.4	33.1108
2666x30	12 (748x30)	71.9430	3.45	2.1	31.8353
2666x30	14 (573x30)	78.5071	2.47	1.0	31.0696
2000x40	4 (1229x40)	38.55	4.42	3.5	38.0493
2000x40	6 (943x40)	52.85	4.07	2.9	35.2927
2000x40	8 (689x40)	65.55	3.72	2.4	33.1645
2000x40	10 (540x40)	73.0	3.32	2.1	31.9541
2000x40	12 (419x40)	79.0500	2.42	1.4	31.1528
2000x40	14 (313x40)	84.3500	1.37	1.0	30.4417
1600x50	4 (1127x50)	29.5625	4.51	4.2	38.1854
1600x50	6 (934x50)	41.6250	4.21	3.8	38.1854
1600x50	8 (715x50)	55.3125	3.72	2.7	33.6974
1600x50	10 (577x50)	63.9375	3.1	2.4	32.4373
1600x50	12 (434x50)	72.8750	2.2	1.4	31.2834
1600x50	14 (327x50)	79.5625	1.8	1.0	30.5261

TABLE 2: TEMPORAL DOMAIN SIMULATION RESULTS FOR 20-SECOND SPEECH CLIP #2 (160 000 SAMPLES)

Original Speech Size in Samples	Threshold Value (Temporal Reference Book Size)	VRLC Speedup Percentage %	MOS For		AVSNR in dB For Reconstructed Speech
			Reconstructed Speech	Speedup Speech	
5333x30	4 (2968x30)	44.3465	4.94	4.2	44.1061
5333x30	6 (2398x30)	55.0347	4.71	3.8	41.2503
5333x30	8 (1759x30)	67.0167	3.58	2.57	38.6783
5333x30	10 (1120x30)	78.9987	2.22	1.7	36.6240
5333x30	12 (812x30)	84.7740	2.17	1.2	35.6663
5333x30	14 (493x30)	90.7557	1.11	1.0	34.8843
4000x40	4 (2119x40)	47.0250	4.85	4.1	42.8094
4000x40	6 (1641x40)	58.9750	4.22	3.7	39.8724
4000x40	8 (1168x40)	70.8000	3.92	2.7	37.6483
4000x40	10 (824x40)	79.4000	2.72	1.53	36.2849
4000x40	12 (522x40)	86.9500	1.58	1.4	35.1951
4000x40	14 (377x40)	90.5750	1.31	1.0	34.6254
3200x50	4 (1274x50)	60.1875	4.51	3.83	41.4586
3200x50	6 (914x50)	71.4375	4.11	2.91	38.8821
3200x50	8 (630x50)	80.3125	3.02	1.61	36.8485
3200x50	10 (470x50)	85.3125	2.37	1.1	35.3257
3200x50	12 (337x50)	89.4688	1.18	1.0	34.4185
3200x50	14 (239x50)	92.5313	1.09	1.0	33.9587

TABLE 3: TEMPORAL DOMAIN SIMULATION RESULTS FOR 10-SECOND SPEECH CLIP #3 (80 000 SAMPLES)

Original Speech Size in Samples	Threshold Value (Temporal Reference Book Size)	VRLC Speedup Percentage %	MOS For		AVSNR in dB For Reconstructed Speech
			Reconstructed Speech	Speedup Speech	
2666x30	4 (2132x30)	20.0300	4.94	4.9	37.4903
2666x30	6 (1820x30)	31.7329	4.81	4.5	34.6119
2666x30	8 (1658x30)	37.8095	4.48	4.3	33.1075
2666x30	10 (1517x30)	43.0983	4.22	4.1	32.0210
2666x30	12 (1395x30)	47.6744	4.17	3.9	31.0080
2666x30	14 (1308x30)	50.9377	4. 1	3.3	30.3925
2000x40	4 (1652x40)	17.4000	4.9	4.8	36.9465
2000x40	6 (1397x40)	30.1500	4.62	4.5	34.2410
2000x40	8 (1272x40)	36.4000	4.42	4.3	32.9266
2000x40	10 (1164x40)	41.8000	4.22	4.05	31.7651
2000x40	12 (1053x40)	47.3500	4.12	3.87	30.6585
2000x40	14 (982x40)	50.9000	4.07	3.81	29.8763
1600x50	4 (1323x50)	17.3125	4.84	4.55	36.1781
1600x50	6 (1137x50)	28.9375	4.61	4.29	33.9830
1600x50	8 (1024x50)	36	4.52	4.16	32.6444
1600x50	10 (947x50)	40.8125	4.47	4.06	31.6918
1600x50	12 (884x50)	44.7500	4.41	3.9	30.8059
1600x50	14 (820x50)	48.7500	4.18	3.8	30.0159

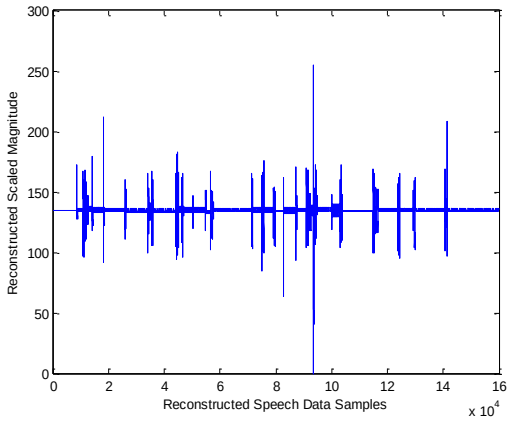

Fig. 6. Clip #2 Reconstructed Speech Time Data Signal (MOS=1.09) Vector size=50, Threshold=14 and Speedup Percentage = 92.53

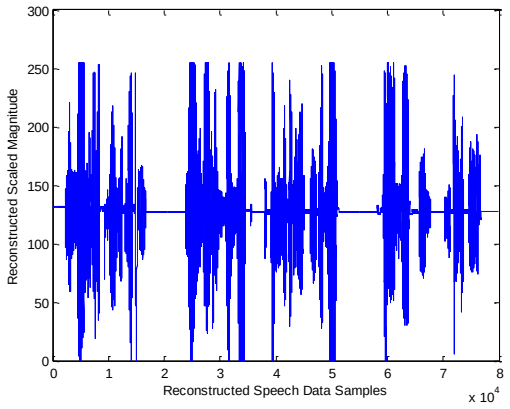

Fig. 7. Clip #3 Reconstructed Speech Time Data Signal (MOS=4.1) Vector size=30, Threshold=14 and Speedup Percentage =50.94

of fourteen with completely imperceptible unrecognizable speedup speech version. For a given threshold value, on the other hand, the AVSNR is decreased by less than 1 dB for clip#1 and clip#2 and by less than 0.5 dB for clip# 3 on the average.

Clip #1 has maximum speedup percentage ratio of 84.35 obtained at speech vector size of 40 samples, while for clip #2, the maximum speedup percentage of 92.53 is obtained at speech vector size of 50 samples and the maximum speedup percentage for clip #3 of 50.94 is obtained at speech vector size of 30 samples. Figures 5, 6 and 7 show the original-length reconstructed speech time signals for the maximum speedup percentages for clips #1, 2 and 3 respectively. While the reconstructed speech is a fair to poor perceptible recognizable version (MOS=1.37 and 1.09) in both figures 5 and 6, it is a good high perceptible recognizable version (MOS=4.1) in Fig. 7. It is clear that the best results in terms of a high speedup percentage with good quality of both the speedup speech and the original-length reconstructed speech obtained from clip #3 which has the most powerful original signal. Note that in general, the AVSNR values do not reflect the real subjective quality of the reconstructed speech. Thus, MOS is calculated for each case of the 18 different cases of each speech clip for both the speedup speech and the original-length reconstructed speech.

3.3 SPEECH PROCESSING USING VRLC ALGORITHM IN 1D_DCT DOMAIN

In order to reduce the size of the reference book (the speedup speech array) and hence reduce the necessary bit rate for reference book transmission and improve the reconstructed speech quality, each speech vector is first transformed using 1D_DCT. Some irrelevant high frequency coefficients can be zeroed and a good quality speech signal can still be reconstructed using inverse 1D_DCT. This reduction in the number of the DCT coefficients should yield a smaller transform domain reference book and remove some noise in the original speech message. Based on a comprehensive study for the reconstructed speech, it is found that the DC coefficient and the first nine AC coefficients for all temporal speech vector lengths 30, 40 and 50 are adequate to obtain a high quality reconstructed speech. Therefore, the compression percentages resulted from applying 1D_DCT as a pre processing for the original speech message vectors are 66.67%, 75% and 80% respectively.

The proposed VRLC is applied on the reduced transform speech domain. The resulted 1D_DCT coefficients reference book represents the speedup speech in the transform domain. The resulted num vector holds the runlength of each corresponding 1D_DCT 10-coefficient reference vector. The resulted 1D_DCT coefficients reference book and its associated num vector are then permutated using a secret permutation key. This secret permutation represents the key which should be transmitted to the receiver over a secure channel while both the permutated 1D_DCT coefficients reference book and its associated permutated num vector can be transmitted over an open channel.

At the receiver, the received secret key will be used to inverse permutate both the permutated 1D_DCT coefficients reference book and its associated permutated num vector. Then 1D_inverse DCT (1D_IDCT) is applied to each row in the resulted reference book in order to obtain the corresponding temporal reference book which represents the speedup speech data array. A recognizable version of the original-length speech message can be reconstructed from the obtained temporal reference book and its associated num vector.

VRLC algorithm is implemented in 1D_DCT domain using Matlab 7. Tables 4, 5 and 6 show the transform domain simulation results for speech files clip #1, 2 and 3 respectively.

The VRLC speedup percentage for a given threshold value is calculated as before, i.e. [(sum of num vector – size of the num vector)/ sum of num vector]*100%, while the total compression percentage for a given threshold value is calculated as following [(#of samples in the original message - # of the 1D_DCT coefficients in the transform reference book)/ (#of samples in the original message)]* 100%.

It can be seen that for a given vector length, increasing the threshold value from 4 to 14 (three and half times) will increase the VRLC speedup percentage by about 573% on the average, but on the expense of decreasing the reconstructed speech quality by only about 5 dB on the average. In general, both MOS and AVSNR for all the 18 cases of Tables 4, 5 and 6 have higher values than their counterparts in Tables 1, 2 and 3. Comparing Tables 1, 2 & 3 and their counterparts in tables 4, 5 & 6 respectively, shows the superiority of applying the VRLC algorithm in 1D_DCT domain over temporal domain in terms of the better achieved quality of both the speed up speech and the reconstructed speech. Although applying the VRLC algorithm in 1D_DCT domain has achieved a high total compression percentage, the speedup percentage is reduced on the average by about one fifth. Clip #1 has the maximum speedup percentage of 49.25% (compression ratio of 87.31%) which is achieved at threshold value of 14 and speech vector length of 40 samples but the reconstructed speech is a faire recognizable version (MOS=2.37). Clip #2 has the maximum speedup percentage of 71.78% (compression ratio of 94.36%) which is achieved at threshold value of 14 and speech vector Clip #3 has the maximum speedup percentage of 20.03% (compression percentage of 73.34%) which is achieved at threshold value of 14 and speech vector length of 30 samples and the reconstructed speech is a good recognizable version (MOS=3.82).

TABLE 4: TRANSFORM DOMAIN SIMULATION RESULTS FOR 10-SECOND SPEECH CLIP #1 (80000 SAMPLES)

Original Speech Size in Samples	Reduced Transform Speech Size in Coefficients	Threshold Value (Transform Reference Book Size)	VRLC speedup Percentage %	Total compression Percentage %	MOS For		AVSNR in dB For Reconstructed Speech
					Reconstructed Speech	Speedup Speech	
2666x30	2666x10	4 (2463x10)	7.6144	69.2048	4.84	4.31	45.1466
2666x30	2666x10	6 (2275x10)	14.6662	71.5554	4.42	4.12	43.7049
2666x30	2666x10	8 (2069x10)	22.3931	74.1310	4.28	3.9	41.7213
2666x30	2666x10	10 (1884x10)	29.3323	76.4441	3.97	3.5	40.4866
2666x30	2666x10	12 (1685x10)	36.7967	78.9322	3.85	3.1	38.6956
2666x30	2666x10	14 (1503x10)	43.6234	81.2078	2.67	2.5	36.9431
2000x40	2000x10	4 (1878x10)	6.1000	76.5250	4.82	4.31	43.1156
2000x40	2000x10	6 (1733x10)	13.3500	78.3375	4.37	4.1	42.0531
2000x40	2000x10	8 (1552x10)	22.4000	80.6000	4.12	3.7	40.6158
2000x40	2000x10	10 (1371x10)	31.4500	82.8625	3.82	3.2	38.9113
2000x40	2000x10	12 (1195x10)	40.2500	85.0625	3.12	2.4	37.3185
2000x40	2000x10	14 (1015x10)	49.2500	87.3125	2.37	2.13	35.5002
1600x50	1600x10	4 (1541x10)	3.6875	80.7375	4.71	4.3	41.9276
1600x50	1600x10	6 (1490x10)	6.8750	81.3750	4.51	3.9	41.4042
1600x50	1600x10	8 (1412x10)	11.7500	82.3500	3.32	3.4	40.551
1600x50	1600x10	10 (1339x10)	16.3125	83.2625	2.77	3.1	39.6678
1600x50	1600x10	12 (1190x10)	25.6250	85.1250	2.41	2.94	37.6914
1600x50	1600x10	14 (1033x10)	35.4375	87.0875	2.55	2.8	35.7622

TABLE 5: TRANSFORM DOMAIN SIMULATION RESULTS FOR 20-SECOND SPEECH CLIP #2 (160000 SAMPLES)

Original Speech Size in Samples	Reduced Transform Speech Size in Coefficients	Threshold Value (Transform Reference Book Size)	VRLC Speedup Percentage %	Total Compression Percentage %	MOS For		AVSNR in dB For Reconstructed Speech
					Reconstructed Speech	Speedup Speech	
5333x30	5333x10	4 (3630x10)	31.9332	77.3111	4.84	4.39	44.2546
5333x30	5333x10	6 (3341x10)	37.3523	79.1174	4.71	4.16	43.6881
5333x30	5333x10	8 (3088x10)	42.0964	80.6988	4.38	3.68	41.8900
5333x30	5333x10	10 (2812x10)	47.2717	82.4239	4.12	3.31	40.7088
5333x30	5333x10	12 (2485x10)	53.4033	84.4678	3.97	3.21	38.6621
5333x30	5333x10	14 (2134x10)	59.9850	86.6617	3.51	2.73	37.2334
4000x40	4000x10	4 (2787x10)	30.3250	82.5813	4.75	4.37	42.6232
4000x40	4000x10	6 (2486x10)	37.8500	84.4625	4.54	3.94	41.4238
4000x40	4000x10	8 (2211x10)	44.7250	86.1812	4.32	3.81	40.0649
4000x40	4000x10	10 (1905x10)	52.3750	88.0938	3.95	3.2	38.1789
4000x40	4000x10	12 (1639x10)	59.0250	89.7563	3.76	2.47	36.3906
4000x40	4000x10	14 (1419x10)	64.5250	91.1312	3.27	1.81	34.4268
3200x50	3200x10	4 (1994x10)	37.6875	87.5375	4.56	3.64	41.2984
3200x50	3200x10	6 (1679x10)	47.5313	89.5062	4.31	3.17	39.5607
3200x50	3200x10	8 (1426x10)	55.4375	91.0875	3.94	2.41	38.0321
3200x50	3200x10	10 (1200x10)	62.5000	92.5000	3.47	1.85	36.6045
3200x50	3200x10	12 (1037x10)	67.5938	93.5188	3.28	1.71	35.0034
3200x50	3200x10	14 (903x10)	71.7813	94.3563	2.98	1.52	33.8365

TABLE 6: TRANSFORM DOMAIN SIMULATION RESULTS FOR 10-SECOND SPEECH CLIP #3 (80000 SAMPLES)

Original Speech Size in Samples	Reduced Transform Speech Size in Coefficients	Threshold Value (Transform Reference Book Size)	VRLC Speedup Percentage %	Total Compression Percentage %	MOS For		AVSNR in dB For Reconstructed Speech
					Reconstructed Speech	Speedup Speech	
2666x30	2666x10	4 (2550x10)	4.3511	68.1170	4.97	4.83	38.6725
2666x30	2666x10	6 (2480x10)	6.9767	68.9922	4.81	4.82	38.2270
2666x30	2666x10	8 (2405x10)	9.7899	69.9300	4.58	4.49	37.7471
2666x30	2666x10	10 (2318x10)	13.0533	71.0178	4.22	4.24	37.2216
2666x30	2666x10	12 (2230x10)	16.3541	72.1180	3.97	4.11	36.6263
2666x30	2666x10	14 (2132x10)	20.0300	73.3433	3.82	3.97	35.9108
2000x40	2000x10	4 (1957x10)	2.1500	75.5375	4.84	4.79	36.7646
2000x40	2000x10	6 (1882x10)	5.9000	76.4750	4.63	4.57	36.1220
2000x40	2000x10	8 (1836x10)	8.2000	77.0500	4.42	4.39	35.7650
2000x40	2000x10	10 (1778x10)	11.1000	77.7750	4.28	4.11	35.2430
2000x40	2000x10	12 (1733x10)	13.3500	78.3375	3.82	3.88	34.8871
2000x40	2000x10	14 (1678x10)	16.1000	79.0250	3.57	3.63	34.4355
1600x50	1600x10	4 (1572x10)	1.7500	80.3500	4.74	4.81	35.4122
1600x50	1600x10	6 (1523x10)	4.8125	80.9625	4.51	4.44	34.8832
1600x50	1600x10	8 (1479x10)	7.5625	81.5125	4.22	4.23	34.4217
1600x50	1600x10	10 (1429x10)	10.6875	82.1375	3.97	3.91	33.9492
1600x50	1600x10	12 (1384x10)	13.5000	82.7000	3.78	3.52	33.4485
1600x50	1600x10	14 (1350x10)	15.6250	83.1250	3.28	3.31	33.0840

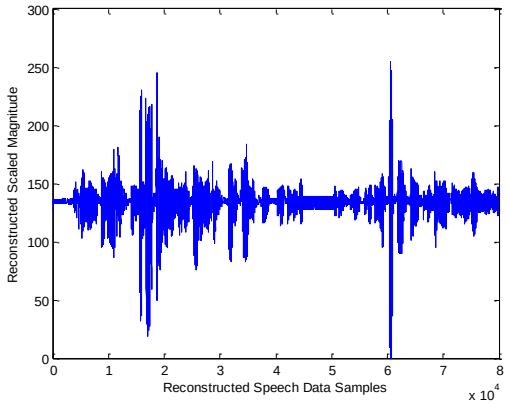

Fig. 8. Clip #1 Reconstructed Speech Time Data Signal (MOS=2.37)

Vector size=40, Threshold=14 and Speedup Percentage =49.25%

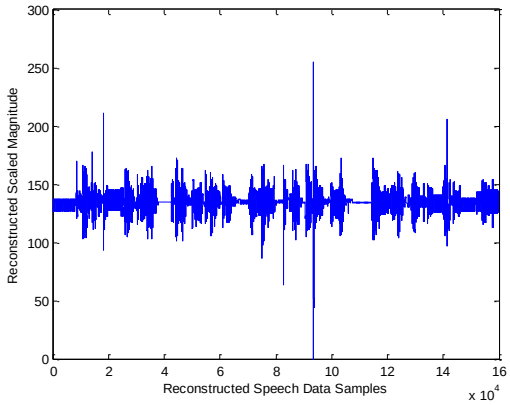

Fig. 9: Clip #2 Reconstructed Speech Time Data Signal (MOS=2.98)

Vector size=50, Threshold=14 and Speedup Percentage =71.78%

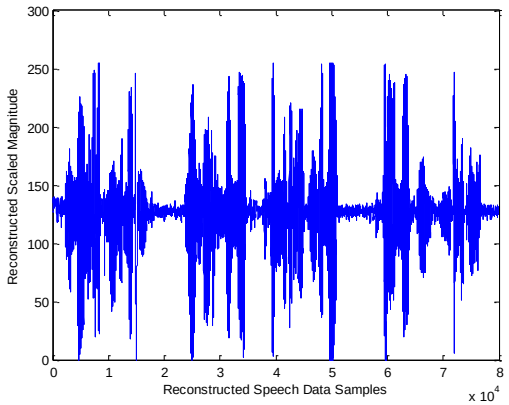

Fig. 10: Clip #3 Reconstructed Speech Time Data Signal (MOS=3.28)

Vector size=30, Threshold=14 and Speedup Percentage =20.03%

Figures 8, 9 and 10 show the reconstructed speech time signals for the maximum speedup percentages for clip # 1, 2 and 3 respectively.

Comparing Figures 5 & 8, 6 & 9 and 7 & 10 show the superiority of applying the VRLC algorithm in 1D_DCT domain over temporal domain in terms of the higher achieved compression percentage and the better perceptible quality of the reconstructed speech.

The three speech clips used in this research are also speedup using the Falso algorithm for the purpose of comparison. In Faslo application, there is no attempt to reconstruct the original length speech message from the speedup one. Therefore, the comparison will be on the perceptible quality using MOS for some speedup files at some chosen speeds as shown in Table 7.

From Table 7, it can be seen that Faslo speech speedup processing is more competitive to the proposed algorithms at low speeds. At high speeds, the proposed algorithm in time domain has better perceptible quality where pitch is more maintained unchanged. Moreover, from these speedup speech files, the proposed algorithms can still reconstruct the original length speech messages with acceptable perceptible quality.

TABLE 7: COMPARISON USING MOS FOR SOME SPEEDUP SPEECH FILES (NA ≡ NOT APPLICABLE)

Clip#	speed	Corresponding Table and parameters: Vector length=VL and Threshold value= Thr	MOS for speedup speech		
			Proposed VRLC in Temporal Domain	Proposed VRLC in Transform Domain	Faslo
1	2.63	Table 1, VL=30 & Thr=10	2.4	NA	1.2
	1.13	Table 4, VL=50 & Thr=8	NA	3.4	3.7
2	4.76	Table 2, VL=30 & Thr=10	1.7	NA	1.02
	2.24	Table 5, VL=50 & Thr=8	NA	2.41	3.6
3	1.76	Table 3, VL=30 & Thr=10	4.1	NA	3.6
	1.08	Table 6, VL=50 & Thr=8	NA	4.23	4.3

4. CONCLUSION

Speech speedup has been achieved by applying the new proposed Vector Runlength Coding (VRLC) for multimedia applications. Moreover, secure transmission is also achieved by using a secret permutation key in transmission of the speedup speech. The algorithm is applied in both temporal domain and one dimensional discrete cosine transform domain. The simulation results are compared with Faslo speech speedup algorithm which shows the superiority of applying the new proposed VRLC technique in terms of the achieved speed and the high quality of both the speedup speech and the reconstructed speech with secure transmission.

For further increasing the speedup percentage of the processed speech file, Number of zero crossings in a speech vector may be used as a specific speech characteristic in future work. The number of zero crossing per reference vector can be calculated. Each two consecutive reference vectors in the reference book having the same number of zero crossings can be combined together by averaging them so that the speedup speech reference book is reduced by one. In this case, their corresponding num values in the num vector should be added so that the reconstructed speech file has the same length as the original one.

REFERENCES

[1] M. Rabbani and P. W. Jones, Digital Image Compression Techniques, SPIE, Bellingham, Washington, 1991.

[2] Chengjie Tu Jie Liang Tran, T.D. "Adaptive runlength coding", IEEE Signal Processing Letters, vol. 10, issue 3, pp 61-64, March 2003.

[3] S. Tucker and S. Whittaker, “Novel techniques for time-compressing speech: An exploratory study,” in Proc. Int. Conf. Acoustic, Speech, Signal Process, pp. 477–480, 2005.

[4] S. Tucker and S. Whittaker, "Temporal Compression of Speech: An Evaluation," IEEE transactions on audio, speech, and language processing, vol. 16, no. 4, pp 790-796, may 2008

[5] http://www.faslo.com

[6] http://www.sharewareconnection.com/faslo-player.htm.

[7] Matlab 7 Help line command.

ضغط بيانات الصوت بناءً على برنامج تسريع للصوت بإستخدام الترميز التكراري للمتجهات

أحمد عبد الوهاب

قسم الهندسة الكهربائية ، جامعة القصيم، المملكة العربية السعودية

الملخص:

يقترح هذا البحث تقنية جديدة لضغط بيانات الصوت بناء على مقترح جديد لتسريع الصوت بإستخدام الترميز التكراري للمتجهات. لقد تم إقتراح وبرمجة بإستخدام الماتلاب لتسريع الصوت معتمدا على بعض خصائص الصوت في كل من المجال الزمني و مجال التحويل أحادي البعد لتحويل جيب التمام المتقطع زمنيا. هناك غرضان لهذا البحث تم تحقيقهما. أولهما ضغط بيانات الصوت المرسل لتقليل نطاق قناة الإتصال اللازمة للنقل وثانيهما تحقيق تأمين للصوت المرسل عبر قناة الإتصال بحيث يكون صوت الإشارة المرسلة مبهم تماما وغير مفهوم. يمكن للمستقبل الذي يعلم التبديل السري كمفتاح لكتاب الترميز التكراري لمتجهات إشارة الصوت المرسلة أن يقوم بإعادة تركيب إشارة الصوت لنفس الفترة الزمنية للإشارة الأصلية. ولقد تم إعادة تركيب إشارة الصوت بجودة عالية وبنسبة ضغط عالية في مجال التحويل.