Asian Journal of Information Technology 5(5) : 536-539, 2006

Text and Grammatical Coding: Voice Compression Algorithm with High Compression Rate

Mohammad Moniruzzaman, Faisal Ahmed and Mohammad Mahmudur Rahman
Department of Computer Science and Engineering, International Islamic University Chittagong,
Dhaka Campus Dhaka-1209, Bangladesh

Abstract: Text Coding (TC) is a voice compression algorithm of very high compression rate. In this coding, speech is recognized and broken into phonemes. The phonemes are encoded into text and transmitted. In the receiver the encoded text is decoded into speech. Text Coding takes very little bandwidth as it transmits text message instead of compressed voice. Consistent pronunciation is not a goal, while predictable pronunciation is. Using the phoneme set, an application developer may guarantee a minimal pronunciation, but not the exact expression. In Grammatical Coding (GC) system, words of text are identified by grammar rules. Each word is then encoded with a predefined number. The encoded number is then again compressed with suitable compression algorithm and transmitted by the communication media. The compressed data is at first decompressed in the receiver. Then from the encoded numbers, the words are decoded by Reverse Grammatical Coding (RGC), which is opposite to Grammatical Coding. Finally the decoded words are converted to speech by Text To Speech (TTS) engine.

Key words: Text Coding, Grammatical Coding voice compression, speech recognition, text to speech, phoneme

INTRODUCTION

Developments in electronics, computers, telecommunications and broadcasting have been rapid and spectacular during the past decade. For many people this has led to a new era of social and economic change. However, more than two-thirds of the world's population does not have affordable access to either voice or data communication. Text and Grammatical Coding based systems may be a low cost alternative in this scenario.

Approximately 50 to 100 bps data rate is sufficient for Text Coding based voice transmission. So one GSM voice channel of 13,000 bps may provide 130 Text Coding voice channels.

As the cost of Text Coding is very little, so it can be used as a major communication system for the poor people in the third world countries.

For transmission backbone, this system may use a specialized network designed only for Text Coding based transmission, which may be wireless or wired. It may also use other existing networks like cellular networks, PSTN, Internet, local IP networks, paging networks etc. as its transmission backbone.

Text coding system: In Text Coding system the transmitter contains speech recognition software and the receiver contains TTS (Text To Speech) software. In the transmitter, speech is recognized and broken into phonemes by speech recognition software. The phonemes are encoded into text and transmitted. In the receiver the encoded text is decoded into speech by TTS software.

Text Coding takes very little bandwidth as it transmits text message instead of compressed voice. For example, the speech: “hello” is converted to phonetic text: hehlow. This phonetic text is transmitted.

Speaking hello takes nearly 0.5 second. 32 characters are sufficient to represent phonetic text. If each character in the phonetic text requires 5 bits for transmission, then the 9 characters text h eh l ow requires 45 bits. Thus 1-second speech requires 90 bits. So 100 bps data rate is sufficient for phonetic text based voice transmission.

Consistent pronunciation is not a goal, while predictable pronunciation is. Using the phoneme set, an application developer may guarantee a minimal pronunciation, but not the exact expression. So, the word first may always be pronounced as first, never as fist or feast, etc, but the accent may be slightly different.

Grammatical coding system: In the transmitter, Input Source (IS) takes the input in text format and encodes in numbers by Grammatical Coding software. In the receiver, the encrypted text is decrypted by Reverse Grammatical Coding (RGC), which is opposite to GC.

In this system several short Tables and one long Table exist. The length of short Table may be 128 (28-1) and long Table may be 32768 (216-1).

Corresponding Author: Mohammad Moniruzzaman, Department of Computer Science and Engineering, International Islamic University Chittagong, Dhaka Campus Dhaka-1209, Bangladesh

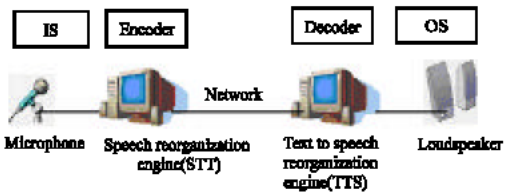


Fig: Text coding system

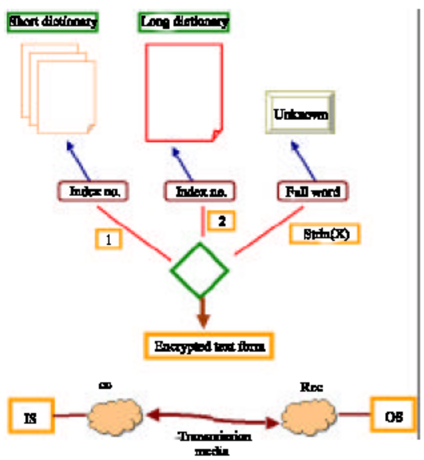


Fig: Grammatical coding system

For each type of word there is one short Table. For example the Table for 'Noun' contains 128 nouns, which are the mostly occurred in everyday talking. Similar for pronoun, verb adverb etc. The long Table contains 32768 words, which are the mostly occurred words of the full dictionary.

When a sentence comes to GC, it identifies the sentence rule and writes the rule number in the encoded data stream. Then for each word, GC checks if the word is in short dictionary. For example, if a word is 'Man' then its type is Noun. So 'Man' is searched in the 'Noun' table. Length of Noun Table is 128 as a short Table. If 'Man' is found in the index number 5 of Noun Table, then '0-000.0101' is written in the encrypted text. Size of this encrypted Man is 1 byte. The MSB bit: 0- of '0-000.0101' means that, the length of the encoded word is 1 byte. The LSB 7 bits: -000.0101 means that, the index number of the word is 5 and the word is Man.

If Man is not found in the short Table for Noun, then the long Table is searched. Length of long Table is 32768. If Man is found in the long Table in the index number 9, then 1-000.0000.0000.1001. Size of this encrypted Man now is 2 bytes. The MSB bit: 1- of'1-000.0000.0000.1001' means that, the length of the encoded word is 2 bytes. The LSB 15 bits: -000.0000.0000.1001 means that, the index number of the word is 9 and the word is Man.

If Man is not found in even the long dictionary, then it is directly written in the encrypted text. Size of this encrypted Man now is 3 bytes.

In the Figure of GC, IS means Input Source and OS means Output Source.

Some examples of grammatical coding

Example # 1:

Input string	What is your name?
Rule no	5
Rule	Ques+Verb+Object+Object
Encrypted output	5 1 6 5 4
Size of encrypted words in bytes	1 1 1 2 2
Total size of output in bytes	7

Example # 2

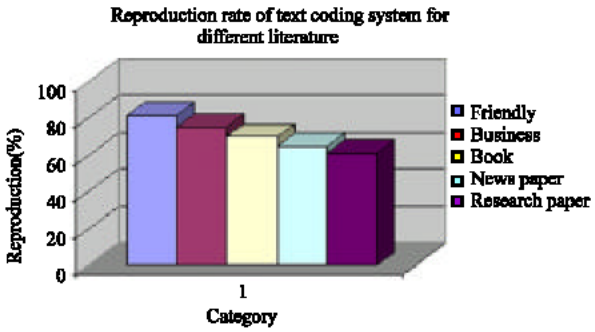
Input string	I m Badal
Rule no	0
Rule	Sub+Unknown+Unknown
Encrypted output	0 40 m Badal
Size of encrypted words in bytes	1 1 1 5
Total size of output in bytes	8

The combined system: In the complete system, the following steps may be used:

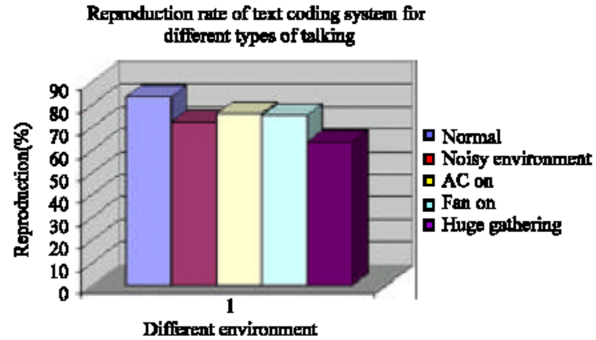
1. Getting Input (Analog to digital)
2. Text Coding (Speech recognition)
3. Grammatical Coding
4. Compression
5. Transmission
6. Decompression
7. Reverse Grammatical Coding
8. Reverse Text Coding (Text to speech)
9. Generating output

Comparison with other encoding techniques: Text Coding is a highly efficient voice compression technique comparing with others. The RPE-LPC coding used in GSM is the mostly used voice encoding technique today. It requires 13000 bps data rate. But TC requires only 100 bps. So one GSM voice channel may provide 130 Text Coding voice channels. Thus the telephone cost in TC may be reduced to an amount of one-hundredth with slight quality degradation. However developments in TC may allow better quality with lower data rate.

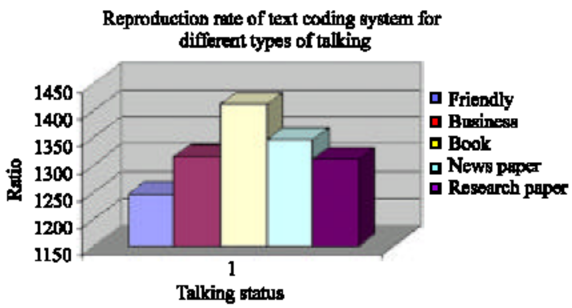
Human voice frequency is in the range of 500 to 4000 Hz. According to the Nyquist theorem, the sampling rate



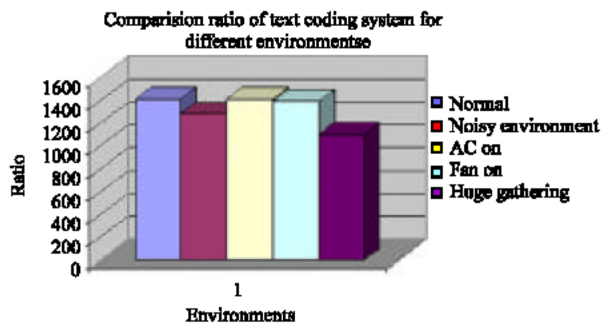
Friendly	82
Business	75
Book	70
News paper	65
Research paper	61



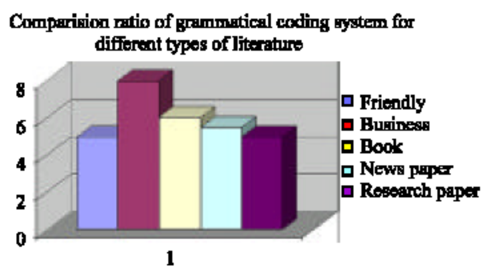
Normal	83
Noisy environment	72
AC on	76
Fan on	75
Huge gathering	63



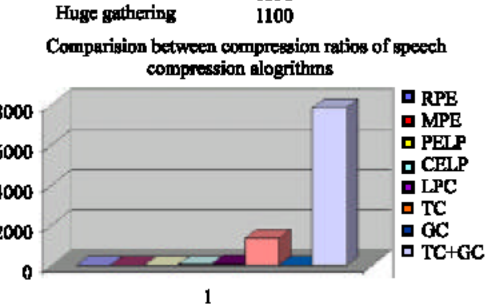
Friendly	1246
Business	1315.06
Book	1411.76
News paper	1347.54
Research paper	1311.02



Normal	1411.86
Noisy environment	1282.05
AC on	1411.86
Fan on	1390
Huge gathering	1100



Friendly	5
Business	8
Book	6
News paper	5.0
Research paper	5



RPE	4.92
MPE	6.8
PELP	6.8
CELP	13.33
LPC	26.66
TC	1333.33
GC	5.9
TC+GC	7866.65

must be at least two times of the highest frequency.

So the sampling rate
= 2* Highest frequency
= 2* 4000
= 8000 samples/sec

Assuming 8 bit sampling, the transmission bandwidth
= 8 X 8000 bits/sec
= 64 kbps

Taking this 64 kbps as the base, we have calculated compression ratio of different compression algorithms. Transmission bandwidth of CELP is 4.8 kbps[1]. So its compression ratio

= 4.8 Kbps: 64Kbps
= 1: 13.33

For TC, we found 56 words came per minute and this 56 words having on and average 346 characters per minute, which is 6 characters per second. If 8 bits are needed for 1 char, then 48 bits (8*6) are needed for 6 chars. So compression ratio of TC

= 48 bps: 64 kbps
= 48: 64000
= 1: 1333.33

CELP	= 4.8 Kbps: 64 Kbps = 1:13.33
RPE	= 13 Kbps: 64 Kbps = 1:4.92
MPE	= 9.6 Kbps: 64 Kbps = 1:6.8
RELP	= 9.6 Kbps: 64 Kbps= 1:6.8
LPC	= 2.4 Kbps: 64 Kbps= 1:26.66
TC	= 48 bps: 64 Kbps = 48: 64000 = 1:1333.33
TC+GC	= 1333.33 * 5.9 = 7866.65

Performance study: The performance of TC, GC and combined system is observed in different environments and for different types of conversations. Although TC has a high compression ratio but it cannot reproduce all the words it encoded. So its reproduction rate is also observed. Reproduction rate is the percent of words reproduced, after encoding and decoding a speech by TC.

Carrier: For transmission backbone, Text coding may use a specialized network designed only for Text Coding based transmission, which may be wireless or wired. It may also use other existing networks like cellular networks (GSM, CDMA etc.), PSTN, Internet, local IP networks (Intranet), paging networks (Mobitex, Datatac etc.) as its transmission backbone.

Cost estimation: Specialized communication network may be designed only for Text Coding based system. Assume two 100 Mbps channels for uplink and downlink are assigned for this purpose. Each subscriber requires 100 bps data rate for uplink and similar for downlink for one conversation. Then one transmission tower may allow one million subscribers to talk simultaneously at a time. Assume each subscriber talks averagely 10% time in a day (i.e., 2.5 hrs). Then 10 million subscribers may use the network to talk 2.5 hours in a day. Allowing 10 million subscribers is nearly sufficient for a whole country or territory.

If initial cost of a transmission tower is 10 million US$ including all necessary equipment, then each of 10 million subscribers has to pay only 1 US$ initially to take connection.

If monthly cost of a transmission tower is 0.1 million US$, then each of 10 million subscribers has to pay only 0.01 US$ (1 cent) per month to talk 2.5 hours everyday!

Achievement: Using text coding only 100 bps is sufficient for one-way continuous voice communication. If two callers are talking simultaneously then each caller talks approximately 50% time and each requires 50 bps. So again 100 bps is sufficient for a two-way voice communication.

Most other communication systems are designed in the developed countries. They focus the requirement and financial capability of people of those countries. These high cost and high-featured communication systems are out of reach of the poor people of the third world who are the 80% of the world. Almost 80% people of the world live in 3 countries like China, India and Indonesia. They also need a communication system but most of them are unable to afford a GSM, CDMA or a 3G phone. Text and grammatical coding based phones may be an affordable communication device for them considering their financial capability.

CONCLUSION

Text and Grammatical Coding based systems require very little data rate. It can be used on any exiting or newly designed backbone. It easily affordable by the poor people of the third world, who are the most population of the world. So it may be a successful technology to make wireless for the mass.

REFERENCES

1. Stallings, W., 2002. Wireless Communications and Networking, 2nd Indian Reprint, Pearson Education (Singapore) Pte. Ltd., India, Delhi.