SYMMETRIC ENCRYPTION ALGORITHM IN SPEECH CODING FOR DEFENCE COMMUNICATIONS

Akella Amarendra Babu[1] and Ramadevi Yellasiri[2]

[1]Progressive Engineering College, Hyderabad, Andhra Pradesh, India
aababu.akella@gmail.com
[2]Department of CSE, Chaitanya Bharathi Institute of Technology, Hyderabad, India
yrd@cbit.ac.in

ABSTRACT

In battlefield, messages must be encrypted to provide protection from enemy interception. Several speech coding algorithms have been developed to provide secure communications. FS1015 LPC in 1984, FS1016 CELP in 1991and FS MELP in 1997 became official federal standards. In 2002, the US DoD adopted enhanced MELP (MELPe). Later on in 2007, US Naval Research Laboratories have designed Variable Data Rate (VDR) voice processor.

Although certain degree of inherent security is ensured in all the above coding algorithms by way of compression techniques, strength of security in these algorithms is weak as the codecs using the above algorithms are vulnerable to interception. Explicit encryption gadgets need to be provided as accessory to provide strong secure communications.

In this paper, we have described an algorithm which provides robust and secure communications. This Robust Secure Coder (RSC) is backward compatible with the existing codec's and operates at marginally higher bit rates when switched to secure mode.

KEYWORDS

Enhanced MELP, Data Encryption standards (DES), FS1015 LPC, Triple DES, Robust Secure Coder (RSC)

1. INTRODUCTION

Speech coders are classified into waveform coders, parametric and hybrid coders. Waveform coders like Pulse Code Modulation (PCM) and Adaptive differential PCM (ADPCM) attempt to preserve the original shape of the input signal and work at bit rate of 32 kbps and above. Parametric coders parameters of input speech signal are estimated and these parameters are used to synthesize the speech signal. This class of coders work typically in the range of 2 to 5 kbps. Example coders of this class linear prediction coding (LPC) and Mixed Excitation linear Prediction (MELP). Hybrid coder combines the strength of waveform coder with that of parametric coder. These coders typically operate between 5to 32 Kbps. Code excited Linear

Natarajan Meghanathan, et al. (Eds): ITCS, SIP, JSE-2012, CS & IT 04, pp. 369–376, 2012.
 DOI : 10.5121/csit.2012.2134

prediction algorithm, its variants, mixed excitation linear prediction algorithm and its variants belong to this class.

In parametric speech coding, 256 samples of input speech signal are buffered into frames and passed through linear prediction filter. The frame can be represented by ten filter coefficients, plus scale factor. 4096 bits corresponding to 256 samples of original speech frame are converted into 45 bits per frame.

The speech coding procedure is summarised as under:

- Encoding

Derive the filter coefficients from the speech from
Derive the scale factor from the speech frame.
Transmit filter coefficients and scale factor to the decoder.

- Decoding

Generate white noise sequence.
Multiply the white noise samples by the scale factor.
Construct the filter using the coefficients from the encoder and filter the scaled white noise sequence. Output speech is the output of the filter.

LPC coder uses a fully parametric model and produces intelligible speech at 2.4 kbps. However, it generates annoying artefacts such as buzzes, thumps and tonal noises. MELP utilizes additional parameters to capture the underlying signal dynamics. MELP voice encoder is reviewed in this paper.

In symmetric encryption system, both sender and receiver use the same key. If the sender and receiver each use different keys, the system is referred to as asymmetric or public-key encryption system. A block cipher processes the plain text input in fixed size blocks and produces a block of cipher text of equal size for each plain text block.

Block symmetric encryption is suitable for use with parametric speech coders because both buffer the input data into frames and process the same frame by frame.

2. Mixed Excitation Linear Prediction (MELP) Coder

2.1. Block Diagram of MELP [1]

A block diagram of the **MELP** model of speech production is shown in Figure 1, which is an attempt to improve upon the LPC model. MELP decoder utilizes a sophisticated interpolation technique to smooth out inter frame transitions. A randomly generated period jitter is used to perturb the value of the pitch period so as to generate an aperiodic impulse train. The MELP coder extends the number of classes into three: unvoiced, voiced, and jittery voiced. The latter state corresponds to the case when the excitation is aperiodic but not completely random, which is often encountered in voicing transitions. This jittery voiced state is controlled in the MELP model by the pitch jitter parameter and is essentially a random number. A period jitter uniformly distributed up to +/- 25% of the pitch period produced good results. The short isolated tones, often encountered in LPC coded speech due to misclassification of voicing state, are reduced to a minimum.

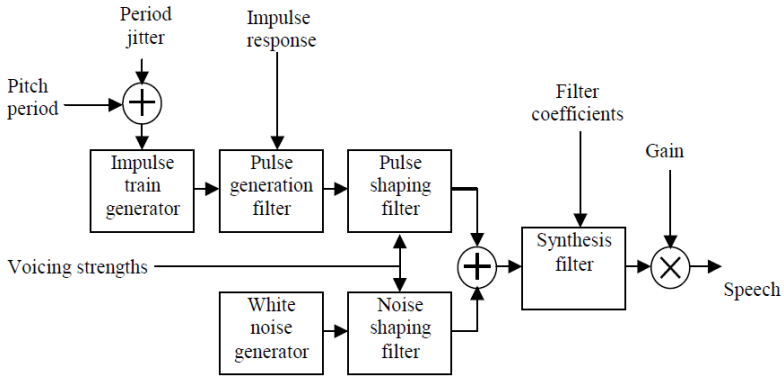

Figure1. The MELP model of speech production

Shape of the excitation pulse for periodic excitation is extracted from the input speech signal and transmitted as information on the frame. The shape of the pulse contains important information and is captured by the MELP coder through Fourier magnitudes of the prediction error. These quantities are used to generate the impulse response of the pulse generation filter (Figure 1), responsible for the synthesis of periodic excitation.

Periodic excitation and noise excitation are first filtered using the pulse shaping filter and noise shaping filter, respectively; with the filters' outputs added together to form the total excitation, known as the mixed excitation, since portions of the noise and pulse train are mixed together.
In Figure 1, the frequency responses of the shaping filters are controlled by a set of parameters called voicing strengths, which measure the amount of ''voicedness.'' The responses of these filters are variable with time, with their parameters estimated from the input speech signal, and transmitted as information on the frame.

2.2. Shaping Filters

The MELP speech production model makes use of two shaping filters (Figure 1) to combine pulse excitation with noise excitation so as to form the mixed excitation signal. Responses of these filters are controlled by a set of parameters called voicing strengths; these parameters are estimated from the input signal. By varying the voicing strengths with time, a pair of time-varying filters results. These filters decide the amount of pulse and the amount of noise in the excitation, at various frequency bands.

In FS MELP, each shaping filter is composed of five filters, called the synthesis filters, since they are used to synthesize the mixed excitation signal during decoding. Each synthesis filter controls one particular frequency band, with pass bands defined by 0–500, 500–1000, 1000–2000, 2000–3000, and 3000–4000 Hz. The synthesis filters connected in parallel define the frequency responses of the shaping filters. Figure 2 shows the block diagram of the pulse shaping filter, exhibiting the mechanism by which the frequency response is controlled. VS 1 to 5 are the voiced strengths.

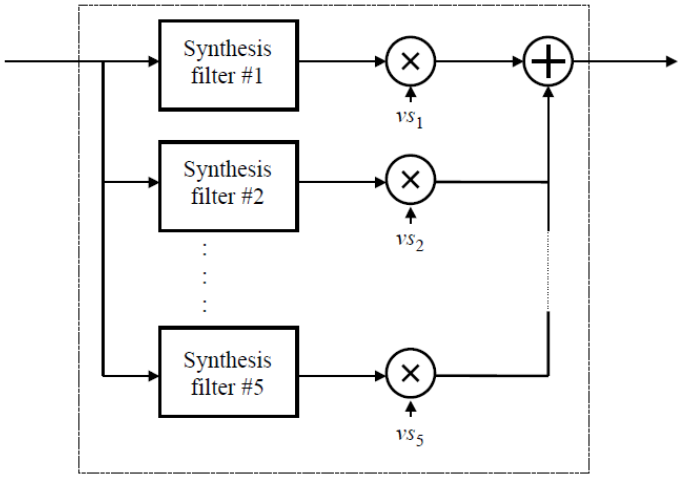

Figure 2: Block diagram of the pulse shaping filter

Thus, the two filters complement each other in the sense that if the gain of one filter is high, then the gain of the other is proportionately lower, with the total gain of the two filters remaining constant at all times.

2.3. 1.2Kbps / 2.4 Kbps MELP Speech Coders [2]

The MELPe or enhanced-MELP (Mixed Excitation Linear Prediction) is a United States Department of Defence speech coding standard used mainly in military applications and satellite communications, secure voice, and secure radio devices. In 2002, the US DoD adopted MELPe as NATO standard, known as STANAG-4591, enabling the same quality as the old 2400 bit/s MELP at half the rate.

The 2.4Kbps MELP algorithm divides the 8Kbps sampled speech signal into 22.5ms frames for analysis, whereas The 1.2Kbps MELP algorithm divides the 8Kbps sampled speech signal into groups of three 22.5ms frames into a 67.5ms super frame for analysis. Depending upon the type of speech present in the signal, inter-frame redundancy can be exploited to efficiently quantize the parameters.

2.4. Bit Allocation

The allocation scheme of FS MELP [1] is summarised in Figure 3. A total of 54 bits are transmitted per frame, at a frame length of 22.5 ms. 2.4 kbps bit-rate is required to transmit 54 bits per frame.

Parameter	Resolution	
	Voiced	Unvoiced
LPC	25	25
Pitch period/low-band voicing strength	7	7
Bandpass voicing strength	4	—
First gain	3	3
Second gain	5	5
Aperiodic flag	1	—
Fourier magnitudes	8	—
Synchronization	1	1
Error protection	—	13
Total	**54**	**54**

Figure 3. Bit allocation for the FS MELP Coder

3. Encryption Algorithms

3.1. Data Encryption Algorithm (DEA)

An encryption scheme computationally secure if the cost of breaking the cipher text generated by the scheme exceeds the value of the encrypted information and the time required to break the cipher exceeds the useful lifetime of the information. In battlefield, the lives of soldiers depend on information and therefore value of the information incalculable. However, the useful lifetime of the information is known and the time required to break the cipher can be calculated.

Assuming there are no inherent mathematical weaknesses in the algorithm, brute-force approach makes reasonable estimates about the time. Brute-force approach involves trying every possible key until intelligible translation of the cipher text into plain text is obtained. Assuming that it takes 1 micro second to perform single decryption, it takes 10.01 hours [3] to break a 56-bit key size DES, and 5.4 x 10**30 years to break a 168-bit key size DES.

In 1999, Triple DES (3DES) was incorporated as part of the Data Encryption Standard and published as FIPS PUB 46-3. 3DES uses three keys and three executions of the DES algorithm. 3DES is very resistant to cryptanalysis and makes the system robust.

3DES processes the input data in 64-bit blocks. 54 bits are required to encode one frame of input speech of 22.5 ms. Remaining 10 bits are utilised for error protection to make the coder more robust.

3.2. Error Protection [4]

In 1950, Hamming introduced the (7, 4) code. It encodes 4 data bits into 7 bits by adding three parity bits. Hamming (7, 4) can detect and correct single-bit errors. With the addition of an overall parity bit, it can also detect (but not correct) double-bit errors.

In MELP algorithm, Forward Error Correction (FEC) is implemented in the unvoiced mode only. The parameters that are not transmitted in the unvoiced mode are the Fourier magnitudes, band pass voicing and the aperiodic flag. FEC replaces these 13 bits with parity bits from three

Hamming (7, 4) codes and one Hamming (8, 4) code. However, no error correction is provided for the voiced mode MELP coder.

The DES/3DES encryption algorithms process input data in 64-bit blocks. 54 bits are allocation for MELP encoded speech frame. Remaining 10 bits are utilised for FEC parity bits for voiced mode with two Hamming (31, 26) codes.

4. Robust Secure Coder (RSC) Algorithm

Figure 4 gives the block diagram of RSC voice coder with 3DES encryption scheme and 10-bit FEC incorporated in its algorithm.

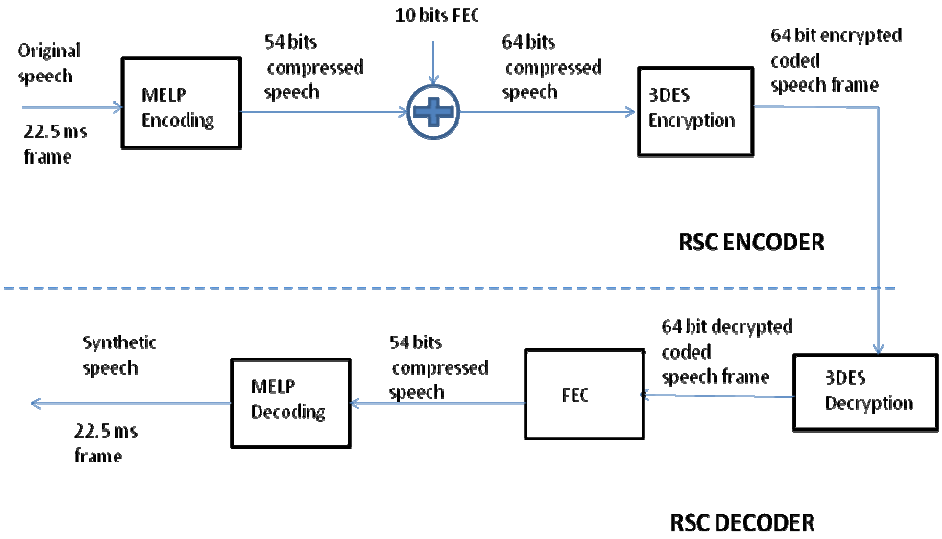

Figure 4. RSC voice processor

Step 1: Data Compression

The original speech is buffered into 22.5 ms frames and passed through MELP coding filter. The 22.5 ms frame coded into 54 bits compressed speech frame.

Step 2: Forward Error Correction

In MELP algorithm, Forward Error Correction (FEC) is implemented in the unvoiced mode only. RSC algorithm uses 10 parity bits to provide error correction for voiced mode also. It uses two Hamming (31, 26) codes. LPC parameters are coded with 25 bits (Refer Figure 3 above). Hamming (31, 26) code is applied to 25 bits of LPC parameter bits and one MSB bit of band pass voicing parameter. The second Hamming (31, 25) code is applied to 5 bits of second gain parameter, 3 bits of first gain parameter, 8 bits of Fourier magnitudes, 7 bits of pitch period, low band voice strength and three LSB bits of band pass voicing parameter. Total 2 bits are corrected over 52 bits of data which cover all parameters except sync bit and aperiodic flag. Thus 10 parity bits are used to correct 2 errors over 52 bits out of 54 bits.

Step 3: Encryption

54 bits of compressed speech, 10 bits of forward error correction are buffered into a 64 bits compressed speech frame. This block of 64 bits is encrypted with 3DES and resulting 64 bit encrypted compressed speech is transmitted to receiver end.

Step 4: Decryption

The 64 bits of encrypted speech is input to 3DES decryption. The resulting 64-bit decoded speech is passed to the next stage.

Step 5: Application of FEC

10 parity bits are used to correct errors, if any. 54 bits of compressed speech is separated and given to MELP Decoding filter.

Step 6: Speech Synthesis

54 bits of compressed speech frame is passed through the MELP Decoder which produces the synthesized speech frame of 22.5 ms.

5. CONCLUSIONS

RSC voice processor uses 64-bit allocation scheme for 22.5 ms frame which would get translated to 2844 bps bit-rate. However, the RSC voice processor is interoperable with existing MELP based communication systems in non-encryption mode. The secure mode can be optionally switched over at the extra cost of 444 bps.

The encryption algorithm will introduce very small delay in processing time. Advances in microelectronics and the vast availability of low cost programmable processors and dedicated chips have enabled rapid technology transfer to product development. Assuming one microsecond for encryption / decryption, the delay added to the processing time is negligible and overall delay would be less than acceptable 150 ms from speaker to receiver and the conversation will not be impaired after switching to encryption mode.

Net-centric communications are accessed by large number of users and therefore, there is a need to provide protection against security attacks and suitable security systems should be introduced to match with the speed of migration to net-centric communications.

In this paper, we briefly reviewed the coding algorithms suitable for secure net-centric communications in the battlefield and suggested a robust and secure voice processor with explicit encryption and error correction features.

REFERENCES

[1] Wai C. Chu, (2003), Speech Coding Algorithms, Wiley Interscience.

[2] John S. Collura, Diane F. Brandt, Douglas J. Rahikka (2002), The 1.2Kbps/2.4Kbps MELP Speech Coding Suite with Integrated Noise Pre-Processing, National Security Agency.

[3] William Stallings, (2009), Network Security Essentials Applications and Standards, Pearson Education

[4] Lann M Supplee, Ronald P Cohn, John S. Collura from US DOD and Alan V McCree from Corporate R&D , TI, Dallas, MELP: The New Federal Standard at 2400 bps.

[5] PENG Tan,CUI Huijuan,TANG Kun (2010); Speech coding and transmission algorithm based on multi folded barrel shifting majority judgment; Journal of Tsinghua University (Science and Technology)

[6] JI Zhe,LI Ye,CUI Huijuan,TANG Kun (2009); Leaping frame detection and processing with a 2.4 kb/s SELP vocoder; Journal of Tsinghua University (Science and Technology)

[7] Arundhati S. Mehendale and M. R. Dixit, (2011), Speaker Identification, Signal & Image Processing: An International Journal (SIPIJ) Vol.2, No.2, June 2011

[8] Jelena NIKOLIC, Zoran PERIC, (2008), Lloyd–Max's Algorithm Implementation in Speech Coding Algorithm Based on Forward Adaptive Technique, INFORMATICA, 2008, Vol. 19, No. 2, 255–270

[9] Chetana Prakash, Dhananjaya N., and S. V. Gangashetty, (2011) "Detection of Glottal Closure Instants from Bessel Features using AM-M Signal," IWSSP 2011, pp 143-146

[10] Sri Harish Reddy M., Kishore Prahallad, S. V. Gangashetty, and B. Yagnanaryana (2011) "Significance of Pitch Synchronous Analysis for Speaker Recognition using AANN Models", Eleventh INTERSPEECH 2010, pp 669-672

Authors

Y Rama Devi received B.E. from Osmania University in 1991 and M.Tech (CSE) degree from JNT University in 1997. She received her Ph.D. degree from Central University, Hyderabad in 2009. She is Professor, Chaitanya Bharathi Institute of Technology, Hyderabad. Her research interests include Image Processing, Soft Computing, Data Mining, and Bio-Informatics. She is a member for IEEE, ISTE, IETE, and IE. She has published more than 25 research publications in various National, International conferences, proceedings and Journals.

Akella Amarendra Babu received B. Sc., degree from SV University in 1973, B. Tech (ECE) degree from JNU in 1984 and M. Tech (CSE) degree from IIT Madras, Chennai in 1988. He served Indian Army for 23 years as Lt Colonel in Corps of Signals and has 12 years of senior project management experience in corporate IT industry. He has research experience on mega defence project in DLRL, DRDO for two and half years and is currently Professor and HOD of CSE department in Progressive Engineering College, Hyderabad. He is pursuing Ph D degree from JNTUA under the guidance of Dr. Y Ramadevi.
His research interests include Speech processing, Information Security, Computer Networks, Telecom and Electronic Warfare. He is a Fellow of IETE.