Implementation and evaluation of low-speed speech compression LPC and MELP algorithms

Duong Phuc Phan, Chu Thi Ngoc Quynh, Nguyen Thanh Ngoc

***Abstract*— Currently, the design of a speech signal compression module that responds to low speed but still ensures signal quality for communication devices is a topic of interest for research. LPC (Linear Prediction Coding) and MELP (Mixed-Excitation Linear Prediction), two common approaches, are extensively researched and used, particularly in the vocoder of radio communication equipment like HF or VHF equipment. Implementing security solutions for these devices are challenging due to the crucial role played by the vocoder in signal processing prior to encryption. The paper first analyzes the two algorithms LPC and MELP and implements them on the current popular hardware technologies ARM and DSP to give an evaluation direction when choosing technology to serve the design of vocoders. There are many different methods of speech quality assessment in this paper, the PESQ method according to the ITU-T P862 standard was used to evaluate the results of the implementation of the proposed algorithms. The results show that the quality of the compression rate reduction of about 2 kbps is still satisfactory, and dedicated DSP technology for better speech quality results is a basis for helping vocoder device developers choose the right hardware technology when designing products.**

***Tóm tắt*— Hiện nay, việc thiết kế module nén tín hiệu thoại đáp ứng tốc độ thấp nhưng vẫn đảm bảo chất lượng tín hiệu cho thiết bị thông tin là chủ đề đang được quan tâm nghiên cứu. Trong đó có 2 kỹ thuật tiêu biểu là LPC và MELP được nghiên cứu và ứng dụng rộng rãi, đặc biệt là trong bộ vocoder của thiết bị liên lạc thông tin vô tuyến như HF hay VHF. Việc triển khai các giải pháp bảo mật cho các thiết bị này là một thách thức do bộ nén tiếng nói đóng vai trò quan trọng trong việc xử lý tín hiệu trước khi mã hóa. Bài báo trước hết phân tích về hai kỹ thuật LPC và MELP, thực thi trên các công nghệ phần cứng phổ biến hiện nay là ARM và DSP để đưa ra một hướng đánh giá khi lựa chọn công nghệ phục vụ thiết kế các bộ vocoder. Có nhiều phương pháp đánh giá chất lượng thoại khác nhau, trong bài báo sẽ sử dụng phương pháp PESQ theo chuẩn ITU P862 nhằm đánh giá kết quả thực thi các kỹ thuật đề xuất. Kết quả cho thấy bộ nén tiếng nói có thể nén xuống tốc độ xấp xỉ 2 kbps mà vẫn đạt yêu cầu và trên công nghệ chuyên dụng DSP cho kết quả chất lượng thoại tốt hơn, điều này là một cơ sở để giúp các nhà phát triển thiết bị vocoder lựa chọn công nghệ phần cứng phù hợp khi thiết kế sản phẩm.**

***Keywords*— speech compression; LPC; MELP; PESQ .**

***Từ khóa*— nén tiếng nói; LPC; MELP; PESQ .**

I. INTRODUCTION

The topic of research on the compression of speech signals on any system of terminals is still an open field, and there will be many different approaches and solutions. Currently, the development of radio security products is being pursued in the field of national security, especially security devices for shortwave and microwave radio. When it comes to designing a speech signal compression module for security devices that need to maintain signal quality while operating at a low rate, numerous algorithms for source coding are available. These algorithms include channel coding, formant coding, and parametric coding. Typical low-speed speech compression algorithms are LPC and MELP coding algorithms [1].

There are usually two approaches to these algorithms. The initial approach involves enhancing algorithms from a theoretical standpoint by providing mathematical proofs and introducing a novel execution model that

---

This manuscript is received on April 07, 2023. It is commented on May 08, 2023 and is accepted on June 05, 2023 by the first reviewer. It is commented on April 26, 2023 and is accepted on June 02, 2023 by the second reviewer.

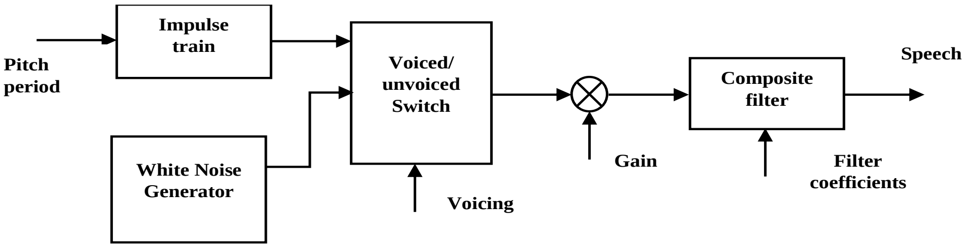


Figure 1. LPC Model of speech production

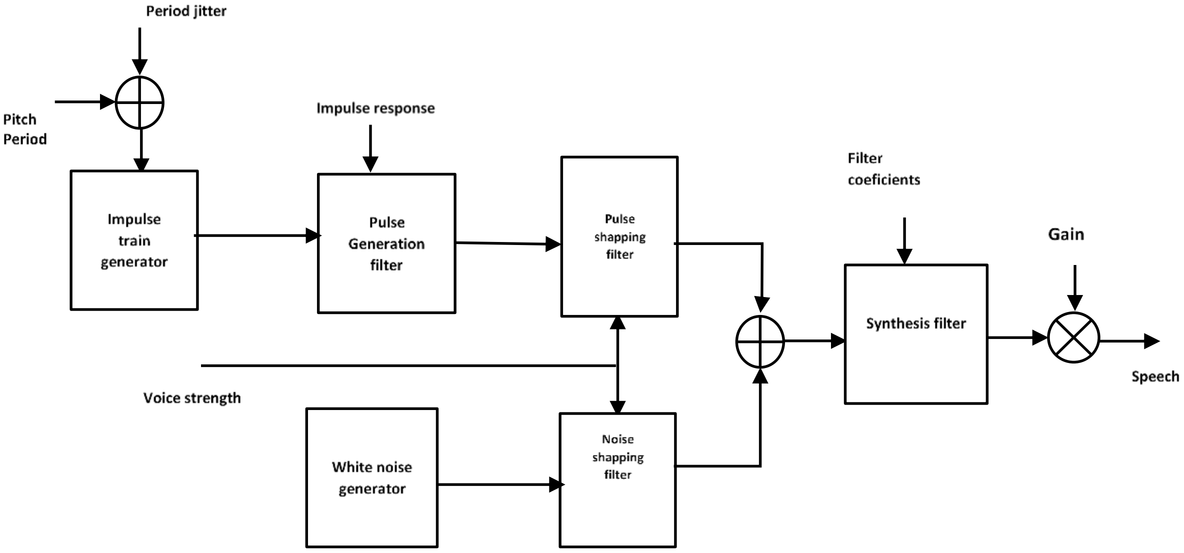


Figure 2. MELP Model of speech production

addresses speed, quality, and latency requirements, such as in [2, 3]. The second is to algorithms, and then evaluate the results, as in some studies [4, 5, 6]. This paper also approaches in the second direction, which is to implement the existing algorithms based on different technologies such as ARM and DSP and provide a solution to evaluate the results based on the PESQ method.

The LPC method [1, 7, 8] is one of the earliest standardized coders that operates at low bit rates and is based on a greatly simplified model for speech generation. The FS1015 LPC coder [1] is a breakthrough in speech coding technology since it can decode speech at a speed of 2.4 kbps while maintaining high audio fidelity. Since then, any algorithm using the LPC model of speech creation has been referred to as “linear prediction coding” with the FS1015 standard serving as its most notable example. Linear prediction coding relies on a highly simplified model for speech production, with the block diagram shown in Figure 1.

The required parameters in this model are: speech recognition, energy calculation, zero overshoot rate, gain prediction, and speech recognizer design. To get around some of LPC's drawbacks, there is a coder called MELP [1, 8, 9]. It utilizes a more intricate speech production model that incorporates additional parameters in order to more precisely capture the underlying dynamics of the signal. The fundamental concept is the creation of a mixed excitation signal, where the "mixing" is the fusion of a filtered periodic pulse sequence with a filtered noise sequence, as input to the synthesis filter. The advantages necessitate an additional computational cost that can only be met by the potent Digital Signal Processors (DSPs) that were first launched in the mid-1990s. It uses a more complex speech production model shown in Figure 2.

The pulse generator will generate pulses with variable periods according to the pitch and oscillate the period; the pulse sequence is then passed through the pulse generator filter and to the pulse shaping filter. Simultaneously, in the

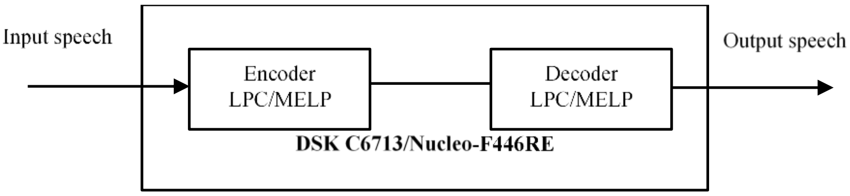


Figure 3. The block diagram of implementing 2 LPC/MELP algorithms on hardwares

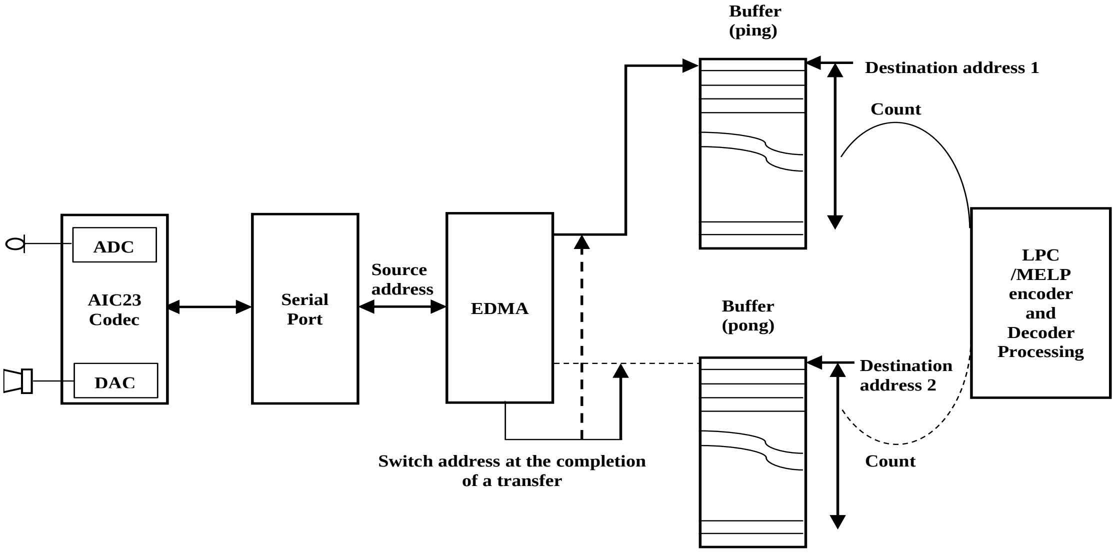


Figure 4. The block diagram describing the implementation LPC and MELP on DSK C6713

second arm of the model, the white noise generator feeds to the noise shaping filter. These two signals are added before passing through the synthesis filter, and the synthesized speech signal after the filter will be changed according to the Gain.

II. IMPLEMENTATION OF LPC AND MELP ALGORITHMS

A. The overall model of implementing

The content of this section will present the implementation of two LPC and MELP algorithms on the ARM Nucleo-F446RE STM32 [10] and DSK C6713 [11] (Figure 3). The input speech is encoded and decoded using LPC and MELP algorithms, respectively, on different types of hardware, and then the corresponding output is obtained for evaluation in the following section.

The block diagram describing the implementation of the speech compression system model on DSP is shown in Figure 4. A microphone is used to feed audio data to the DSP via the 3.5mm connector. AIC23 and AIC23 audio codecs communicate with DSP using two serial ports, McBSP1 and McBSP2. McBSP1 is configured in unidirectional mode to control codec parameters by accessing its setting registers. On the other hand, the McBSP2 is configured as a bidirectional port to transmit audio data to the ADC and from the codec's DAC, which is interfaced with the stereo input and output ports. The DSP uses enhanced direct memory access (EDMA) to transmit and receive audio data through a buffer that is accessible in the C6713's internal memory. Synchronization between EDMA and the C6713 chip is provided by buffer implementation using the Ping Pong caching technique [10, 14].

The process of implementing speech compression and speech decompression is done on an ST chip, the STM32F446RE. The STM32F446RE chip uses 32-bit RISC-based ARM Cortex-M4 technology with a clock speed of 180 MHz, floating-point support, built-in high-speed embedded memory (Flash memory up to 512 KB, SRAM up to 128 KB), redundant SRAM up to 4 KB, and an assortment of advanced I/Os and peripherals connected to the APB bus, two AHB buses, and a 32-bit multi-

AHB bus matrix. All devices provide 3 12-bit ADCs, 2 DACs, a low-power RTC, and 12 general-purpose 16-bit timers.

B. Implement the LPC algorithm

The LPC engineering implementation model is depicted in terms of encryption and decryption in Figure 5 and 6, respectively.

The LPC-10 model offers two levels of speech compression. The first level of compression is achieved by processing the original speech data frame to obtain 10 LPC coefficients, pitch, a power gain constant, and speech/speechless parameters.

TABLE 1. TABLE OF PARAMETERS OF THE LPC-10 ALGORITHM

Parameters	**Value**
Sampling frequency	8000Hz
Data rate	2400bit/s
Frame	22,5ms
Number of bits per frame	54 bits
LPC coefficient	10

TABLE 2. LPC-10 BIT ALLOCATION TABLE ACCORDING TO FS1015 STANDARD

Parameters	Speech signal	Speechless signal
Sound pitch interval	7	7
Power	5	5
LPC	41	20
Synchronized	1	1
Error protection	-	21
Summary	54	54

Comparing the types of coefficients and other parameters with the actual speech data, this processing provides an almost 1:8 compression ratio. This compression comes at the expense of losing the “naturalness” of the speech signal. Second-level compression is achieved by quantizing these parameters. After quantization, the encoded frame consists of only 54 bits, as shown in Table 2, 54 bits per frame were specifically allocated.

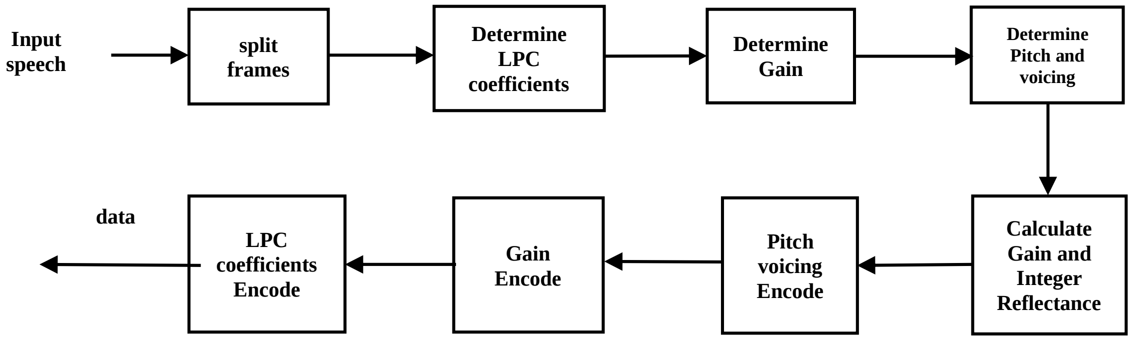


Figure 5. LPC encoder implementation block diagram

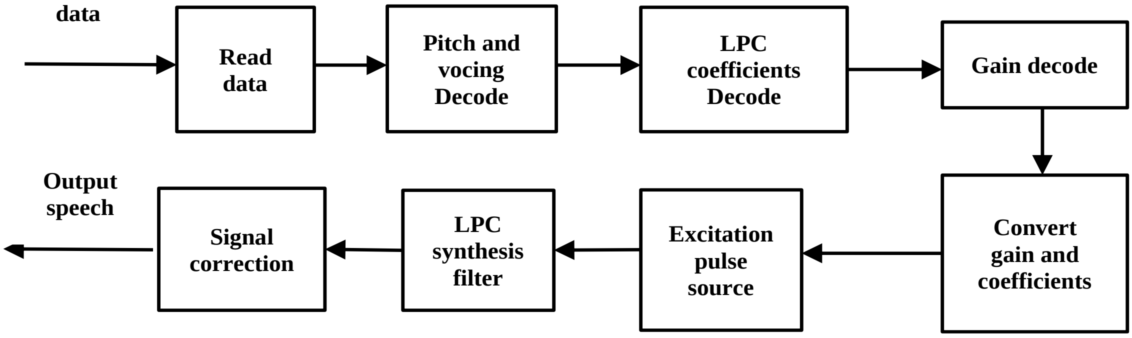


Figure 6. LPC decoder implementation block diagram

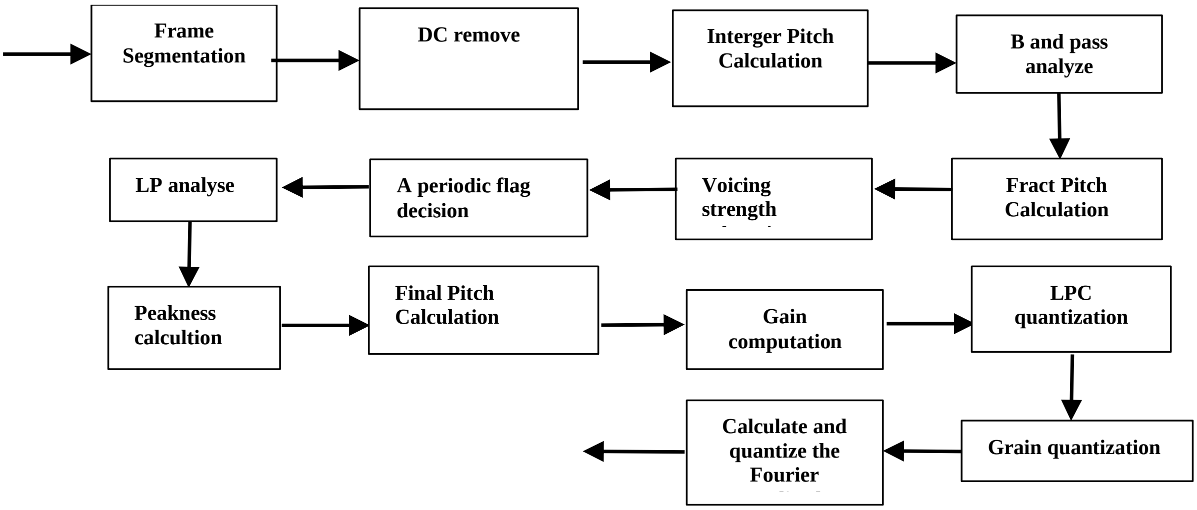


Figure 7. MELP encoder implementation block diagram

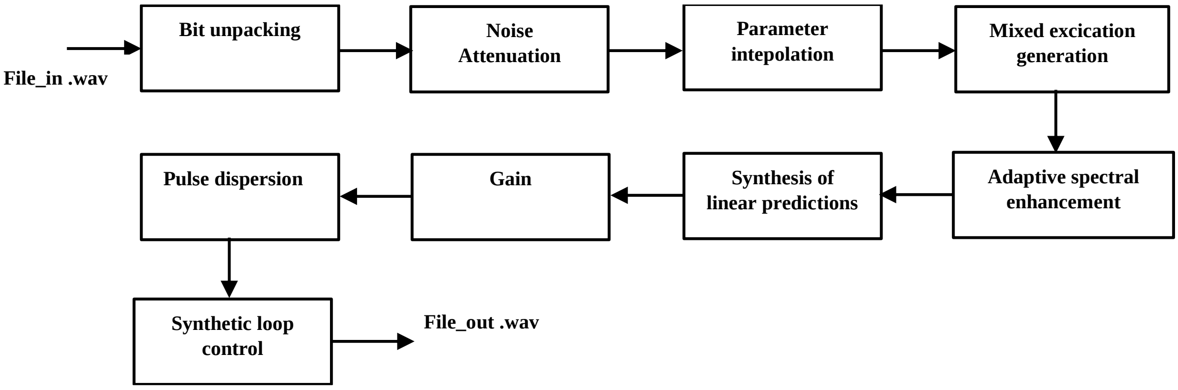


Figure 8. MELP decoder implementation block diagram

C. Implement the MELP algorithm

The MELP engineering execution encoder model is depicted in Figure 7.

The MELP decoder is executed in a sequential stream, and the algorithms are executed sequentially, as shown in Figure 8.

III. EVALUATION OF SPEECH QUALITY AFTER IMPLEMENTING 2 ALGORITHMS (LPC AND MELP)

A. Evaluation of speech signal quality using PESQ

The method of assessing speech quality has been standardized by many standards organizations, such as ITU-T, ETSI, and 3GPP. PESQ is a comparative speech quality assessment method, which is described in ITU-T recommendation P.862 to be used as an alternative to ITU-T recommendation P.861 [12,13]. This method compares the original signal x(t) with the attenuated signal y(t) as a result of the transmission of signal x(t) through the communication system. The output of PESQ is an estimate of the received speech quality of signal y (t).

Figure 9 is a block diagram of the PESQ algorithm. It performs a series of signal delays between the original input signal and the specified output signal; each delay value is calculated for a time interval that has a difference in delay from the previous time segment. For each time segment, the start and end points are determined. A sorting algorithm is based on the principle of comparing the probability of two delays in a time interval with the probability of one delay in that time interval. This algorithm can handle delay changes in both silence and active talk. Based on the set of identified delays, PESQ compares the original input with the sorted output using a sensory model. The key to this process is to convert both

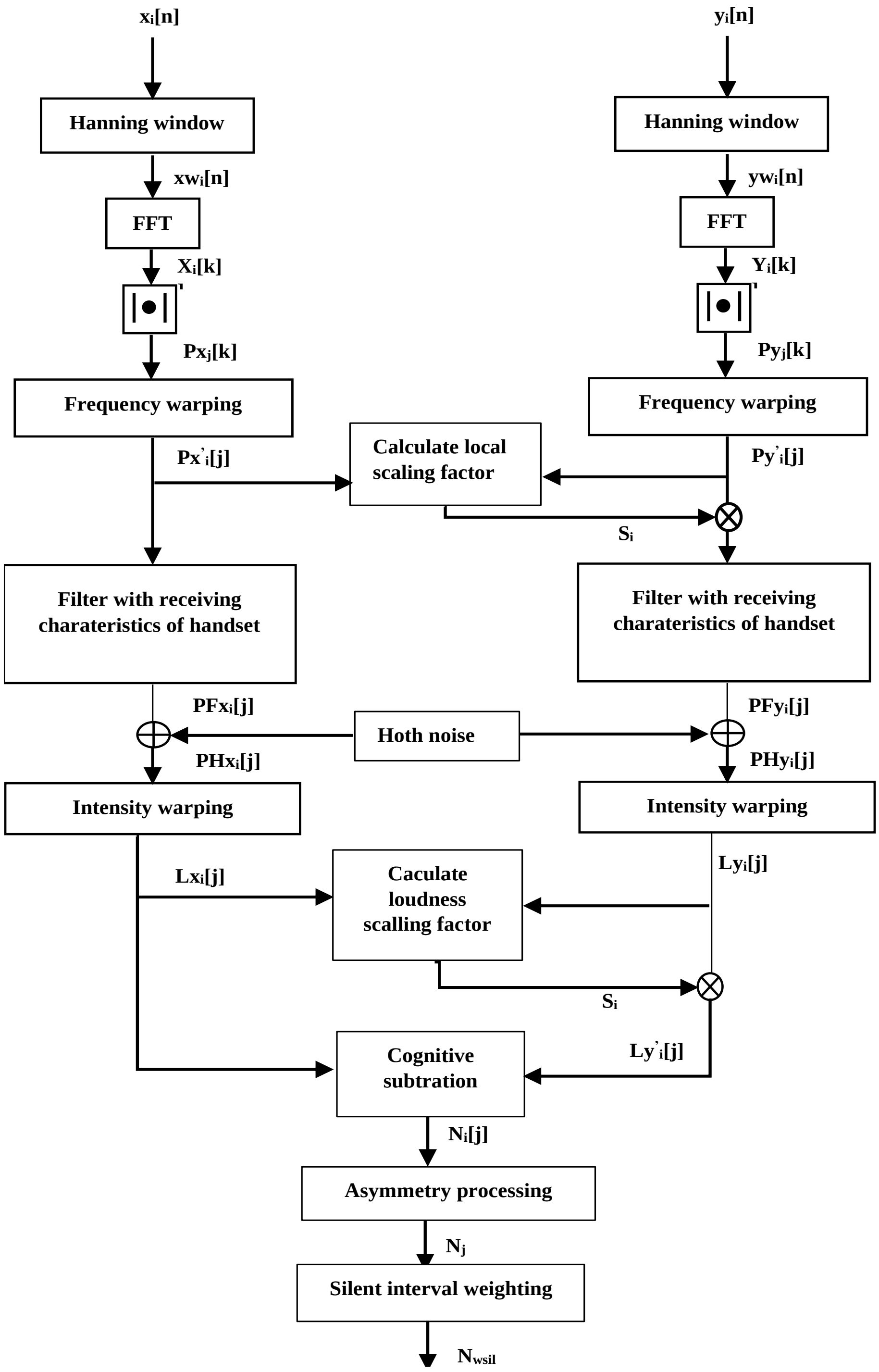
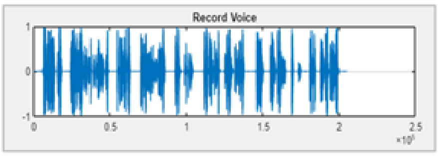


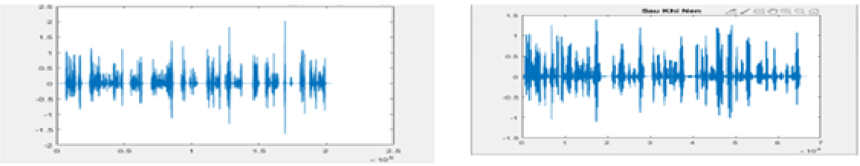
Figure 9. Block diagram of PESQ algorithm

the original signal and the attenuated signal into a representation of the audio signal in the human auditory system that takes into account auditory frequency and pitch. This process is done in several stages: time alignment, signal level alignment to a calibrated listening signal level, time-frequency mapping, frequency warping, and pitch correction.

In PESQ, two error parameters are calculated in the empirical model and combined to estimate the MOS score. A computer model of the subject consisting of a sensory model and a heuristic model is used to compare the output signal with the original signal using alignment information derived from the timing signals in the module timing arrangement.

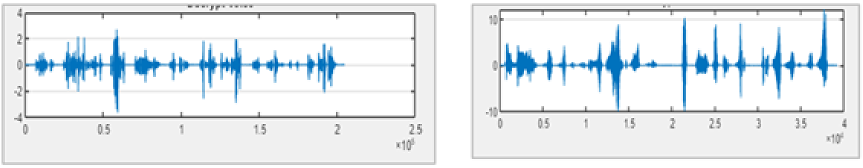


Nam_01.wav



Nam_01_LPCDSP.wav

Nam_01_LPCARM.wav



Nam_01_MELPDSP.wav

Nam_01_MELPARM.wav

Figure 10. Spectrum analysis of input and output signals

B. Figures and tables

The input audio files are pre-recorded from the standard speech over from the radio station, as shown in Table 3.

TABLE 3. AUDIO FILES USED

	File name	**Male/Female**
1	Nam_01.wav	Male
2	Nam_02.wav	Male
3	Nam_03.wav	Male
4	Nu_01.wav	Female
5	Nu_01.wav	Female
6	Nu_01.wav	Female

An approach to the formation of the voice signal informative features of the Vietnamese language on the basis of stationary autoregressive model coefficients is described [12]. The audio files are an 8000 Hz PCM-sampled audio signal with quanta (16 bits per sample). The audio signal passes through the encoders and will be analyzed into 180 sample data frames, analyzed into parameters (LPCs, pitch, voicing strengths, gain, Fourier quantities, aperiodic flags). These parameters are saved as an output structured array. This data array will be the input to the decoders. The decoder's output will be recorded to an audio file with an 8000 Hz sampling rate and 16-bit quantization per sample.

Perform and evaluate the audio files in Table 3 in turn with the execution model as shown. Outputs after execution are evaluated

TABLE 4. PESQ EVALUATION RESULTS WHEN EXECUTING LPC ON DSP AND ARM

Original file	**LPC_DSP**	**LPC_ARM**	**PESQ_DSP**	**PESQ_ARM**
Nam 01.wav	Nam 01 LPCDSP.wav	Nam 01 LPCARM.wav	2.341	2.135
Nam_02.wav	Nam_02 LPCDSP.wav	Nam_02 LPCARM.wav	2.496	2.326
Nam_03.wav	Nam_03 LPCDSP.wav	Nam_03 LPCARM.wav	2.457	2.148
Nu 01.wav	Nu 01 LPCDSP.wav	Nu 01 LPCARM.wav	1.952	1.887
Nu_02.wav	Nu_02 LPCDSP.wav	Nu_02 LPCARM.wav	1.968	1.889
Nu_03.wav	Nu_03 LPC.wav	Nu_03 LPCARM.wav	1.935	1.885

TABLE 5. PESQ EVALUATION RESULTS WHEN EXECUTING MELP ON DSP AND ARM

Original file	**MELP_DSP**	**MELP _ARM**	**PESQ_DSP**	**PESQ_ARM**
Nam 01.wav	Nam 01 MELPDSP.wav	Nam 01 MELPARM.wav	2.341	2.135
Nam_02.wav	Nam_02 MELPDSP.wav	Nam_02 MELPARM.wav	2.496	2.326
Nam_03.wav	Nam_03 MELPDSP.wav	Nam_03 MELPARM.wav	2.457	2.148
Nu 01.wav	Nu 01 MELPDSP.wav	Nu 01 MELPARM.wav	1.952	1.887
Nu_02.wav	Nu_02 MELPDSP.wav	Nu_02 MELPARM.wav	1.968	1.889
Nu_03.wav	Nu_03 LPC.wav	Nu_03 LPCARM.wav	1.935	1.885

TABLE 6. EXECUTION TIME (REAL-TIME IMPLEMENTATION)

	LPC speech coder			**MELP speech coder**		
	Analysis	Synthesis	Analysis + Synthesis	Analysis	Synthesis	Analysis + Synthesis
Avg.no.of cycles (N)	622125	4096350	4718475	1863675	2035125	3898800
Execution time (ms)	2.765	18.206	20.971	8.283	9.045	17.328

by the PESQ method (based on the algorithm shown in Figure 9.

The results of spectrum analysis of 5 audio files include: an original file (nam_01.wav), a file obtained after LPC compression on DSP, a file after LPC compression on ARM, a file after MELP compression on DSP, and a file after compressing MELP on ARM, shown in Figure 10.

Observing the input and output audio signals with the remaining audio files, we see that the output audio signal has the same amplitude and that the amplitude variation is relatively similar. The processed signal is still able to hear the word clearly, even though the signal has been garbled by the algorithm. Below is a table to evaluate speech quality by the PESQ method. The results of the evaluation statistics are shown in Table 4 and 5.

The result files after analyzing and synthesizing MELP are all played back with quite clear results, but there is loss or interference between words. Results with Vietnamese are close to or more than 2 points, which is temporarily acceptable with MELP 2.4 kbps.

Next, the article evaluates the execution time and resources required when implementing these two algorithms on Kit DSP. The Table 6 lists the measured execution times for the two programmers. In comparison to a MELP coder, an LPC coder greatly reduces the time spent analyzing the speech.These two speech coding algorithms' total analysis and synthesis times, however, are quite similar. Both coders are meeting the real-time requirement because their total processing time for each frame is less than the frame duration of 20 ms.

Using the DSP/BIOS configuration file, it is also possible to determine how little memory must be allocated for the speech algorithm to be implemented. The estimated memories correspond to the memory used by the code, i.e., program memory, data memory, and stack memory. The measured memory space needed for the LPC and MELP algorithms is listed in

Table 7. It demonstrates that the LPC coder uses less program memory and data memory than does the MELP coder. The reason for this is that the computational complexity of MELP is higher than that of LPC, but better quality is achieved.

TABLE 7. MEMORY CONSUMPTION

Memory (Byte)	**LPC coder**	**MELP coder**
Program memory	240908	368366
Data memory	14325	18058
Stack memory	640	640
Total	255873	387,064

IV. CONCLUSION

The paper has modeled and implemented two speech compression algorithms, LPC and MELP, then used the PESQ standard to evaluate in detail the quality of speech compression of these two algorithms. Compression quality results when used on dedicated DSP technology give better compression quality than ARM hardware implementation, but this comparison only stops on the two available hardware devices without any difference in similar hardware configuration. The results show that if these two algorithms are effectively implemented, they can compress the speech signal to a very low speed of 2–3 kbps while still ensuring the signal quality; this rate completely meets the criteria for the design of the speech signal in speech encoders in radio security devices.

REFERENCES

[1] Wai C. Chu, “Speech coding algorithms – Foundation and evolution of standardized coders”, 2003.

[2] Min-Jae Hwang and Hong-Goo Kang, “Parameter enhancement for MELP speech codec in noisy communication environment”, arXiv:1906.08470v1, 2019.

[3] Rhutuja Jage, Savitha Upadhya, “Imlementation of CELP and MELP Speech Coding Techniques”, IJARCCE, 2016.

[4] Xiaochen Wu, Longxiang Guo, Yang Yang, et al, “Improvement of mixed excitation linear prediction speech coding algorithm and implementation on DSP system”, The Acoustical Society of America, Vol. 21, 2014.

[5] Nguyễn Thanh Bình, Nguyễn Thành Vinh, Nguyễn Xuân Liêm. “Phân tích, thiết kế tích hợp hệ mã thoại Vocoder dựa trên chuẩn MELP cải tiến phục vụ bảo mật thoại và dữ liệu qua kênh vô tuyến HF chuyên dụng”, Tạp chí Khoa học và Công nghệ (Journal of Science and Technology), Số 115, bài số 10, 11/2016.

[6] Phạm Văn Hậu, Đinh Văn Ngọc, Nguyễn Anh Đức, Thái Trung Kiên, “Real-time implementation of melp vocoder on ti fixed-point TMS320C55X DSP”, Tạp chí nghiên cứu khoa học và công nghệ quân sự, 02-2014.

[7] Mr. Mohit Narayanbhai Raja, Miss. Priyanka Richhpal Jangid, Dr. Sanjay M. Gulhane, LINEAR PREDICTIVE CODING, IJESRT, 2015.

[8] Carl Kritzinger, Low Bit Rate Speech Coding, the University of Stellenbosch, 04-2016, pp 42-50, 52-72.

[9] U. S. DoD, “MIL-STD-3005”, Department of Defense Telecommunications Systems Standard, 1999.

[10] ARM Nucleo-F446RE STM32 datasheet.

[11] DSK C6713 datasheet.

[12] Novikov, E., & Trubitsyn, V. (2020). Applicaton of parameters of speech singal autoregressive models to solve speaker recognition problems. Journal of Science and Technology on Information Security, 10(2), 25-35. https://doi.org/10.54654/isj.v10i2.63

[13] ITU P.862, “Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs”, ITU Recommendation P.862, 2000.

[14] Sen M. Kuo, Bob H. Lee, Wenshun Tian, “Real-time digital signal processing fundamentals, implementations and applications third edition”, Wiley, 2013.

ABOUT THE AUTHOR

Duong Phuc Phan

Workplace: The University of Electro-Communications, Chofu, Tokyo.

Email: duongphucphan@gmail.com

Education: He received his master-degree in Telecommunications Engineering from Posts and Telecommunications Institute of Technology in 2014. PhD student - Department of Computer and Network Engineering.

Recent research direction: Electronics; IoT; Telecommunications.

Cơ quan làm việc: Đại học Điện tử Truyền thông Chofu, Tokyo.

Email: duongphucphan@gmail.com

Quá trình đào tạo: Nhận bằng thạc sĩ Kỹ thuật Viễn thông tại Học viện Công nghệ Bưu chính Viễn thông năm 2014. Nghiên cứu sinh - Khoa Kỹ thuật máy tính và mạng.

Hướng nghiên cứu hiện nay: Điện tử; IoT; Viễn thông.

Chu Thi Ngoc Quynh

Workplace: Academy of Cryptography Techniques.

Email: ctnquynh@gmail.com

Education: She received her master-degree in Telecommunications Engineering from Posts and Telecommunications Institute of Technology in 2012.

Recent research direction: Digital signal processing; Telecommunications.

Cơ quan làm việc: Học viện Kỹ thuật Mật mã.

Email: ctnquynh@gmail.com

Quá trình đào tạo: Nhận bằng thạc sĩ Kỹ thuật Viễn thông tại Học viện Công nghệ Bưu chính Viễn thông năm 2012.

Hướng nghiên cứu hiện nay: Xử lý tín hiệu số; Viễn thông.

Nguyen Thanh Ngoc

Workplace: Academy of Cryptography Techniques.

Email: thanhngoc8818@gmail.com

Education: She received his master-degree in Telecommunications Engineering from Posts and Telecommunications Institute of Technology in 2015.

Recent research direction: Digital electronics; FPGA design; Digital signal processing.

Cơ quan làm việc: Học viện Kỹ thuật Mật mã.

Email: thanhngoc8818@gmail.com

Quá trình đào tạo: Nhận bằng thạc sĩ Kỹ thuật Viễn thông tại Học viện Công nghệ Bưu chính Viễn thông năm 2015.

Hướng nghiên cứu hiện nay: Điện tử số; thiết kế FPGA; Xử lý tín hiệu số.