International Core Journal of Engineering Volume 10 Issue 5, 2024
ISSN: 2414-1895 DOI: 10.6919/ICJE.202405_10(5).0017

DSP-Based Optimization of the G.729 Codec: Strategies for Complexity Reduction in Voice Encoding and Decoding

Zixian Guo*, Qianwen Lin, Shengguang Liu, Yiming Ma, Luyu Guo

Department of Electrical and Information, Shandong University of Science and Technology, Jinan, Shandong 250031, China

Abstract

This study presents an in-depth analysis of the G.729 voice encoding and decoding algorithm alongside the principles of the TMS320C55x Digital Signal Processor (DSP), offering a comprehensive optimization strategy to mitigate the algorithm's complexity. With its eminent 8kb/s transmission rate and superior synthesized voice quality, the G.729 codec stands as a significant standard in voice compression. However, its practical implementation is hindered by its intrinsic complexity, necessitating optimization to ensure real-time processing capabilities. Through the strategic use of Code Composer Studio (CCS) for C and assembly level optimizations, the elimination of unnecessary overflow protections, direct function inlining, and macro definitions, this paper addresses these challenges. Additionally, it underscores the integration of the G.729 codec with DSP technology, demonstrating the creation of cost-effective systems with vast applications in digital communications. The proposed optimization methods significantly enhance the codec's efficiency and applicability, heralding advancements in the technology of digital communication by facilitating improved, reliable voice transmission across diverse platforms.

Keywords

Communication Engineering; G.729 Codec Optimization; TMS320C55x DSP; Voice Encoding and Decoding; Digital Signal Processing.

1. Introduction

In the era of digital revolution, the swift advancement of multimedia information technology and network technology has led to an exponential growth in the volume of information. This surge in data generation has subsequently made the already scarce channel resources even more precious. Faced with the challenge of transmitting a burgeoning amount of information within these limited channel resources, voice compression has emerged as an indispensable tool[1].

In 1996, the International Telecommunication Union (ITU) introduced the G.729 protocol, officially known as the Conjugate-Structure Algebraic-Code-Excited Linear Prediction (CS-ACELP) algorithm. This protocol, with its encoding rate of 8kbps, has been crafted to meet the demands of network communication while maintaining high voice quality. It exhibits a strong adaptability across different application scenarios, marking it as a superior international standard for voice compression[2]. Its exceptional performance has led to its widespread application in various fields, including personal mobile communication and satellite communication, among others[3].

This paper aims to delve into the G.729 protocol, exploring its significance, technological underpinnings, and its broad applications in the field of telecommunications. By providing a comprehensive analysis, this study seeks to underscore the protocol's critical role in facilitating

ISSN: 2414-1895 DOI: 10.6919/ICJE.202405_10(5).0017

efficient information transmission in an era characterized by limited channel resources and growing data volumes.

2. G.729 Codec Algorithm Principles

(1) Voice Coding Principles

The G.729 codec serves as a pivotal technology in digital voice communication, leveraging advanced signal processing techniques to compress voice signals efficiently. The initial phase of processing involves filtering the input analog signals to confine them within the telephony bandwidth. This critical step ensures the preservation of frequency components vital to human speech, thereby optimizing the encoding process that follows.

The subsequent stage involves sampling the signal at a rate of 8kHz, adhering to the Nyquist criterion. This criterion can be expressed as:

$$f_s > 2B$$

where(f_s) is the sampling frequency and (B) is the maximum bandwidth of the signal. For human speech, this ensures the capture of essential characteristics without the introduction of aliasing artifacts. The sampled signal is then converted into a 16-bit linear Pulse Code Modulation (PCM) format, striking an optimal balance between signal fidelity and computational efficiency for the encoder.

A cornerstone of the G.729's encoding strategy is its robust preprocessing module, which employs high-pass filtering to reduce low-frequency noise and applies signal amplitude attenuation to normalize the dynamic range of the input signal. Such preprocessing not only elevates the quality of the encoded signal but also enhances the encoder's stability across different speech inputs.

The encoding process adopts a frame-based approach, segmenting the signal into 10ms frames, each containing 80 samples. This segmentation enables efficient analysis and encoding by organizing the data into manageable units. For each frame, the codec extracts Linear Prediction (LP) parameters, encapsulating the signal's spectral envelope. These parameters are subsequently transformed into Line Spectral Pairs (LSPs), enhancing stability and efficiency in quantization and transmission.

An innovative feature of the G.729 codec is its predictive two-stage vector quantization of LSP parameters, compressing spectral information into 18 bits per frame. This process involves a delicate balance between bit rate and speech quality. The search for the optimal excitation signal, aimed at minimizing the perceptual weighted error, utilizes a synthetic approach to align closely with the original speech's tonal qualities.

The codec intricately produces a reconstructed signal through a synthetic filter, employing complex techniques to derive the residual signal. When processed through a perceptual weighting filter, this residual undergoes spectral shaping influenced by auditory perception models, thereby ensuring the prioritization of perceptually significant components. This step significantly enhances the intelligibility and naturalness of the synthesized speech.

Pitch analysis, an essential aspect of the codec, utilizes autocorrelation to determine the speech signal's fundamental frequency, guided by the formula:

$$R(\tau) = \sum_{n=0}^{N-1} s(n) \cdot s(n+\tau)$$

where $(R(\tau))$ is the autocorrelation function at lag (τ), and $(s(n))$ is the speech signal. This analysis underpins both the adaptive codebook strategy for pitch prediction and the generation of the

ISSN: 2414-1895 DOI: 10.6919/ICJE.202405_10(5).0017

excitation pattern, essential for emulating natural-sounding speech. The codec dynamically adjusts the LP filtering coefficients to match the signal's evolving spectral properties.

Finally, the encoding process culminates in the assembly of all calculated parameters into an 80-bit compressed data frame, efficiently encapsulating the original speech signal's essence. This compact representation, designed for efficient transmission, prepares the encoded signal for decoding and reconstruction at the receiving end, maintaining high voice quality and intelligibility.

(2) Voice Decoding Principles

Upon reception, the G.729 codec undertakes the complex task of transforming compressed data frames back into audible speech. This transformation process, or decoding, commences with the extraction of encoded parameters from the received bitstream. This step reveals the carefully encoded spectral and temporal characteristics embedded within the speech signal.

At the heart of the decoding process lie the Line Spectral Pair (LSP) coefficients. These coefficients undergo interpolation to reconstruct the Linear Prediction (LP) filters for each subframe, a process mathematically represented as follows:

$$LP_{coeff}^{(i)} = \text{Interpolate}\left(LSP_{coeff}^{(i-1)}, LSP_{coeff}^{(i)}\right)$$

where $(LP_{coeff}^{(i)})$ denotes the LP filter coefficients for the (i^{th}) subframe, and $(LSP_{coeff}^{(i-1)})$ and $(LSP_{coeff}^{(i)})$are the LSP coefficients of the adjacent subframes. These LP filters are crucial for defining the spectral envelope of the decoded signal and are updated every 5ms to accurately replicate the original speech's tonal quality.

Following this, the excitation signals are produced by amalgamating the outputs of the adaptive and fixed codebooks, scaled by their respective gains. This is a pivotal step in re-establishing a continuous time signal from the sparse spectral representation, thus closely mirroring the original speech waveform. The mathematical representation of this step can be simplified as:

$$s(n) = \sum_i G_i \cdot CB_i(n)$$

where $(s(n))$is the synthesized speech signal, (G_i) are the gain factors for the (i^{th}) codebook, and $(CB_i(n))$ represents the codebook vectors.

Moreover, the decoded speech undergoes a series of post-processing stages to further refine its quality. These include adaptive post-filtering and high-pass filtering, which enhance the naturalness and clarity of the decoded speech while reducing artifacts introduced during the compression process. Such post-processing ensures a high-quality listening experience.

Through the intricate interplay of these signal processing techniques, the G.729 codec achieves efficient compression and decompression of voice signals. This capability allows for high-quality voice communication over bandwidth-constrained networks, underscoring the codec's significance in the realm of digital telecommunications. Its innovative design and dedicated optimization for speech signals render the G.729 codec an indispensable tool for facilitating clear and efficient voice transmission worldwide..

3. TMS320C55x Architecture

The TMS320C55x represents the forefront of Texas Instruments' (TI) C5000 series, showcasing a 16-bit fixed-point DSP core engineered for optimized performance in digital signal processing. It inherits software compatibility with the TMS320C54x series, yet it leaps forward in code efficiency

ISSN: 2414-1895 DOI: 10.6919/ICJE.202405_10(5).0017

through the adoption of variable instruction lengths. This feature, combined with an enhanced parallel mechanism, significantly improves the efficiency of loops, a critical factor in DSP applications.

This DSP core is celebrated for its trifecta of advantages: low power consumption, high efficiency, and high code density. Such traits make the TMS320C55x an ideal choice for a myriad of applications, particularly in the realms of telecommunications and consumer electronics, where these qualities are paramount.

The architectural design of the TMS320C55x is distinguished by its multi-bus structure. It encompasses one program bus, five data buses, and six address buses. This configuration facilitates a sophisticated data management and instruction flow, essential for handling complex digital signal processing tasks efficiently. The 32-bit program bus is responsible for transporting instruction codes and immediate values from the program memory, utilizing a unified program/data storage space structure. With a 24-bit wide address bus, the TMS320C55x supports an 8M×16bit addressable storage space, providing ample room for program and data storage.

One of the core strengths of the TMS320C55x is its versatile data addressing modes, coupled with a 7-stage pipeline. These features contribute to a powerful and flexible instruction set, notably through the implementation of three-operand instructions that allow for simultaneous processing of three operands. This capability substantially reduces the number of instructions needed for code execution, enhancing overall efficiency.

The TMS320C55x processor core is compartmentalized into four specialized units, each dedicated to a specific aspect of processing:

I Unit (Instruction Buffer Unit): This unit houses a 64-byte instruction buffer and is tasked with decoding instructions, serving as the first step in the execution pipeline.

P Unit (Program Flow Unit): Responsible for controlling the flow of the program, the P Unit executes hardware loops, branches, and conditional transfers, ensuring smooth and efficient program execution.

A Unit (Address-data Flow Unit): This unit generates the access addresses for reading and writing data, incorporating a 16-bit Arithmetic Logic Unit (ALU) to facilitate address calculations and basic data operations.

D Unit (Data Computation Unit): The powerhouse of the CPU, the D Unit, is equipped with a 40-bit shifter, a 40-bit ALU, and two Multiply-Accumulate units (MACs). This configuration allows for complex data computations, including arithmetic and logic operations, essential for DSP tasks.

Through this specialized architecture, the TMS320C55x DSP core delivers unparalleled performance in digital signal processing, adeptly handling the rigorous demands of various applications with efficiency and flexibility.

4. Algorithm Optimization Strategies

The complexity of the G.729 algorithm necessitates code optimization to achieve real-time processing of voice information. Optimization strategies can be approached from two angles: optimizing high-level programming code and simplifying the algorithm while maintaining accuracy requirements.

(1) High-Level Programming Code Optimization

Code Composer Studio (CCS) provided by Texas Instruments offers a C optimizer that facilitates both C-level and assembly-level optimizations. By porting the program code directly to the DSP development platform, the optimizer can enhance code compactness and speed to about 70% of the original. However, this optimization is often insufficient for demanding applications.

Further optimization at the C language level involves several strategies:

1) Eliminating Unnecessary Overflow Protection: In fixed-point C programming for G.729, every operation is followed by an overflow check, especially in basic operations used frequently. Not all operations risk overflow; thus, it's efficient to include overflow protection only where there's a significant risk. Typically, analyzing the range of data values can identify potential overflow

scenarios. For ambiguous cases, extensive testing with actual voice samples can verify the necessity of overflow protection.

2) Direct Function Inlining: Inlining functions can save the overhead of function calls and, once inlined, allows the CCS C optimizer to further optimize the function and its contextual code. However, direct inlining increases code length, particularly when inlined functions are called multiple times. Therefore, for smaller functions that are infrequently called, inlining should be considered within the permissible code size.

3) Macro Definitions: In the fixed-point C program for G.729, basic operations are executed through function calls, which, while standardizing program design, significantly slows execution. Replacing these frequently called functions with macro definitions can reduce overhead without sacrificing program readability, thus accelerating computation. Although this may increase code size, it remains within the acceptable limits for the C55x chip's storage space.

Despite these C-level optimizations enhancing efficiency to some degree, the inherent limitations of C code mean that its performance significantly lags behind hand-optimized assembly. Additionally, controlling hardware directly with C is challenging, making C suitable primarily for algorithms that are less complex and have less stringent real-time requirements. For complex, real-time algorithms like G.729, assembly-level optimization is essential.

(2) Assembly-Level Optimization

The assembly instruction set of the C55x DSP chip series supports a wide range of signal processing algorithms and offers various efficient instructions that maximize the use of DSP hardware resources. These instructions enable high code efficiency and speed, ideal for computation-intensive and real-time critical sections of the algorithm. However, assembly programming is complex, with poor readability and portability, making modifications and upgrades challenging.

To leverage the advantages of DSP software and hardware resources fully, mixed programming with C and assembly is often necessary. There are four main methods for mixed programming:

1) Separate C and Assembly Development: Independently write C and assembly programs, compile and assemble them separately, and then link them together. This approach is common but requires adherence to the fixed-point C compiler's function calling and register rules to avoid disrupting the C module's runtime environment.

2) Embedding Assembly Statements in C Code: This method is straightforward, allowing for clear structure and convenient debugging. C manages program entry and exit points, enabling hardware control functions that are difficult to achieve in C alone. However, modifying C variable values directly with embedded assembly can disrupt the C environment.

3) Using Internal Functions for Direct Assembly Calls: Internal functions composed of assembly statements recognized by the C compiler can easily replicate several C statements' functionality, allowing them to be called like regular C functions within the C environment.

4) Manual Optimization of Assembly Generated from C: CCS's C compiler can convert C programs to assembly in one step, but manual optimization often yields better results. This approach requires thorough knowledge of the C compiler and careful modifications to avoid conflicts with other parts of the program.

If these optimizations do not meet real-time requirements, simplifying the algorithm while maintaining accuracy becomes necessary. For example, in G.729, the transformation of LP coefficients to LSP coefficients, quantization of LSP coefficients, and pitch delay search employ multi-level search or quantization strategies. Simplifying the initial search process in multi-level queries or quantizations can reduce the total search time, such as coarse-graining the initial range in pitch delay determination to decrease computational load.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202405_10(5).0017

5. Conclusion

This paper has delineated the foundational principles of the G.729 algorithm, highlighting its primary advantages such as a transmission rate of 8kb/s and superior synthesized voice quality. These benefits position the G.729 as a mainstream standard in voice compression encoding. The optimization methods proposed herein address the challenges posed by the algorithm's inherent complexity, enhancing its practicality and efficiency.

The integration of the G.729 codec with DSP technology paves the way for creating highly cost-effective systems. Such systems are poised for widespread application across various fields, particularly in digital communications, where they promise significant improvements in performance and quality. The broad applicability and promising prospects underscore the importance of continuous refinement and optimization of the G.729 codec, ensuring its adaptability and relevance in the evolving landscape of digital technology.

The advancements in G.729 codec optimization not only contribute to the technical domain by enhancing algorithm efficiency but also by broadening the codec's applicability. As digital communication networks continue to expand and diversify, the demand for efficient, high-quality voice compression standards will remain paramount. The work presented in this paper serves as a cornerstone for future research and development efforts aimed at meeting these demands, ultimately facilitating smoother, clearer, and more reliable voice communication across the globe.

References

[1] Shen Min, Deng Yibing, Zheng Jianhong, et al. Principle of DSP and its Application in Mobile Communication [M]. Beijing: Posts and Telecommunications Press, 2001:8-9.

[2] Li Xiao-wen, LI Gui-yong, Chen Xian-liang, et al. Third Generation Mobile Communication System, Signaling and Implementation [M]. Beijing: Posts and Telecommunications Press, 2003:109.

[3] 3GPP TS 25.212 V6.2.0 (2004-12), Technical Specification; 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Multiplexing and channel coding (TDD) (Release 6).