Transcoding of Voice Codecs G.711 to G.729 and Vice-versa Implementation on FPGA

C. Varun[1*], RS. Dudhe[2]

[1]School of Engineering and IT, Manipal University, DIAC, Dubai
[2]School of Engineering and IT, Manipal University, DIAC, Dubai
Corresponding Author: varunc2424@gmail.com

Abstract –Now-a-days more applications and services provided over the internet such as e-mail, file sharing, e- commerce, etc., helps people from all over the globe to exchange data, do business, and communicate voice and videos in a simple way. For this tremendous growth in web is achieved with the development of Voice over Internet Protocol (VoIP). This protocol is a new model of telephony service, in which people make use voice over the Internet to help communicate with each other without the access of Plain Old Telephone Service (POTS). VoIP provides high-quality services that greatly depend on the delay between capturing the voice data and the playback of the voice data. Transcoding voice calls between different networks and end-point gadget is a vital task. G.711and G.729 are VoIP codecs accessible in most of the events. A codec is chosen by the customer based on its quality, power requirements, bandwidth utilization, and tolerance to network conditions. Plentiful VoIP hardware, switches, and media gateways support variety of codecs, the issue emerges when there is a need to change over starting with one codec then onto the next. The call is initiated with G.711 and the end network tolerates with G729, and the service provider is confronted with the test of changing over to finish the call. For this transcoding is a preferred method, which is simulated using Quartus II tool and is implemented in Altera Cyclone V FPGA kit. Transcoding permits gadgets, similar to the IVR platform and the cell phone in the case, to communicate with each other even when they support different codecs.

Keywords – G.711, G.729, Transcoding, Voice CODEC.

I. Introduction

VoIP technology usage is not only restricted to organizations but also for personal usages such as Skype and Google Talk. Compared with conventional network such as Integrated Switched Digital Network (ISDN) which is being used for Public Switched Telephone Networks (PSTN), the VoIP technology has a sort of voice quality related issues, bandwidth and the network tolerance [1]. For the successful deployment of VoIP, various strategies of implementation will affect the overall performance of the finished or end product in various aspects such as user agent mobility, end-to-end voice communication quality, system availability.
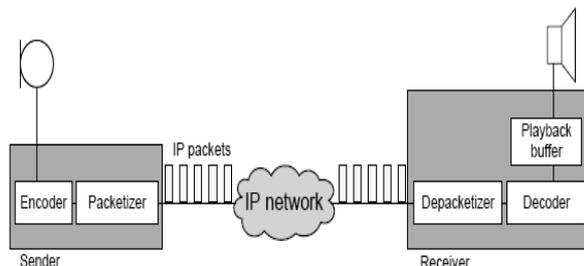

Figure 1 Voice over IP System

Figure 1 shows a VoIP system in which the caller section voice signal is sampled from a microphone. To save the bandwidth voice signal is coded by an encoder. The coded bit stream is packed into IP packets and is transferred along the IP network. On the receiver the received IP packets are unpacked from network headers and IP. Then the voice data which is in coded form is decoded to obtain samples of the voice signal. These samples are placed into a buffer for playback [2].

On seeing protocols used for IP telephony, Session Initiated Protocol (SIP) is an important protocol of choice. But ITU-T H.323 can be considered as an alternative choice for the same application. Another factor taken into contemplation is the option between voice codecs for VoIP communication. The most wide spread codecs for VoIP are, G.729A,G.711 andG.723.1.These codecs operates at different bit rates. Their encoding-decoding delay is not similar. This delay will contribute to the overall end-to-end delay and finally leads to the VoIP Quality of Service (QoS) [3].Codecs determine the sound quality and bandwidth used for communication in which the voice traffic is encoded and encapsulated. Codecs help to determine how the communication has to be put into small packets and sent into a network. Each codec is implemented using different packetization interval times that are a different number of frames per packet. This paper implemented with the best suited codec for VoIP which uses a transcoding gateway that convert G.711 to G.729 and vice-versa.

II. System Design

Transcoding is the transformation of data from one format to another format. VoIP transcoding particularly is the transformation between one computerized representation of voice, or "codec," and another. This is used when two communicating IP-based telephones, or end-point gadgets, have no similar codec that they both support to enable voice communication [4]. Transcoding of voice calls is especially requesting on the grounds that it must occur continuously, supporting quick necessities of comprehensible human conversation in which delays in the communication is approximately 150 milliseconds.

This blocks clear communication and make a telephone discussion troublesome. Transcoding is time consuming under a variety of conditions. Transcoding can commonly be required for applications that cross various networks, prominently including worldwide calling, broad organizational networks, and Open Source and wireless applications. Transcoding is likewise regularly required for call recording, tone detection and play announcements or tones. Other than various system traversals, another key sign for transcoding necessities is bandwidth scarcity, for example, in instances of remote interconnection, upstream Digital Subscriber Line (DSL) and some universal situations.

For instance, a remote end-point may utilize G.729 [5] or another high-compression codec to save bandwidth while the system backbone is utilizing G.711 [6], along these lines requiring transcoding. With progressively complex interconnection between networks, adaptable transcoding usefulness is progressively fundamental. Particularly amid the present time of far reaching move to all-IP networks, VoIP suppliers, in the same way as other others, should definitely course calls over different systems and cross various IP networks; expanding transcoding necessities.

The various codecs are determined by the ITU-T standards. Codecs have execution and effect which is diverse on the voice quality because of various degrees of compression. Codecs determine the sound quality of the call. High level of compression will cause higher compression delay and will expand loss sensitivity compared with codecs having no or low compression. In opposition to this, codecs whose compression levels are high will have less bandwidth requirements, and accordingly have a better execution during a system blockage circumstances. Hence, it is important to choose the proper codec to get best voice quality with the most minimal bandwidth requirements. 64 Kbit/s of sound transmission capacity and gives great quality level. The G.711 codec uses no compression, it has sampling rate of 8Khz, it needs 64 Kbit/s of audio bandwidth and will provide good voice quality level. The codecG.729 is computationally perplexing, yet gives remarkable transfer speed reserve funds. It has compression of 8:1 and will need only 8 Kbit/s of audio bandwidth. Figure 2 shows the transcoding of G.711 to G.729 and vice-versa.
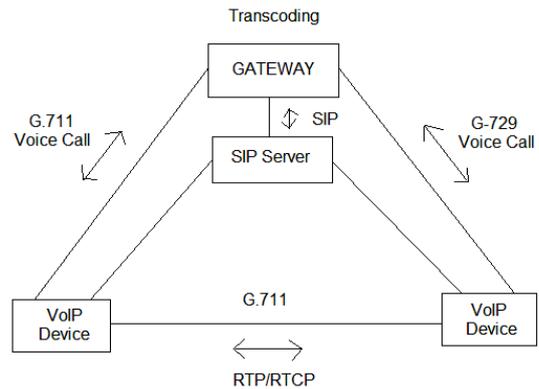

Figure 2 Transcoding of G.711 to G.729 and vice-versa

SIP [7] – is a control protocol operating in the application layer that can modify, set up and terminate multimedia sessions. Web communication calls is an example of this application. Also SIP welcome members to sessions which now exist. Multicast conference calls are an example for this. Media can be either evacuated or added to an officially existing session. SIP additionally supports redirection services which supports personal mobility. This mainly operates between gateway and SIP server.

SIP Server – is the fundamental part of an IP PBX system. It manages the setup of all SIP calls in the network. The SIP server is often referred to as a Registrar or SIP proxy. The SIP server efficiently handles call setup and the call tear down mechanisms. This doesn't really transmit or get any audio information. All the functions are provided using the media server using the Real Time Transport Protocol (RTP).

Gateway - is a network node that connects two networks using two different protocols. Gateway is also known as protocol translators, impedance coordinating gadgets, rate converters, fault isolators and signal translators, as important to give framework interoperability. The functions of a gateway are more complex than that of the switch or router as it conveys utilizing more than one protocol. When both the VoIP gadgets are compatible with each other, both are utilizing G.711 then the transmission of information streams happen from specifically between devices utilizing RTP and RTCP. Though if both the gadgets are working on an alternate codec like one gadget working on G.711 and other working on G.729 then the information streams are transmitted to the gateway wherein the transcoding happens and voice packets are converted over to required format to support the respective VoIP device.

RTP –is utilized to exchange multimedia data and RTP Control Protocol (RTCP) is utilized to periodically send control data and QoS parameters. A codec is utilized as a system to maximize the transfer speed/bandwidth by compressing the information of the call and afterward decompressing it when the call is delivered.
In gateway network trancoding of G.711 to G.729 and vice-versa support is provided by hardware accelerated Digital Signal Processor (DSP) engine either by using DSP

platform or Field Programmable Gate Array (FPGA) platform. This paper describes the development and implementation of G.711 to G.729 trancoding and vice-versa on FPGA.

G.729 – provide good level of call quality at a low bit rate of 8 Kbps, wherein more calls can be got however assigned data transfer capacity if G.711 codec is utilized. One thought to make before actualizing the G.729 codec is that it takes little amount of bandwidth to transmit call information and requires considerably more CPU handling time to make a call. Because of this imperative a considerable lot of the accessible VoIP telephones can deal with just a single call at once because of processing power identified with G.729 calls. Thus transcoding is required at VoIP empowered gateways and gadgets.G.729a is a hybrid speech coder which makes use of Algebraic Code Exited Linear Prediction (ACELP). Figure 3 shows the architecture of G.729 codec.
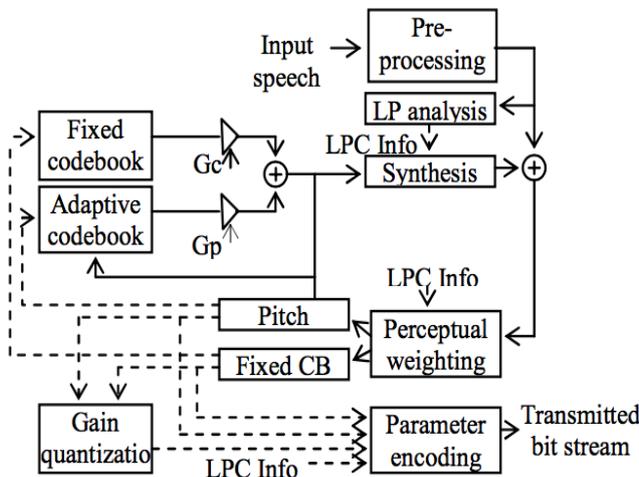

Figure 3 G.729 Architecture

CS–SCELP Encoder – Figure 4 shows the CS-SCELP encoder. The scaling and the high pass filtering of the signal happens in the pre-processing block. The signal which is pre processed is considered to be the input signal for each and every further examinations/tests. Once per 10 millisecond frame LP test is done to ascertain the LP filter coefficients. The excitation signal is selected making use of analysis by synthesis search methodology in which the error between the first and the reconstructed speech is made minimal based on perceptually weighted distortion measure. This is acheived by filtering the error signal with a perceptual weighting filter, which as coefficients derived from the quantized LP filter. The excitation parameters will be determined per sub frame of 5 millisecond each. A open-loop pitch delay will be estimated once per 10 millisecond frame based upon the perceptually weighted speech signal. Later subsequent operations will be repeated for each sub frame. The target signal is computed by filtering the LP residual through the weighted synthesis filter. The impulse response of the weighted synthesis filter is computed. Closed-loop pitch analysis is then carried out, utilizing the target and impulse reaction, via searching around the

estimation of the open-loop pitch delay. A partial pitch delay with 1/3 resolution is utilized. Pitch delay is usually encoded with 8 bits in the principal sub frame and is differentially encoded with 5 bits in the second sub frame. The objective/ target signal is upgraded by subtracting the contribution made by the adaptive-codebook and this new target, is made use in the fixed-codebook search to locate the ideal excitation. An algebraic codebook consisting of 17 bits is utilized for the fixed-codebook excitation. The increases of the fixed-codebook and adaptive contributions are vector quantized with 7 bits. At last, the channel recollections are upgraded utilizing the decided excitation flag.
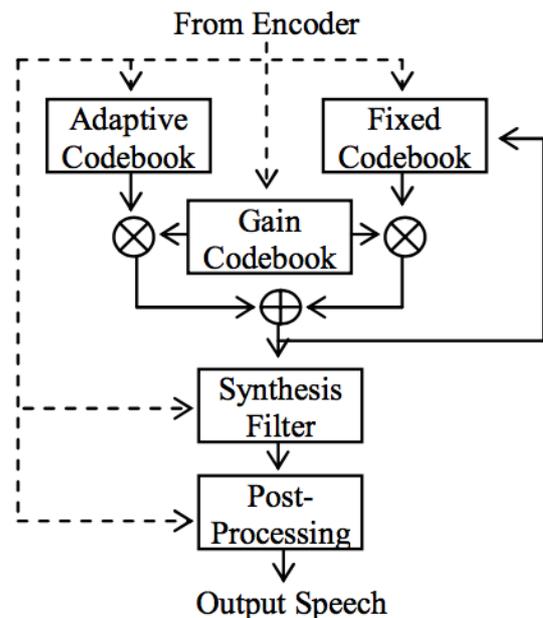

Figure 4 CS-SCELP Encoder

CS-ACELP Decoder –The parameter indices shall be extricated from the bit stream which is received. Then these parameters will be decoded to acquire the coder parameters relating to a 10 millisecond discourse outline. These parameters are the two fixed-codebook vectors, LSP coefficients, two fractional pitch delays and two sets of adaptive and fixed- codebook gains. Then the LSP coefficients are then introduced and changed over to LP channel coefficients for each and every sub frame.

G.711 – The sampling frequency of G.711 is 8 KHz, with 64 Kbit/s bit rate and 0.125 ms delay. There are two sorts in this codec, U-Law and A-Law. U-Law is particularly utilized in Japan and North America, A-law is utilized by the rest of the world. G.711 will give the best of call quality for VoIP as it utilized with no compression and subsequently the call quality is great as in ISDN telephone. This kind of codec is upheld/ supported by most VoIP gadget. G.711 passes sound signals within range 300–3400 Hz and samples them at the rate of 8000 samples per second. The main concept behind operation of G.711 is Pulse Code Modulation (PCM). The voice signal or human speech is in analog format which needs to be converted into

digital format for the purpose of processing. The three main steps involved in the PCM are sampling, quantization, and coding.

Sampling is the process of encoding analog signal into digital form by reading its levels at precisely spaced intervals of time. Quantization is the process of converting amplitude of the infinite samples to finite number of discrete values. There are two types of quantization uniform and non/uniform quantization. In uniform quantization all the intervals have same width. In uniform quantization type the noise is independent of the sample amplitude. Coding is a process of representing the quantized sample by a binary number (1 or 0).Basically in telephony to represent each and every possible samples values (for example for G.711 or µ law and A law)we require 256 intervals of quantization. Hence 8 bits are needed to represent all the intervals ($2^8 =$ 256). The device where quantization and coding happens is the encoder.

III. Simulation Results

Transcoding core of G.729A codec performs coding and decoding of 16-bit LPCM sound examples as per ITU-T G.729A standard. When coding, the core/center takes as input information packet made out of 80 16-bit LPCM sound samples and outputs compacted information packets comprising of 5 16-bit words. When decoding, the core takes as compacted information bundles comprising of 5 16-bit words and yields information parcels made out of 80 16-bit LPCM sound specimens. The core can work in both half-duplex and full-duplex modes. The core can deal with different channels by intermixing processing of their information packets, corruption of state information identified with every channel is avoided from by saving them after processing a packet that belongs to a channel, and storing them back before handling the next packet that belong to the same channel. The most extreme number of channels which can be intermixed is constrained just by the core operating frequency. Channel state saves/reestablishes operations that are not required for single-channel operations. All source code records are composed in synthesizable VHDL dialect.

The modules are implemented by connecting self-test module clock and reset inputs to the board FPGA clock and CPU user reset pins and to connect self-test module using Altera Cyclone V FPGA board. If the conversion is successful, the LED 0 on the board goes ON and if the conversion is unsuccessful LED 1 goes ON. Cyclone V has 77 K programmable logic elements, with 4884 Kb memory, 6 fractional PLLs and 63.125 GHz transceivers. The executed modules are single data port, codec top level, sub program controller, CPU, memory initialization file, sub program controller, loop control stack management logic, instruction fetching logic, instruction fetch Q, instruction decoder, pipe line execution logic, pipe line stall logic, pipeline A decoder, pipeline B decoder, branch/jump execute logic, load store unit, load unit, adder unit, pipe B

adder/subtractor unit, 2-cycle multiply unit, shift unit, and the logic unit.

Figure 5 to 8 shows the timing diagrams obtained after the codes of the core module is compiled, simulated, when RLT simulation for the respective code is run on Quartus II.
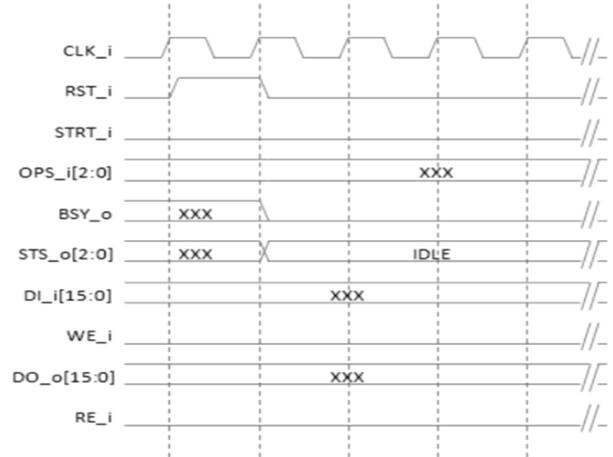

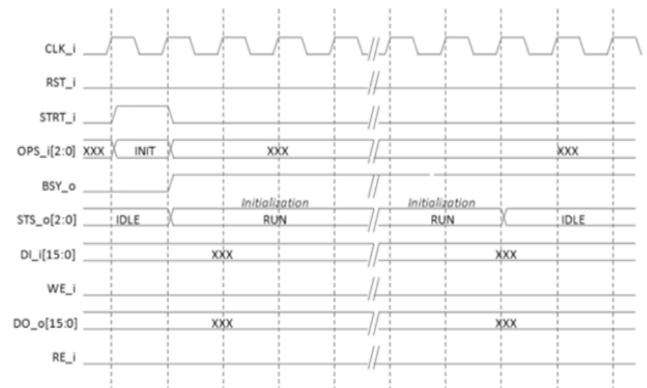Figure 5 Core Reset Timing Diagram
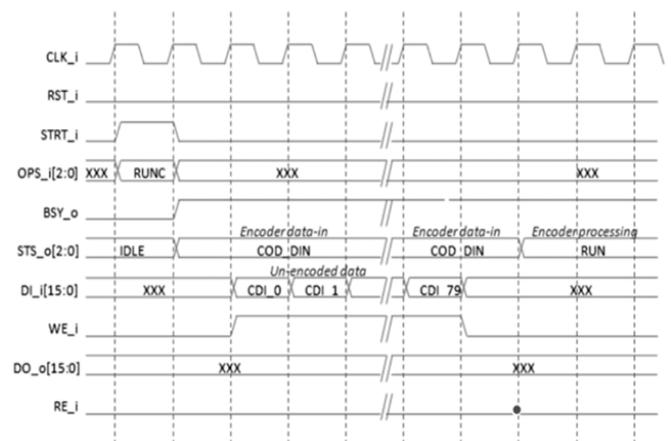

Figure 6 Initialization Timing Diagram

Figure 7a Encoding Timing Diagram
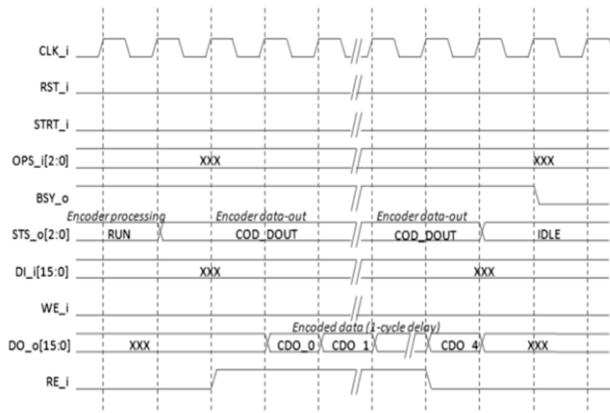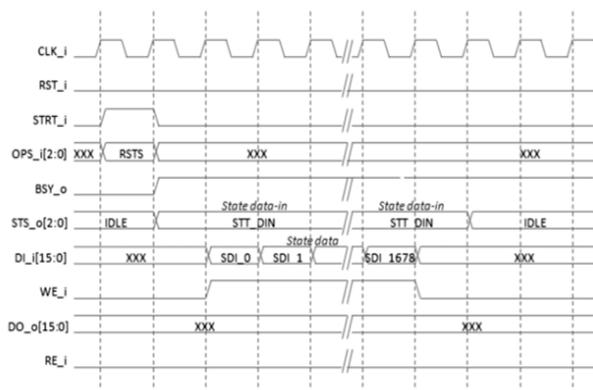

Figure 7b Encoding Timing Diagram

Figure 8 Channel State Restoring Timing Diagram

IV. Conclusion

In this paper, voice transcoding from G.711 to G.729 and vice-versa is implemented using Altera Cyclone V FPGA board. FPGA hardware is needed to increase the processing power of conventional instruction driven DSP chips, while maintaining the ability to upgrade as well as flexibility of software implemented DSP algorithms. The concept of transcoding is designed to cater the needs in military applications considering the cost and reliability of the system. FPGA is a better choice than the DSP because it reduces the overall cost of manufacturing. FPGA improve the performance levels in the order of magnitude that is with architectures having pipelined data flow. Data will flow from one processing unit to the other with minimal signal loading and less overhead for data fetching or instruction. Simulations executed with Quartus II clearly give the predicted output. Transcoding will help the military operating in terrain regions especially when low bandwidth communication is needed or when communication is being made on a weak network.

REFERENCES

[1]   F. D. Rango, M. Tropea, P. Fazio, S. Marano, "*Overview on VoIP: Subjective and Objective Measurement Methods*", IJCSNS, Vol. 6, No.1, pp.34-41, 2006.

[2]   M. Narbutt, L. Murphy, "*Improving Voice over IP Subjective Call Quality*", IEEE Communications Letters, Vol.8, Issue.5, pp. 308-310, Mar 2004.

[3]   R.V. Pawar, S.S. Mahajan, "*Performance Analysis of Wireless Sensor Network Using Cognitive Radio Concept*", International Journal of Computer Sciences and Engineering, Vol.2, Issue.4, pp.136-137, 2014.

[4]   L. Malfait, J. Berger, M. Kastner, "*P.563 – The ITU-T standard for single-ended speech quality assessment*", IEEE Transaction on Audio Speech and Language Processing, Vol. 14, No. 6, pp. 1924-1934, 2006.

[5]   Noor M. Sheikh, KI. Siddiqui, MW. Khan, SF. Tajammul, M. Ashraf, "*Real-time Implementation and Optimization of ITU-T's G.729 Speech Codec running at 8 kbits/sec using CS-ACELP on TM-1000 VLIW DSP CPU*", IEEE International Multi Topic Conference, Pakistan, pp. 23-27, 2001.

[6]   M. Kaur, S. Kumar, S. Sharma, "*DoS Flooding Attacks against SIP based VoIP Systems- a Survey*", International Journal of Computer Sciences and Engineering, Vol.2, Issue.5, pp.112-117, 2014.

[7]   P.N. Kadam, R.R. Gangarde, "*To Evaluate the Performance based on Delay and Jitter of Wired VoIP using a network of assorted number of Routers and Hosts.*", International Journal of Computer Sciences and Engineering, Vol.5, Issue.3, pp.11-15, 2017.