Bitrate Scalable Speech Codec for IP Telephony and IP Broadcast Applications

Ramesh Krishnan, *Ayyanar Packiaraj* and Dr. *Singaram Jayakumar*
Epigon Media Technologies Pvt Ltd,

Abstract:

Emergence of Internet Telephony has presented new challenges and opportunities in Speech Codec design and Development. Key challenges are quality of speech, usage of available low cost DSP/ARM and inbuilt error resilience. For instance, bandwidth of Internet connection is a non stationary parameter and thus the same is expected to vary over time. Devices connected to such a network need to accommodate or able to withstand non-stationary functional feature of bandwidth availability. Components like speech codec, which is prime part of internet telephony application, need to cater or fulfill speech quality requirements. In particular speech quality needs to be improved to match *existing speech quality in conventional telephone* systems. This work introduces a new bitrate scalable speech codec, which is more suitable to meet challenges posed by Internet Telephony.

Bitrate scalability enables each user to receive the best possible speech quality given by the current network condition. More importantly, speech codec is designed to use wideband bandwidth and provide good quality when bandwidth is high. On the other hand designed speech codec provides speech quality of narrow band speech codec when bandwidth is low. Rate converter tool is less complex and the same can be used in internet gateway for rate conversion. Mentioned rate converter tool is critical for heterogeneous network in which speech codec related data is traveling. One way to meet bandwidth requirements is to make encoder itself to produce compatible bitrate for a given bandwidth, but this is not interesting when bandwidth availability is issue at internet gateway instead of source encoding device. Thus, designed speech codec has encoder tools, rate converter tools and decoder tools. With these three tools, internet telephony can become powerful media in which million can communicate each other.

Another significant feature of the designed speech codec is bandwidth scalability. The user at the transmitter can down sample from any standard sampling frequency (16, 32, 44.1, 48 KHz) to 8 kHz sampling rate. The encoding and decoding is done at 8 kHz sampling rate. Similarly the user at the receiver can up sample it to the desired sampling frequency.

Presently existing speech codec's (ITU G72x, GSM AMR, MPEG-4 CELP) use 10th order LPC filter for narrow band applications and 16th – 20th order LPC filters for wide band applications. The computation and quantization of higher order LPC filter coefficients requires very high precision computation to maintain Stability of the LPC Model. Because of mentioned stability problem, present generation speech codec algorithms resort to conservative approach of pushing poles of LPC model very much inside unit circle. Thus there is acute need for better approach to model for voiced speech. Designed speech codec is using multiple 2nd order LPC model to over come mentioned problem. In this process, there are two advantages in the designed speech codec. First one is to provide more stable model (with error resilience as inbuilt) and the second one is to comfortably use present generation processors 16-bit DSP or 32-bit ARM for wide band speech codec based applications.

Contents

1.INTRODUCTION 4

2.BITRATE SCALABILITY 7

3.SUMMARY 10

REFERENCES 10

1. Introduction

Digital speech signal processing is having several advantages over analog counterpart for storage or for communication purposes and so it is rapidly replacing the analog speech processing techniques. To save storage memory or to efficiently utilize channel capacity it is important to compress digital speech samples while retaining the perceptible speech quality. The algorithms that compress and decompress speech signal through various coding techniques are referred to as speech codec's. For compatibility reasons, off late world standards for speech codec's have emerged. Some of the standards are ITU,GSM,MPEG-4 and their codec's are ITU 6.72x, Narrow band and Wide band GSM, very low bitrate HVXC MPEG – 4 and Bitrate and Bandwidth scalable CELP MPEG – 4.

All the presently existing speech codec's use 10th order LPC filter for narrow band applications and 16th – 20th order LPC filters for wide band applications. The computation and quantization of higher order LPC filter coefficients requires very high precision and advanced computation. So, these higher order filters are more susceptible to instability due to quantization errors in LPC coefficients computation. This instability whenever occurs, it severely degrades the performance which may lead to loose the information. To reduce the probability of instability one has to go for higher precision arithmetic and hence these codec's needs 32-bit word length or 64 bit word length to achieve this high precision. But the present popular and industry proven processors are 16 bit DSP processors. So the present existing popular Speech Codec's are not very much native to 16-bit DSP processors. Stability of LPC model becomes very critical and it is worse in the case of 20th order model. To overcome this problem we have innovated a new scalable speech coding technique keeping the processor architecture in mind, which works well with 16 bit DSP processors. This newly developed scalable speech codec ideally suits for the presently available popular 16-bit DSP processors. The use of scalable layers of second order IIR filters to model the speech signal in this scalable codec requires very less computation and it guarantees the stability of the filter coefficients. The stability of the filter is become problem in presently available speech codec's because of the use of higher order IIR filters.

The other main functionality of this speech codec is bitrate scalability. Rate converter tool is less complex and the same can be used in internet gateway for rate conversion. Mentioned rate converter tool is critical for heterogeneous network in which speech codec related data is traveling. One way to meet bandwidth requirements is to make encoder itself to produce compatible bitrate for a given bandwidth, but this is not interesting when bandwidth availability is issue at internet gateway instead of source encoding device. Thus, designed speech codec has encoder tools, rate converter tools and decoder tools. With these three tools, internet telephony can become powerful media in which million can communicate each other.

Depending on the availability of the bandwidth or current network condition, the number of layers to be transcoded is decided. The lowest possible bitrate of this speech codec is 11.37 kbps and highest possible bitrate is 46.67 kbps. The quality of the speech codec depends on the number of layers used in encoded bitstream.

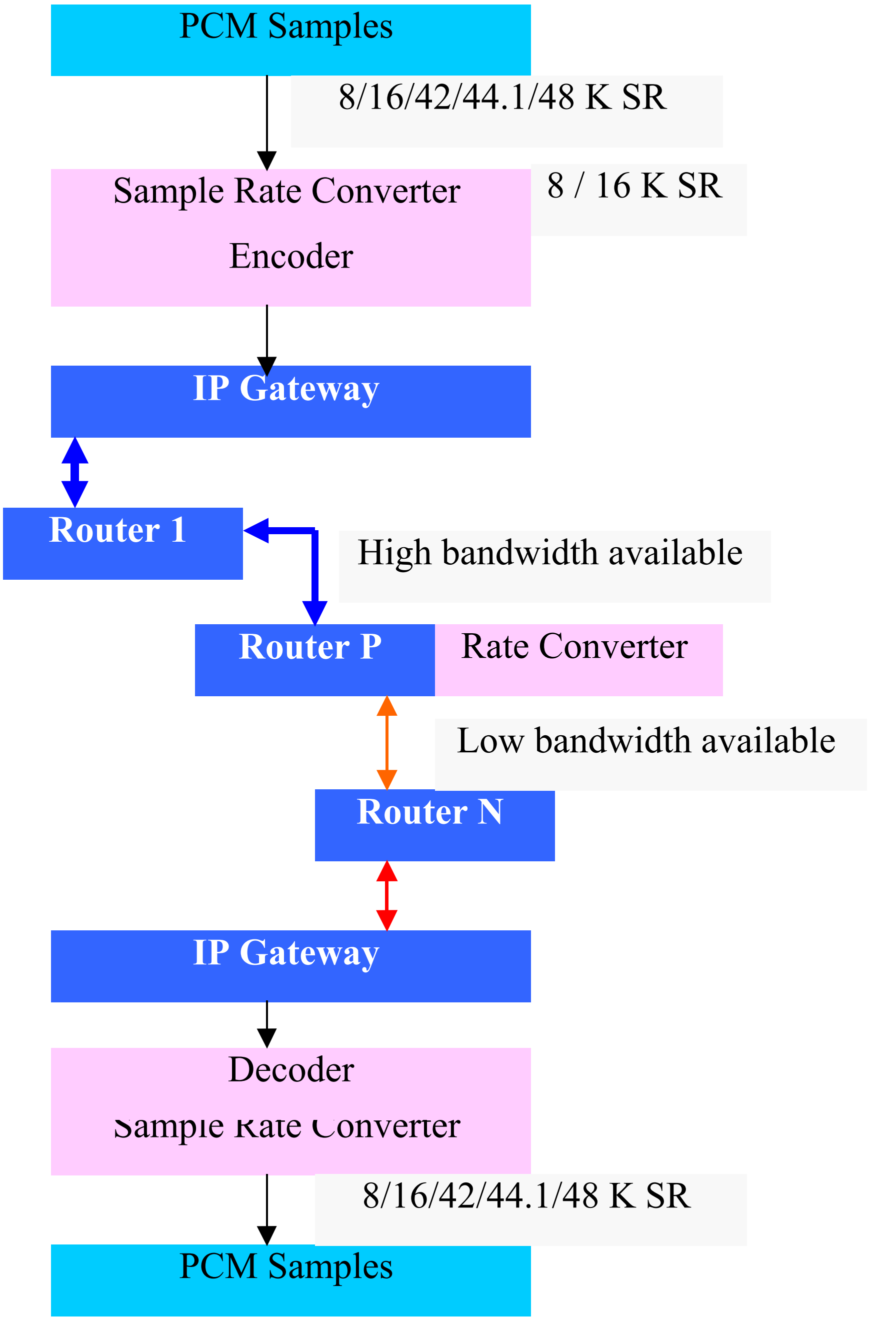
In general the characteristics of the speech signal lies below 4 KHz, therefore the sampling rate of either 8 KHz or 16 KHz is used. The speech signal has dominant resonant characteristics, non-dominant resonant characteristics and mixture of both. Based on these characteristics the speech signal is classified as voiced, unvoiced and combination voiced. The part of the speech signal, which has no spectral activity, is classified as silence. In this speech codec we have combined RMS energy of the speech frame and number of zero crossing of the speech frame, to classify the speech signal as voiced, unvoiced, combination of voiced, unvoiced signal and silence.

The most of the dominant resonant characteristics of the speech signal is well below 2 KHz and non-dominant resonant characteristics of the speech signal is in between 2 KHz - 4 KHz. Speech coding techniques used to model the characteristics of the speech signal, linear prediction analysis, FFT, HMM, Wavelets etc. And these mentioned codec captures the dominant resonant characteristics of the speech signal below the frequency 2 KHz. This speech codec makes use of hybrid model. The dominant resonant characteristics of the speech signal is coded by using layers of second order LPC model which captures the characteristics of the signal below 2 KHz. The residue is coded by using modified discrete cosine transform over the frequency range 2 KHz – 4 KHz. The LPC model does not capture the non-dominant characteristics of the speech signal, to code that we have innovated a new method called impulse response method (like wavelets) to code the non-dominant characteristics of the speech signal.

In this speech codec, the input speech signal is first passed through the high pass filter with cut off 50 Hz and then the windowing is performed. The windowed signal is then passed through the Voiced/Unvoiced/Silence detector. The voiced and unvoiced frames are coded separately. The silence part of the speech frame is assumed to have zero valued samples. Using the layers of second order LPC model, the silence part of the speech signal is coded. In order to improve the speech quality, the residual signal is then coded using modified discrete cosine transform over the frequency range between 2 KHz – 4 KHz.

The unvoiced part of the speech signal is coded by using impulse response method. In this method impulse response of the second order IIR filter is taken to code the non-dominant resonant characteristics of the speech signal. The impulse response is then cross-correlated with the unvoiced part of the speech signal. The phase shift needs to be added with the impulse response, which is estimated from the maximum cross correlation point. The phase shifted impulse response is then compressed/elongated to match unvoiced part of the speech signal. The residue coming from one layer is then encoded in the next layer. The number of layers to be encoded is decided based on the current network condition. The encoded parameters are then scalar quantized.

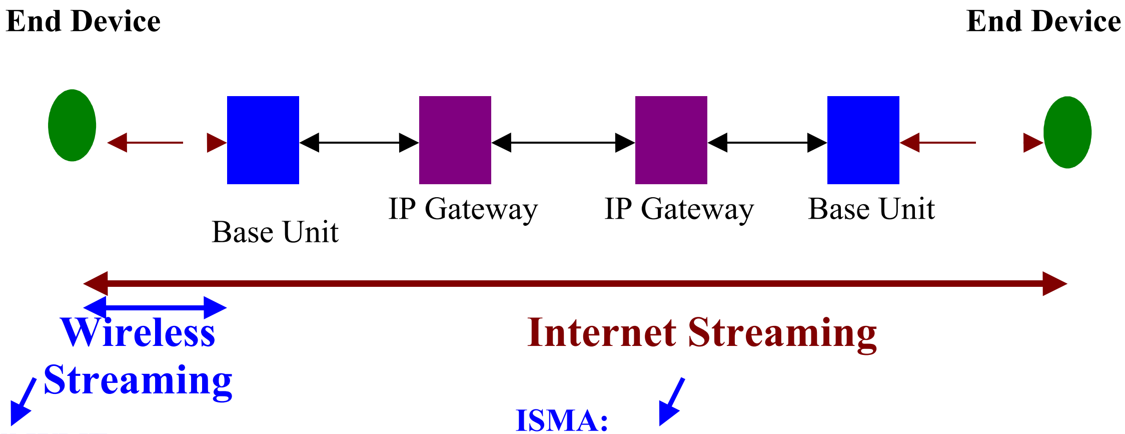
The part of the speech signal, which is classified as the combination of voiced and unvoiced signal is passed through both the voiced and unvoiced part of the speech codec. The number of voiced and unvoiced layers to be encoded depends upon the content of the dominant and non-dominant resonant characteristics of the speech signal. The change in voiced and unvoiced classification will not affect the quality of the codec much, because of the classification of the speech signal as fully voiced, fully unvoiced, 80% voiced and 20% unvoiced, 60% voiced and 40% unvoiced, 40% voiced and 60% unvoiced, 20% voiced and 80% unvoiced. The speech signals classified in between voiced and unvoiced are passed through both the voiced part of the codec and unvoiced part of the codec. Therefore, the quality of the codec was guaranteed irrespective of the changes/errors in classification of speech signal. In the following over view of proposed speech codec is given.



Section 2 provides information on mathematical preliminaries that are used in this paper. Section 3 provides information on voiced or unvoiced or silence detection algorithm. Section 4 provides information on decoder and section 5 provides information on rate converter. Summary is given in section 6.

2. Bitrate Scalability

Bitrate scalability achieved very differently in the proposed codec. For example, MPEG-4 CELP Speech codec have Bitrate Scalability, but encoder need to informed about scalability functional feature and also only three layers of scalability data can be attached along with encoded data such as Narrow band and Wideband data packets. GSM AMR speech codec have multi bitrate functional feature, but it is with in control of source encoder. There are other few speech codec's provides multi bitrate but they are with in the control of source encoder.



Full WMF:
AMR
EVRC

GSM & GPRS: 3GPP
AMR

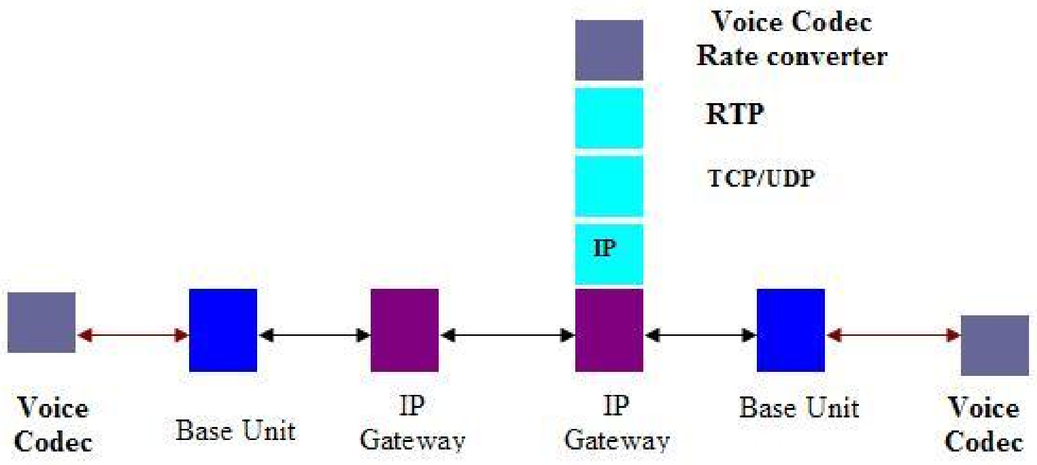
CDMA: 3GPP2
EVRC
GSM-AMR
Pure Voice
SMV

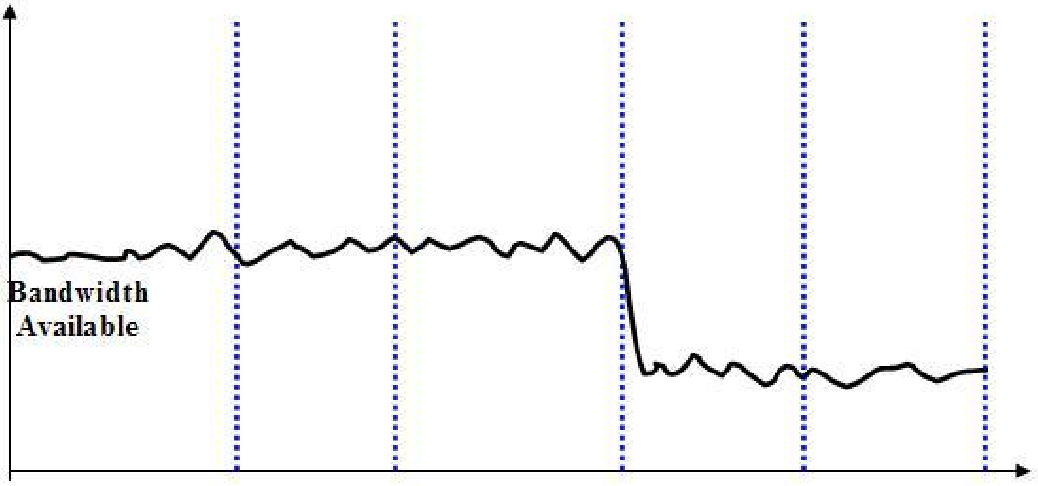
ISMA:
AMR
EVRC

GSM & GPRS: 3GPP
AMR

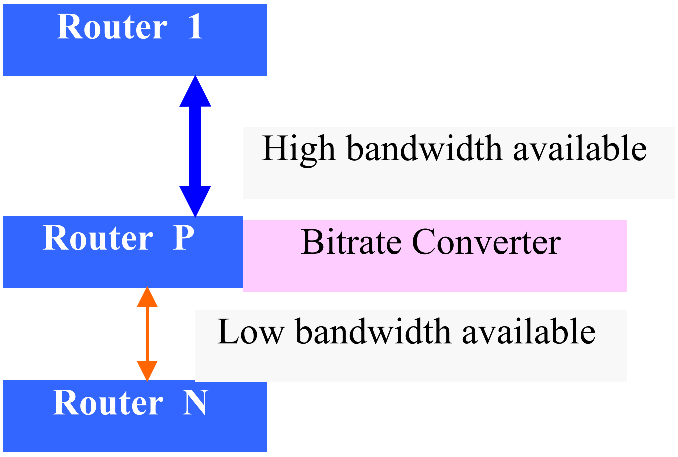
CDMA: 3GPP2
EVRC
GSM-AMR
Pure Voice
SMV

In the case of Internet streaming speech data, it appears that the above speech codec's are put in use. But this essentially brings up issue of "using available bandwidth efficiently". For example, bandwidth available may not be constant from end to end. There might be the case in which bandwidth availability is changing after encoded data traveled across few gateways. And in fact this is the case in most of IP telephony based data streaming.

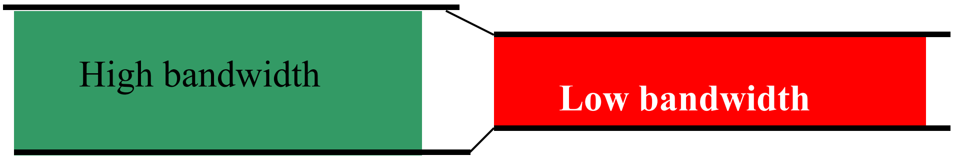




Bandwidth available across network is given in the above diagram. Though the above given diagram is hypothetical and the similar condition is expected in IP telephony network environment. In the proposed speech codec, we have new component called rate converter which will work in Internet gateway machine or in Internet router machine.

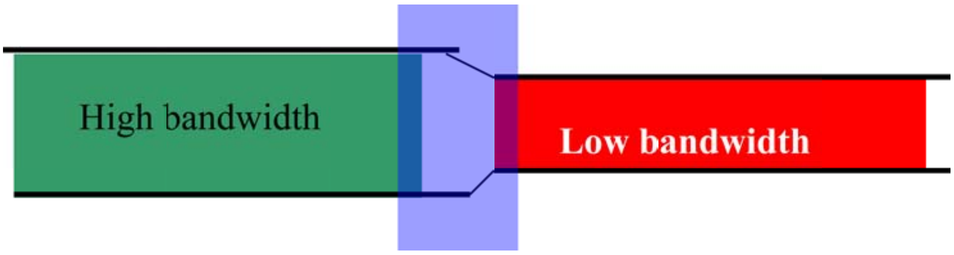


Rate converter is designed to perform the following missing thing in IP telephony.



Missing Something

Solution to the above problem is rate converter which is very new component in speech codec environment.



Rate Converter

The bitrate can be independently selected by simply striping off part of the layers from encoded bitstream. Complexity is less in rate converter and this enables server side machine to run rate converters for many channels (may be in the order of 250) in Pentium 4 machine. Scalabilities necessitate a single encoder to transmit the same data to multiple points connected at the different rates. In the following table, we provide scalability ID versus bitrate.

Scalability ID	Bitrate
0	10.36
1	13.73
2	17.11
3	23
4	24.31
5	31.26
6	34.63
7	38

Encoder will send data with Scalability ID 7 from source encoder. Rate converter can down convert bitrate from higher to lower by picking up lower scalability ID object from given higher scalability ID object. For example, rate converter can form Scalability ID 5 from given Scalability ID 6. Mentioned conversion is less complex because the kind of mathematical model is used in encoding speech data.

3. Summary

Performance evaluation of scalable speech codec is carried out by using well known methods such as **segmental signal to noise ratio**(SSNR) and RMS gain. The SSNR value lies in between 5.7 to 12.2 for different combinations. The SSNR value was found to be comparable with other speech codec such as MPEG-4 CELP. Bitrate scalability is the key contribution in this speech codec. Unlike conventional speech codec, proposed speech codec have components such as encoder, rate converter and decoder. Rate converter is very new components in speech codec design and also this is the central part of requirement from IP telephony applications. MIPS required for encoder appear to be high in reference implementation and more research requires to reduce MIPS requirement for encoder. Rate converter implementation need be tested in server with real time data in and out. May be these are the areas in which some work is required.

References

[1] ISO/IEC 14496-3 CELP Speech Codec Standards Documentation
[2] ISO/IEC 14496-4 CELP Speech Codec Confirmation Test documentation
[3] ITU-T G.723.1 and G.729 Implementor s Guide , 02/15/2002
[4] ITU-T G.729 Annex C+ Documentation, 02/17/2000
[5] ITU-T G.729 Documentation, 03/19/2000
[6] ITU-T G.723.1 Documentation 03/19/1996
[7] ITU-T G.723.1 Annex C Documentation 11/11/1996
[8] ITU-T G.723.1 Annex A Documentation 11/11/1996
[9] Enterprise IP Phone Solution TMS320C5472, Texas Instruments
[10] ETSI TR 126 901 V4.0.1 AMR Wideband Speech Codec, 2001
[11] ETSI TS 126 104 V4.2.0(2001-09) ANSI-C Code for AMR Speech Codec.
[12] TIA/EIA/IS-718 Minimum Performance Specification for the Enhanced Variable Rate codec, Speech Service Option 3 for Spread Spectrum Digital Systems.
[13] http://www.ilbcfreeware.org/ iLBC (internet Low Bitrate Codec)