An Improvement of Voice Quality in LTI System Operating in Codec G729 [ANNEXB = NO] for Low Bandwidth Region

Md. Nahid Akter Shafi, Abdullah Al Mamun, and Md. Niamul Hassan

***Abstract*—In voice communications, the speech signal is degraded when gone through the system layers, since the best effort arrangement based IP network prompts the system degradations including Silence, packet loss jitter. This paper represents the impact system debasement calculates on VoIP system and the otherworldly analysis of the VoIP signal. The spectral analysis of VoIP signal is performed through the various signal processing algorithms. The outcomes are accepted through the quality assessment of the VoIP signal utilizing perceptual evolution of speech quality (PESQ) approximation for narrowband signal. For digital signal processing LTI System, an input signal can be mapped to an output signal where in mapping we can transform the signal in a manner that the voice quality can be enhanced than before extraordinarily in low bandwidth region. The mapping process can extract the noise and the resulting amplifying signal will have low SNR as compared to input signal. It can be shown that the voice bit rate can be drawn approximately to 5.3 kbps with excellent voice quality.**

***Index Terms*—Spectrogram, SNR (signal to noise ratio), LTI (linear time-invariant), MOS (mean opinion score), FIR (Finite impulse response).**

I. Introduction

Once a call has been set up between two or more VoIP devices, the caller starts speaking. At this point the voice signal has to be converted into a digital signal, formatted for TCP/IP transmission and sent along the network to the destination, where all of the preceding steps have to be reversed [1].

The frequency range of human speech covers 300 Hz to 3300 Hz. However, the telephone companies use the bandwidth for voice communications is usually limited to 4000 Hz or 4 KHz, which is high enough to capture the major pitch and enough of the texture to make the voice sound human. The most common technique to change an analog signal to digital data (digitization) is called Pulse Code Modulation (PCM). PCM converts the 4 KHz voice signal of a phone call to a digital format by sampling the analog signal 8000 times per second and converting each sample into a digital bit stream.

According to the Nyquist Sampling theorem, the sampling rate must be at least 2 times the highest frequency contained in the signal [2].

Manuscript received November 1, 2015; revised December 17, 2015.

Md. N. A. Shafi and Md. Niamul Hassan are with the Network Operations Center of Inspire Systems Limited, Dhaka, Bangladesh (e-mail: nahid_apee@yahoo.com, nhleon17.nl@gmail.com).

Abdullah Al Mamun is with the TISL, Dhaka, Bangladesh (e-mail: aamamun.deb19@gmail.com).

The analog signal is sampled by the Nyquist sampling rate = 2 × $f_{\max}$ (where, $f_{\max}$ is the maximum voice frequency = 4 KHz). Thus, the sampling rate is 8 KHz. The output of the sampling is converted into a series of amplitude pulses called Quantizing. After each sample is quantized, it is encoded to digital streams with an 8 bit rate known as Encoding.

Thus, the required bandwidth for a telephone call that PCM produces is 8 KHz × 8 bit = 64 KHz or 64 kbps stream of digital data with excellent voice quality, which is mostly used in telecom world. The digitalization process is shown in Fig. 1.

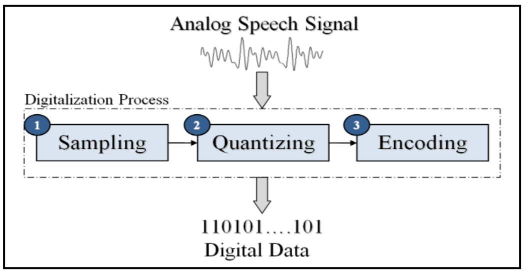

Fig. 1. Analog to digital conversion (ADC) process.

It has been estimated that as much as 60% of a voice conversation is silence. Deleting these empty bits decreases the amount of data needed for the voice transmission. Silence Suppression, also called voice activation detection (VAD), is used in telephony to describe the process of not transmitting information over the network when one of the parties involved in a telephone call is not speaking.

II. Methodology

The impulse response and frequency response are two attributes that are useful for characterizing linear time-invariant (LTI) systems. They provide two different ways of calculating what an LTI system's output will be for a given input signal. A continuous-time LTI system is usually illustrated like this:

In general, the system H maps its input signal $x(t)$ to a corresponding output signal $y(t)$. There are many types of LTI systems that can have apply very different transformations to the signals that pass through them.

Let us assume we have an LTI system with input $x(t)$, and output $y(t)$ as shown in the Fig. 2.

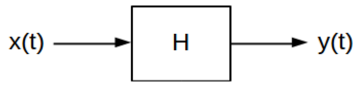

Fig. 2. LTI system.

Let $x(t)$ denote the voice signal with unwanted noise at bit

DOI: 10.7763/IJCTE.2017.V9.1123

rate 5.3 is fed to the H maps, the output will generate a signal (let's denote $y(t)$) will have improved voice quality with the same bit rate.

In a real life environment it is tested practically in research that it is possible to transmit voice satisfactorily using 5.3 kbps by changing some parameter [3], [4]. If the voice with 5.3 kbps is fed to a noise extractor using FIR filter to eliminate unwanted raised peak, it is possible to improve the voice quality in low bandwidth areas. The extended voice quality will have low SNR as compared to the pre-processing signal.

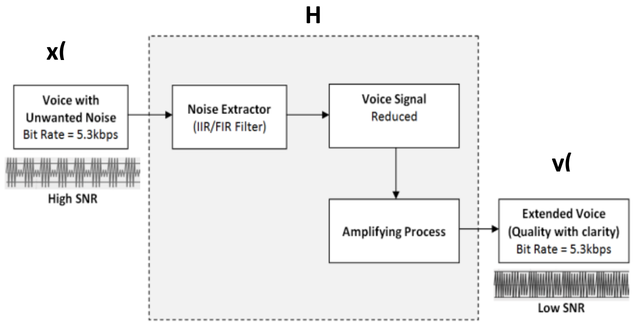

Fig. 3. Input and output signal analysis for LTI system.

For a causal discrete-time FIR filter of order N, each value of the output sequence is a weighted sum of the most recent input values:

$$y[n] = b_o x[n] + b_1 x[n-1] + \cdots + b_n x[n-N]$$
$$= \sum_{i=0}^{N} b_i . x[n-i]$$

where:

$x[n]$ is the input signal,

$y[n]$ is the output signal,

N is the filter order; an Nth-order filter has (N+1) terms on the right-hand side

b_i is the value of the impulse response at the i'th instant for $0 \le i \le N$ of an Nth-order FIR filter. If the filter is a direct form FIR filter then b_i is also a coefficient of the filter [5], [6].

The impulse response of the filter as defined is nonzero over a finite duration. Including zeros, the impulse response is the infinite sequence:

$$h[n] = \sum_{i=0}^{N} b_i . \partial[n-i] = \begin{cases} b_n & 0 \le n \le N \\ 0 & otherwise \end{cases}$$

If an FIR filter is non-causal, the range of nonzero values in its impulse response can start before n=0, with the defining formula appropriately generalized.

The filter's effect on the sequence $x[n]$ is described in the frequency domain by the convolution theorem:

$$\underbrace{F\{x * h\}}_{Y(\omega)} = \underbrace{F\{x\}}_{X(\omega)} * \underbrace{F\{h\}}_{H(\omega)}$$

and

$$y[n] = x[n] * h[n] = F^{-1}\{X(\omega).H(\omega)\}$$

where operators F and F^{-1} respectively denote the discrete-time fourier transform (DTFT) and its inverse. Therefore, the complex-valued, multiplicative function $H(\omega)$ is the filter's frequency response. It is defined by a Fourier series:

$$H_{2\pi}(\omega) \stackrel{\text{def}}{=} \sum_{n=\propto}^{\infty} h[n].(e^{i\omega})^{-n} = \sum_{n=0}^{N} b_0 .(e^{i\omega})^{-n}$$

where the added subscript denotes 2π-periodicity. Here ω represents frequency in normalized units (radians/sample). The substitution $\omega = 2\pi f$ favored by many filter design programs, changes the units of frequency (f) to cycles/sample and the periodicity to 1 [7]. When the $x[n]$ sequence has a known sampling-rate, f_s samples/second, the substitution $\omega = 2\pi f / f_s$ changes the units of frequency (f) to cycles/second (hertz) and the periodicity to f_s. The value $\omega = \pi$ corresponds to a frequency of $f = \frac{f_s}{2}$ Hz $= \frac{1}{2}$ cycles/sample, which is the Nyquist frequency. The output signal will reduce its bit rate with worse voice quality.

Several techniques have been invented for measuring the quality of the voice signal that has been processed by different compression algorithms (CODECs). One of the standard techniques for measuring quality of voice CODECs, which is also an ITU standard, is called Mean Opinion Score (MOS). MOS values, which are subjective and expressed by humans, range from 1 (worst) to 5 (perfect or equivalent to direct conversation).

MOS is an ITU standard method of measuring voice quality based on the judgment of several participants; therefore, it is a subjective method. Table I displays each of the MOS ratings along with its corresponding interpretation, and a description for its distortion level. It is noteworthy that an MOS of 4.0 is deemed to be Toll Quality [8].

TABLE I: MEAN OPTION SCORE

Rating	Speech Quality	Level of Distortion
5	Excellent	Imperceptible
4	Good	Just perceptible but not annoying
3	Fair	Perceptible but slightly annoying
2	Poor	Annoying but not objectionable
1	Unsatisfactory	Very annoying and objectionable

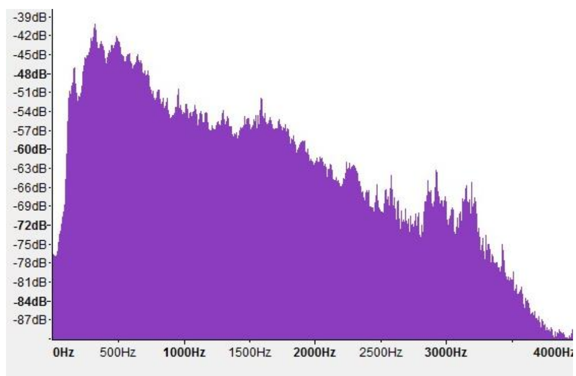

Fig. 4. Input signal $x(n)$.

III. SPECTROGRAM

The spectrogram of the speech signal is an intensity plot of short time Fourier transforms (STFT) magnitude. STFT is a sequence of FFTs of windowed data segments where the windows are usually allowed to overlap in time. The data to be broken up into frames, which usually overlap to reduce artifacts at the boundary and then each frame, is Fourier transformed using K-point FFT [9]. The STFT procedure is shown in Fig. 4.

The short time Fourier transform (STFT) is expressed as:

$$S(\omega_k, l) = \sum_{n=0}^{N-1} x[n+l]\omega[n]e^{-j\omega_k n}$$

where $\omega_k = \frac{2\pi k}{NT}$, $k = 0 \ldots K-1$ at given sampling frequency f_s, $x(n)$ is the time domain signal and $\omega(n)$ is the window function.

If the output signal with reduced bit rate is proceed thru amplification process up to 5.3 kbps, we can achieve imperceptible the voice quality. Also, the output *y*(*t*) signal will have excellent voice quality than *x*(*t*).

IV. Result and Discussion

To examine the impact of packet loss on the quality of the corrupted VoIP output, the spectral investigation was performed in listening to mark dB and frequency both. The signal processing algorithms utilized for spectral analysis is discussed [10].

The spectrogram gives an effective description of the time variety of the spectrogram of voice. The spectrogram of input signal *x*(*n*) and the came about output signal *y*(*n*) is indicated in Fig. 5. The periods where voice is available can be clearly recognized from the pitch structure of the vocal line excitation. In fact looking at histogram, and depending on the pitch structure of speech, it is possible to manually segment the signal into voice periods with ease and with a high degree of accuracy.

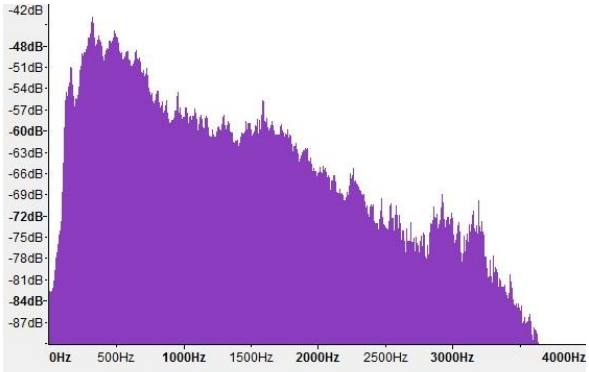

Fig. 5. Output signal *y*(*n*).

As is shown in Fig. 6, the input signal *x*(*n*) has silence suppression and the associated SNR can be reduced up to a reasonable quality. The SNR of output signal *y*(*n*) has reduced silence suppression (Fig. 7). Thus, the voice quality will have an improvement.

Quality of the dialer system administration can be measured by MOS, which gives the subjective perspective. The specialized parameters of the quality bring the more objective information. Signal-to-noise proportion is one of such specialized parameters. It is the proportion of the power of transmitted signal and the noise signal (distinction between transmitted signal and received signal) [11]-[13].

Fig. 6. SNR for input signal *x*(*t*).

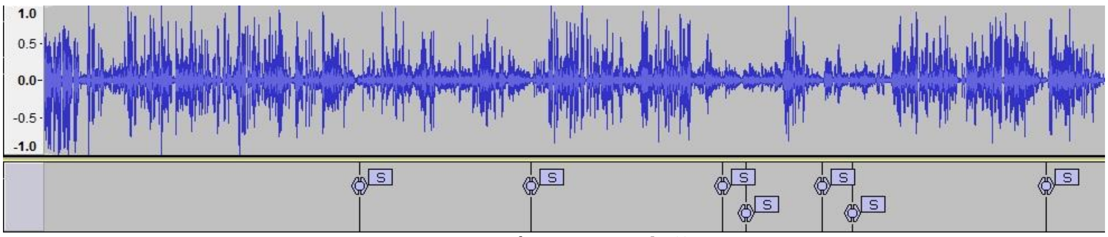

Fig. 7. SNR for output signal *y*(*t*).

V. Conclusion

We have shown an improvement in voice quality for an LTI System through a mapping procedure. The mapping procedure deals with functional analysis that can improve the voice quality. This study is applicable particularly in low bandwidth region where quality is main alarm. The real time implementation of voice quality had been offered through practical experiment and the spectral analysis of speech signal was carried out in this work to evaluate the performance of voice quality with G729 CODEC [annexb=no] analysis. The voice signal quality was unfavorably influenced by packet losses, since with expansion of jitter and delay, sequential packet losses, which diminished the MOS scores of the signal. The spectral analysis results are related to that of the objective measurement results for the designed VoIP system. FIR filter is much of the time utilized as a part of voice communications, since it has the better ability to hide the lost packet amid voice transmission. To improve the execution of the framework, these debasement parameters ought to be appropriately addressed. In future, the study could be utilized for enhancing the VoIP speech signal quality utilizing different signal processing and filtering calculations performed at higher frequency digital signal processing algorithms.

Appendix

VOS switch, VPS switch, various dialer, SQL database, firewall system, STM device and CISCO router & switch, OS-Windows server and Linux server etc.

REFERENCES

[1] B. A. Forouzan, *Data Communications and Networking,* 4th ed. Tata McGraw-Hill, 2009, p. 69.

[2] N. Wittenberg, *Understanding Voice over IP Technology*, 1st ed. Delmar Cengage Learning, 2009, pp. 34-36.

[3] N. A. Shafi, O. Farrok, and M. M. Ali, "An analysis of voip communication and overview of bangladesh practical field," *IJCTE*, vol. 5, no. 3, pp. 574-577, 2013.

[4] N. A. Shafi, Al Kawser, and O. Farrok, "Practical field overview voice quality of rtp packet size analyze on codec G729 Annexb = no in low bandwidth area of Bangladesh," *IJCTE*, vol. 7, no. 1, pp. 62-65, 2015.

[5] A. V. Oppenheim, A. S. Willsky, and I. T. Young, *Signals and Systems*, Englewood Cliffs, New Jersey: Prentice-Hall, Inc, 1983, p. 256.

[6] L. R. Rabiner and G. Bernard, *Theory and Application of Digital Signal Processing*, Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1975.

[7] A. E. Cetin, O. N. Gerek, and Y. Yardimci, "Equiripple FIR filter design by the FFT algorithm," *IEEE Signal Processing Magazine*, pp. 60-64, March 1997.

[8] J. V. Meggelen, L. Madsen, and J. Smith, *CCNP ONT Official Exam Certification Guide*, 2nd ed. Cisco Systems, Inc., 2008, pp. 24-25.

[9] T. F. Quatieri, "Discrete-time speech signal processing: Principle and practice," MIT Lincoln Laboratory, Liexington, Massachusetts, Prentice Hall, 2002.

[10] H. P. Singh, S. Singh, and J. Singh, "Real time analysis of VoIP System under pervasive environment through spectral parameters," *International Journal of Computer Applications*, vol. 31, no. 2, October 2011.

[11] V. K Goyal, "Multiple description coding: compression meets the network," *IEEE Signal Processing Magazine*, September 2001.

[12] K. Bachratá, J. Boroň, and B. Rosival, "Katedra informačných sietí, Fakulta riadenia a informatiky, Žilinská Univerzita, Slovensko, new methods proposal for VoIP transmission," scientific papers of the University of Pardubice, The Jan Perner Transport Faculty 12, 2006.

[13] M. Klimo and K. Bachratá, "Impact of correlated loss on VoIP quality," scientific papers of the University of Pardubice, The Jan Perner Transport Faculty 11, 2005.

Md. Nahid Akter Shafi was born in Rajshahi, Bangladesh on July 23, 1985. He received the B.Sc. and M.Sc. degrees from the Department of Applied Physics and Electronic Engineering, Rajshahi University, Rajshahi, Bangladesh, in 2006 and 2007 respectively. He also certified on ITIL-Foundation, CEH, RHCE and CCNA. He has been working as a senior system engineer in the Network Operations Center of Inspire Systems Limited in Dhaka, Bangladesh since June 2011. He is a member of the American Society for Research (ASR), Science and Engineering Institute (SCIEI) & International Association of Computer Science and Information Technology (IACSIT). His current research interests include VOIP, information security and spintronics.

Abdullah Al Mamun was born in Chittagong, Bangladesh on February 1, 1986. He received the B.Sc. degree from the Department of Electronics and Telecommunication Engineering, Daffodil International University, Dhaka, Bangladesh. Also certified on CEH, RHCE, and ITIL-Foundation. He has 4+ working experience at leading multinational company such as TISL, HP etc. He is an associate member of Institution of Engineers, Bangladesh. His current research interests include information security and secure contact center solution.

Md. Niamul Hassan was born in Sirajganj, Bangladesh on November 17, 1991. He got the BSc. degree from the Department of Electronics and Communication Engineering from BRAC University, Dhaka, Bangladesh in 2015. At his understudy life, he began his career as a support engineer in Network Operations Center of Inspire System Limited. Presently, he has been filling in as System Engineer in the Network Operations Center of Quest Software Solutions Limited in Dhaka, Bangladesh since December 2014. His flow exploration hobbies incorporate VOIP and as of late he has been distributed his work for green technology.