Sound quality assessment in VOIP environment

Darko Lukša[1], Siniša Fajt[2], Miljenko Krhen[2]

[1]Polytechnic of Zagreb, Zagreb, Croatia
[2]Faculty of Electrical Engineering and Computing, University of Zagreb
e-mail address: darko.luksa@tvz.hr

Abstract

The need for measuring the sound quality in VOIP environment is a basic requirement of modern multimedia communication systems due to technical, commercial and legislative reasons. Measuring the sound quality using objective methods is not standardized up till now. Expected measurement results should be in correlation with subjective tests of sound quality. The usage of application for measurement of acoustic qualities of the space is explored, with an emphasis on evaluating speech intelligibility as the main parameter. In this work we present measurement method and influence of codec choice on different hardware platforms. Objectively measured small differences showed possibility to evaluate the sound quality of voice internet protocol.

I. INTRODUCTION

Nowadays, Voice Over IP (VoIP), IP telephony, audio streaming and videoconferencing becomes familiar for a wide range of users. Due to the nature of Internet communication there are still challenges to be overwhelmed to provide the end user with acceptable levels of service quality, especially because of unreliable connections quality and unexpected propagation delays. In the second chapter we presents main characteristic of main components in VOIP systems. Third chapter depicts test environment and methods, and discuss measured results.

II. THE OBJECTIVE OF THE PAPER

Assessing the quality of interactive VoIP streams is a difficult task which than that of assessing the quality of one-way streams. In this case more parameters need to be considered, such as the delay and jitter and various software configurations.

The VoIP tool employed is called Elastix [7], which is based on open-source tool Asterisk. Asterisk is the Open Source software PBX that's widely used in business telephony. It runs on Linux and provides a full-featured approach to voice and data transport over TDM, switched, and Ethernet architectures. Additionally, as one of the fastest growing Internet technologies, Voice over IP can provide a number of specific services compared to traditional telephony such as conferencing, events notification, presence notification, text-to-speech, instant messaging, video telephony and various multimedia transmissions. It could be expected that open source IP telephony products will replace a number of proprietary hardware and software with standard Linux servers and open call control over the coming decade.

Recent advances in embedded systems development qualify them for implementation of Asterisk based system. Such implementation could be found in small IP enabled embedded system boards such as Raspberry PI and various home routers.

Nowadays the most extended protocols for VoIP are SIP [4] and IAX protocol [5]. SIP protocol is widely used and represents the most popular standard until now in many VoIP applications. Before the apparition of IAX, the functionalities of call signaling and multimedia transport were implemented by different protocols acting in parallel (H.323 and SIP for signaling and RTP for the transport). However, IAX as a new concept in VoIP combines both functions in the same protocol. Thus IAX conveys all the signaling and multimedia information for the management of the voice streams into the same session which improves performance. Additionally, IAX also provides the possibility of using Trunk connections for several calls. Using this option, all the protocol overhead is common for all the calls between two IAX nodes, which clearly reduce the bandwidth usage.

During this conversation phase, SIP protocol rely the multimedia transport on Real-time Transport Protocol - RTP. On the other hand, IAX performs multimedia transport in a very different manner. Instead of use the services of RTP, IAX sends multimedia traffic into the one IAX session. Although this leads to the different impact on multimedia transport in this work we are focused on superior impact on bandwidth requirement by codec type choice.

At the transmitter side, the encoder takes an analog voice signal from an input such as a microphone and transforms the signal into a digital format that can be appended as the payload of an RTP packet. On the receiver, the decoder takes that payload content of the RTP packet and converts it back to an analog signal for play out to the receiving user. The broadest division among speech coding algorithms is between waveform quantization and parametric quantization, where the first group represents the signal by the time samples quantization. Second group denotes the signal by a binary representation of a speech model or spectral parameters.

The simplest nonparametric coding technique is Pulse Code Modulation (PCM), which is simply a quantizer of sampled amplitudes. ITU T Recommendation G.711 standardizes the u Law and A Law compression algorithms, used in the United States and Europe, respectively. These algorithms use logarithmic scale quantization and are considered both uncompressed and reference for quality comparison with other algorithms. To enable uses in VOIP environment, ITU T recommendation G.711 incorporated Packet Loss Concealment and Discontinuous Transmission mechanisms for the G.711 codec.

Parametric codecs characteristically involve an analysis and synthesis method. In the analysis phase, speech is represented by a compact set of parameters that are encoded efficiently. At the synthesis phase, these parameters are decoded, and used in conjunction with a reconstruction algorithm to recreate a speech signal. The analysis phase can be open loop or closed loop. In the case of close loop analysis, the parameters are extracted by minimizing an objective metric, typically the square of the error between the original signal and the reconstructed speech. Due to the fact that close loop analysis usually involves the reconstructed speech, it is also called analysis by synthesis.

Some of the most commonly used codecs in the industry today which use Linear Prediction Coefficients are for the representation of the speech signals, and construct the signal using the analysis by synthesis ITU T G.729, ITU T G.723.1 and iLBC. For example, in the case of G.729 codec the autocorrelation coefficients are computed on 10 ms windows, and then converted to LPC coefficients.

Generally, packet networks have some inherent unsuitability with the transport of real time services, such as voice or video. The total delay on the network is compounded by several independent sources, which can be broadly divided in two categories: fixed delay components and variable delay components.
Variable delay are commonly referenced as delay jitter or simply jitter, is caused by the queuing that occurs in the different egress trunks that interface the Local Area Networks (LANs) to the Wide Area Networks (WANs), and add a variable and unpredictable amount of delay.

Coder delay refers to the time taken by the processor to compress a block of sound sample. This factor is manifestly affected by the processing capability of the device performing the selected compression algorithm, and the number of concurrent processes this device should handle. In this process the greatest importance has the algorithmic complexity and the frame size of the coder.

There are two other important factors that add to the overall delay are the decompression delay and the algorithmic delay. Although decompression delay is roughly 10% of the compression delay, but it should be noticed that it is multiplied by the number of frames sent in an RTP packet. Algorithmic delay refers to the unavoidable delay that some coders required in order to acquire future samples of the signal and depends on the number of future frames the coder or decoder needs to look ahead for proper operation.

Unavoidable packetization delay represents the time it takes to accumulate sufficient frames as needed by the designed number of frames in a single RTP packet. This delay is the product of the window size chosen for the compression algorithm (in ms) and the number of frames included in each RTP packet.

Queuing delay is a variable delay which depends on the trunk speed and the state of the queue in the router. It refers specifically to the queue inside the transmitting endpoint or gateway, and should not be mixed up with the propagation or network switching delay.

Network Switching Delay represents the propagation portion of the switching delay which results from the physical distance that needs to be traveled by the transmitted packets. Network congestion may adjust the number of hops required to reach the destination points, consequently the physical distance can vary from packet to packet, making packets arrive in an order different to which they were transmitted.

Specifically, we used Elastix as a unified communication platform that combines many popular sets of software available for Asterisk-based PBXs into a single interface. As a result, Elastix provides a highly user-friendly configuration of different VOIP parameters, especially for those users who are not familiar with Linux and complicated VOIP protocols implemented in Asterisk.

In exploitation it is important to clearly specify all the parameters consuming the VOIP PBX's resources CPU, RAM and hard disk. These parameters include the number of concurrent calls, type of codec used for different trunks, call features such as voice mail, music on hold and call conference to be supported. Even though it is very difficult to specify how much each of these parameters consume the VOIP-PBX resources, this information can benefit us to approximately determine the specification of the hardware for developing the VOIP-PBX. Table 1 presents required resources for main functionality. Due to the nature of different end point devices VOIP PBX an important assignment is converting different coding between them, which is processor and memory consuming task.

Commonly, in practice are used two types of phone for VoIP system: (1) Softphone and (2) specialized VoIP phone. They share the common purpose of making and receiving phone calls that are routed over the Internet, but are quite different in various aspects.

Softphone refers to software that can be installed in a computer device such as PC, notebook, smartphone, or tablet and utilizes its resources. Nowadays, there are plentiful softphones available to download, for example, X-lite, 3CX, Zoiper and many more which could be

Table 1 VOIP functionality and its resource consumption

Functionality	CPU	RAM	Hard disk
Concurrent calls	x	x	
Transcoding	**x**	**x**	
Voice mail, Call log, Backup			x

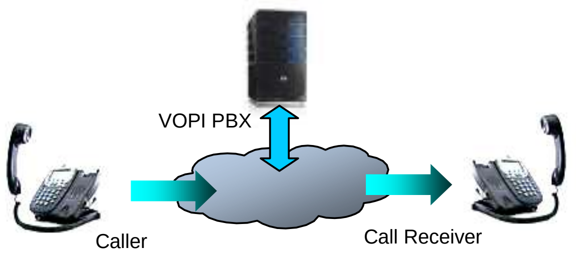

Figure 1 System diagram

Table 2 Codec properties

Codec	Bit rate [kbps]	Frame size [ms]	Data size per frame [bytes]
g.711 (A Law)	**64**	**20**	**160**
g.726	32	20	80
gsm	**13**	**20**	**33**
lpc10	2.4	22.5	7

found in different versions. The call features they provide typically are more flexible when comparing with what VOIP phones can offer.

On the other hand, VOIP hard phone refers to physical equipment that is devoted for telephony services. It is usually implemented using dedicated SOC processors for networking a wide variety of voice coders, jitter buffer management and packet loss concealment techniques, and a library of standard telephony algorithms required for IP phone operations. Comparing with the use of softphone, it is considered to be more user-friendly and due to usage of dedicated speaker and microphone provides better voice quality.

In our study we analyses influence of codec choice on different hardware devices ranging from small embedded systems to high performance computer systems.

III. DESCRIPTION OF SYSTEM

The functional diagram of the system deployed for the purpose of sound quality assessment is presented in Fig. 1. It basically includes three components: an Asterisk server, a Caller, and a Call Receiver.

For the Asterisk server we have used Elastix 2.4 with CentOS 64-bit Linux system on a Pentium 4 (2.4 GHz) with 2 GB RAM Memory. Used engine is Asterisk 1.8.20 using a standard configuration.

Asterisk server support different codec type for transcoding in a case of when receiving endpoints doesn't support used codec.

As a Caller and Call Receiver softphone was used. Choice of softphone over hard phone provides a flexibility of controlling and choosing codec on both endpoint sides. In our test environment free Zoiper softphone was used [6]. In Table 2 are presented main properties of exploited codecs.

III. THE MATERIALS AND RESULTS OF RESEACH

To determine the quality of a VoIP system, it is necessary to perform a number of objective measurements ("classical" methods for signal distortion and SNR) and subjective tests (overall quality score - user opinions).

In this study the performance and quality of a VoIP system dependent on a variety of codecs and parameters was measured, with the emphasis on the objective measurements. Measurements were performed in an anechoic chamber. The measurement system is shown in Fig. 2.

Asterisk were managed through Elastix interface. The software package Easera Pro (Electronic & Acoustic System Evaluation & Response Analysis) installed on Lenovo ThinkPad L512, Windows 7 Professional SP1 64 bit was used

Through Elastix interface more extensions or user (receiver and transmitter) were created. To place the call using the software IP phone (Softphone) ZoIPer. ZoIPer

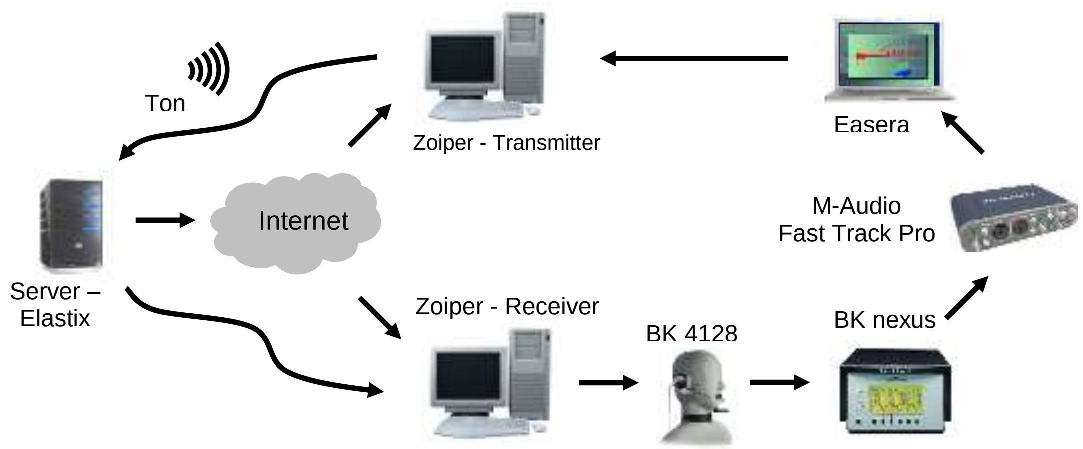

Figure 2 The Measurement System

Table 3 Speech intelligibility

Parameter	**gsm_gsm default**	**gsm_alaw default**	**alaw_alaw default**
STI	0,77	0,78	0,804
AlCons (%)	2,624	2,492	2,18
STI (male)	0,776	0,778	0,807
STI (female)	0,753	0,759	0,792
RASTI	0,855	0,828	0,855
equv. STIPa (male)	0,791	0,781	0,811
equv. STIPa (female)	0,77	0,765	0,8

is a software IP phone that supports both SIP and IAX2 signaling protocols and is the most commonly used software with Asterisk VoIP telephone switchboard. ZoIPer comes in several versions: free, business and OEM. The program has a built-in following audio codecs:

- G.711 µlaw
- G.711 alaw
- GSM
- Speex
- iLBC 20
- iLBC 30
- G.729 (at extra charge)

In this research the parameters by default mode were used or with the included parameters Echo cancellation, AGC and noise suppression for different combinations of codecs such as alaw - alaw, GSM - GSM and alaw - GSM. The amplitude frequency response is measured broadband with the followind codec combinations: alaw - alaw, GSM - GSM and alaw – GSM. It is noticed that there was a difference in the characteristics which can be seen in Figures 3-5.

However, in speech intelligibility by changing combinations of codecs differences were not observed. STI (Speech Transmission Index) during the measurements is ranged from 0.70 to 0.90. Index of speech intelligibility STI is a parameter for the objective evaluation of the quality of voice transmission (Table 3).

However, although in speech intelligibility differences were not observed, subjectively are perceived differences in quality. Specifically in order to examine how the system behaves by octaves, 1/48 octave analysis was made. The review of octave analysis and octave analysis of amplitude characteristics of the combination of codecs a fundamental difference in the color tone or sound is spotted. Octave amplitude characteristics of the combination of codecs GSM-GSM, GSM-alaw and alaw alaw is shown in Figure 6.

IV. CONCLUSIONS

On the basis of measurements it can be concluded that there is no objectively influence the choice of codec and their combinations on the intelligibility of voice, but it was objectively confirmed subjective aural impression of sound color change depending on the combination used codecs.

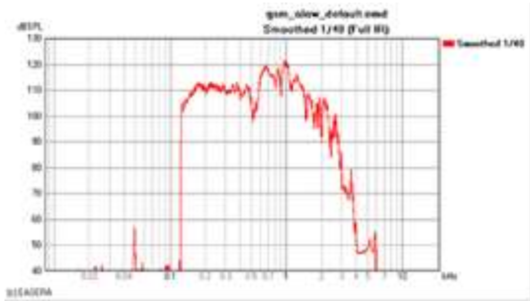
Figure 3 Amplitude characteristic of gsm/alaw codec

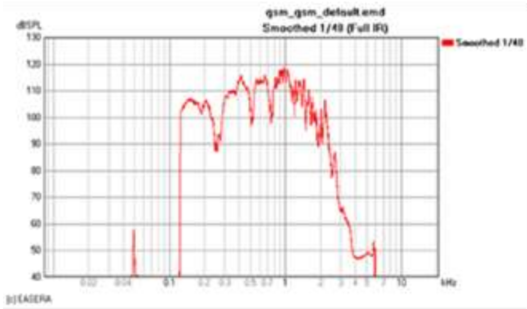

Figure 4 Amplitude characteristic of gsm/gsm codec

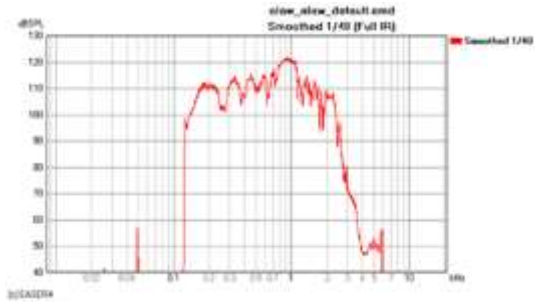
Figure 5 Amplitude characteristic of alaw/alaw codec

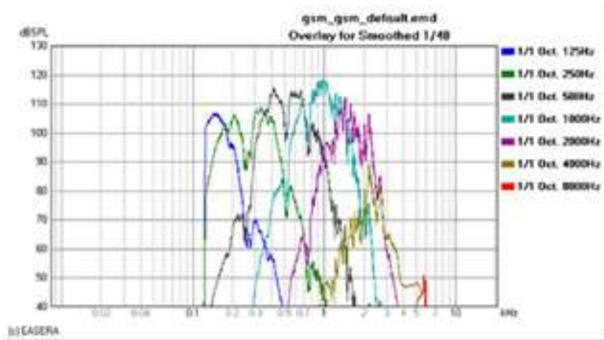

Figure 6 Codecs octal analysis of amplitude characteristics

REFERENCES

[1] Sruk, Vlado; Gašparović, Darko; Fajt, Siniša: Measurements of Speech Quality on VoIP Network, Proceedings of 2nd Congress of Alps-Adria Acoustics Association and 1st Congress of Acoustical Society of Croatia, Opatija,Croatia, 2005. pp. 317 – 320.

[2] Tomaš Goran, Furdek Marija, Fajt Siniša, Olujić Vladimir, Bruči Karlo: Perceptibility Of Aggressive Dynamics Processing In Digital Audio Broadcasting, Graz, Austrija: Joanneum Research Forschungsgeseellschaft mbH, 2007, pp. 1-4.

[3] Jim Van Meggelen, Leif Madsen, Jared Smith, Asterisk: The Future of Telephony, Second Edition, O'Reilly Media, Inc., Sebastopol, 2007.

[4] J.Rosenberg, H.Schulzrinne, G.Camarillo, A.Johnston, J.Peterson, and R.Sparks, M.Handley, E.Schooler, "SIP: Session Initiation Protocol", Internet Engineering Task Force (IETF), Request for Comments (RFC) 3261, Jun.2002.Available: http://www.ietf.org/rfc/rfc3261.txt.

[5] M.Spencer, B.Capouch, E.Guy, F.Miller, and K.Shumard, "IAX: Inter-Asterisk eXchange Version 2, draft-guy-iax-04", Internet Engineering Task Force (IETF), Internet-Draft, Mar.2008.Available: http://www.ietf.org/internet-drafts/draft-guyiax-04.txt.

[6] Zoiper softphone, http://www.zoiper.com/

[7] Elastix Open Source Unified Communications Server, http://www.elastixorg/