Multiple Transcoding Impact on Speech Quality in Ideal Network Conditions

Martin MIKULEC[1], Marcel FAJKUS[1], Miroslav VOZNAK[1], Olle JOHANSSON[2]

[1]Department of Telecommunications, Faculty of Electrical Engineering and Computer Science, VSB–Technical University of Ostrava, 17. listopadu 15/2172, 708 33 Ostrava-Poruba, Czech Republic
[2]Edvina AB, Runbovagen 10, SE-192 48 Sollentuna, Sweden

martin.mikulec@vsb.cz, marcel.fajkus@vsb.cz, miroslav.voznak@vsb.cz, oej@edvina.net

Abstract. *This paper deals with the impact of transcoding on the speech quality. We have focused mainly on the transcoding between codecs without the negative influence of the network parameters such as packet loss and delay. It has ensured objective and repeatable results from our measurement. The measurement was performed on the Transcoding Measuring System developed especially for this purpose. The system is based on the open source projects and is useful as a design tool for VoIP system administrators. The paper compares the most used codecs from the transcoding perspective. The multiple transcoding between G711, GSM and G729 codecs were performed and the speech quality of these calls was evaluated. The speech quality was measured by Perceptual Evaluation of Speech Quality method, which provides results in Mean Opinion Score used to describe the speech quality on a scale from 1 to 5. The obtained results indicate periodical speech quality degradation on every transcoding between two codecs.*

Keywords

Asterisk, MOS, PESQ, speech quality, transcoding.

1. Introduction

Transcoding is an inherent reality in VoIP networks. There are dozens of different codecs in transmission chain between particular VoIP providers. Every codec offers a compromise between the speech quality and required bandwidth. On one hand, there is a G711 codec offering very high speech quality and consuming a lot of bandwidth for each channel. It became the mandatory codec because its usage in fixed ISDN technologies. On the other hand, there is a GSM codec used in mobile networks, which offers sufficient speech quality and consuming five times less bandwidth than the G711 codec. The newer codecs such as G729 are trying to improve the effectivity of the old codecs used during last 20 years. The paper is aimed at the transcoding process on the transmission chain. It eliminates network parameters, which could influence the final speech quality by measurement in ideal laboratory conditions. The new Transcoding Measuring System was developed and described in the paper. The system can originate a call between SIP proxy servers, perform the transcoding between the required codes and evaluate the final speech quality on the transmission chain. The system is useful as the VoIP network administrator's tool to predict the speech quality inside particular network before the implementation of the VoIP into the network. The speech quality is measured by Perceptual Evaluation of Speech Quality [1] method, which provides accurate and repeatable estimates of speech quality degradation in VoIP network. It compares the audio signal input to a network with the corresponding (degraded) audio signal output from the network. The results are evaluated in Mean Opinion Score (MOS) used to describe the speech quality on a scale from 1 (bad quality) to 5 (excellent quality).

2. Transcoding Measuring System

We have developed our measuring system, which consists of the following parts:

- Speech Sample Player.

- Transcoding Cascade System.

- Speech Samples Recorder.
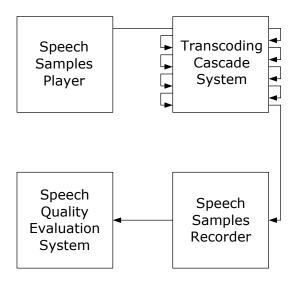
- Speech Quality Evaluation System.

Fig. 1: Transcoding Measuring System schema.

The system originates VoIP call between two nodes player and recorder. The call consists of 10 second speech sample, which is transmitted through Transcoding Cascade System with preconfigured codec translations. The degraded speech sample is recorded and sent to the Speech Quality Evaluation System, where is compared with the original speech sample. The results in MOS are stored into the database for future evaluation.

2.1. Speech Samples Recorder

The Speech Samples Recorder is represented by originating Asterisk PBX [7], which calls particular extension number of Speech Samples Player and records the speech sample obtained from the Speech Samples Player. The recorded speech sample is stored and sent to the Speech Quality Evaluation System [2] to get results.

2.2. Speech Samples Player

The Speech Samples Player is represented by terminating Asterisk PBX. The speech sample is available to playback under particular extension number, which can be called from originating Asterisk PBX.

2.3. Transcoding Cascade System

The Transcoding Cascade System consists of N Asterisk PBX systems running on the same server.

Every Asterisk PBX has own ports reserved for signalling and media transmission. There are SIP trunks configured between the Asterisk PBXs to achieve a unique route for a call. The Asterisk PBXs forwards the call to the destination extension number via localhost. According to that fact, there are no network parameters, which would influence the speech quality such as delay, packet loss, jitter or latency. The system ensures the degradation of the speech sample only from the codec transcoding perspective.
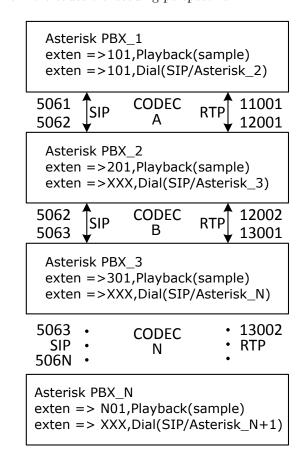

Fig. 2: Transcoding Cascade System schema.

2.4. Speech Quality Evaluation System

According to the good results obtained from the previous measurements [3], the PESQ algorithm is used as the speech quality measuring algorithm. The PESQ method is an intrusive and objective method [4]. It determines the quality of speech according to the comparison of the original signal with the degraded signal. The original signal $x(t)$ is taken from Speech Sample Player and the degraded signal $y(t)$ is taken from Speech Samples Recorder.

The results are evaluated in MOS-PESQ scale. Therefore, it is necessary to use complementary ITU-T P.862.1, which transfer scale from MOS-PESQ

to MOS-LQO. MOS-LQO provides a range of values from 1 to 5, which is more accurate scope for human subjective evaluation. Conversation from MOS-PESQ to MOS-LQO is defined by following equation [5]:

$$y = A + \frac{B - A}{1 + e^{-Cx+D}}, \tag{1}$$

where constants are $A = 0.999$, $B = 4.999$, $C = 1.4945$, $D = 4.6607$.

Inverse score (MOS-PESQ) from MOS-LQO is shown in Eq. (2):

$$x = \frac{D - \ln\left(\frac{B-y}{y-B}\right)}{C}, \tag{2}$$

where constants are $B = 4.999$, $C = 1.4945$, $D = 4.6607$

The Speech Quality Evaluation System [8] is connected to Speech Samples Recorder and Speech Samples Player, which provide the original and degraded speech sample.
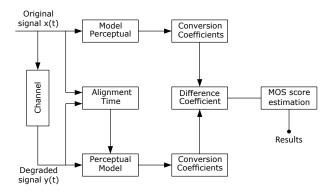

Fig. 3: PESQ schema.

The results are stored in the database with particular configuration parameters such as a number of transcoding and codec sequence in the transmission chain. The results can be exported in well computer processed formats csv or json. The transmission chain can be set via a web interface. When the required number of transcoding and particular codecs are set, the system originates a call with the prepared speech sample, evaluate the speech quality and store the results to the database.

3. Measuring Method

The paper has aimed at the measuring between the most widely used codecs G.711 A-law, GSM and G.729. The G.711 codec is considered as the mandatory codec for VoIP provider interconnection. It becomes from the reverse compatibility of VoIP with ISDN technologies, which uses only the G.711 audio codec. There are two

slightly different versions; μ-law, which is used primarily in North America, and A-law, which is in use in most other countries outside North America.

The GSM codec is the most widely used codec in GSM mobile networks. Apart from the G.711 codec, the GSM has almost the five times lower bit rate. The Full Rate GSM codec option was chosen for this measurement.

The G.729 codec offers a good ratio between speech quality and bit rate. Apart from the G.711 codec, the G.729 has the eight times lower bit rate. The most disadvantage is, that using G.729 requires a license per every channel. The Tab. 1 consists of the parameter comparison of codecs mentioned above:

Tab. 1: Codecs parameters.

Codec	Bit rate (kbit·s^{-1})	Sampling rate (kHz)	MOS
G.711	64	8	4.482
GSM FR	13	8	3.704
G.729	8	8	3.825

We have measured transcoding between all three codecs mutually [6]. The results were taken from 0 to 23 transcoding between two chosen codecs. The maximum transcoding number 23 was taken from the objective and subjective perspective. The objective perspective was the measured MOS value close to 2.0 which is defined as poor quality and the subjective perspective was listening of the recorded speech sample.

4. Results

The Fig. 4, Fig. 5 and Fig. 6 depict the impact of transcoding on the speech quality between all codecs. The lines are named according to the first codec, which is used for the first transcode of the original speech sample. The measurement was performed repeatedly with the similar results, because the call was transcoded locally without any packet loss or delay between endpoints. The first observation from all figures shows that the first transcoding from the better quality codec to the lower quality codec has the most significant impact on the speech quality. The speech quality decreases to the level of the lower quality codec. The second and the following transcoding has the similar impact on the speech quality for the both codecs and the decreasing is quite linear. The translation from the A-law to GSM codec decreases the speech quality for 0.2 MOS in average. The translation from GSM to A-law codec increases the speech quality with insignificant influence.

The translation from the A-law to G.729 codec decreases the speech quality for 0.15 MOS in average.
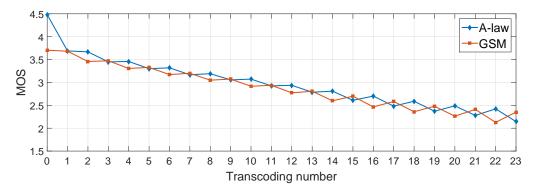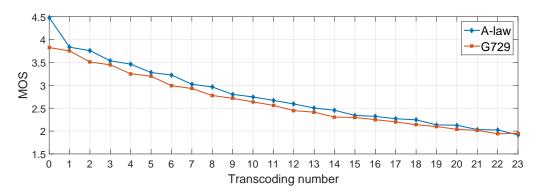
Fig. 4: Transcoding between A-law/GSM codecs.
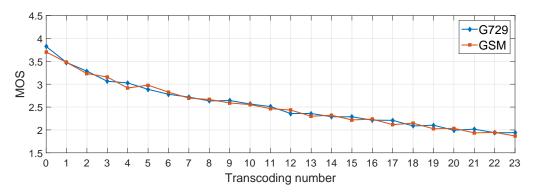

Fig. 5: Transcoding between A-law/G729 codecs.

Fig. 6: Transcoding between G729/GSM codecs.

The translation from G.729 to A-law codec increases the speech quality with the insignificant influence.

The translation from the G.729 to GSM codec decreases the speech quality for 0.1 MOS in average. The translation from GSM to G.729 increases the speech quality with insignificant influence.

Tab. 2: Speech quality variation according to the codecs translations.

Codec	A-law	GSM	G.729
A-law	0	-0.2	-0.15
GSM	0	0	0
G.729	0	-0.1	0

The most understandable results from the administrator's point of view are shown in the Fig. 7. It clearly shows how many percent of the original speech quality will be lose after particular number of transcoding between the different codecs. The A-law codec is the most depended on the loosing of the speech quality because its speech quality before transcoding is very high (4.482). On the other side there is the GSM codec with speech quality 3.704, which loose the least speech quality during the transcoding process.

The results were obtained in ideal network conditions without any influence of network parameters such as delay, packet loss, jitter or latency. The measure-
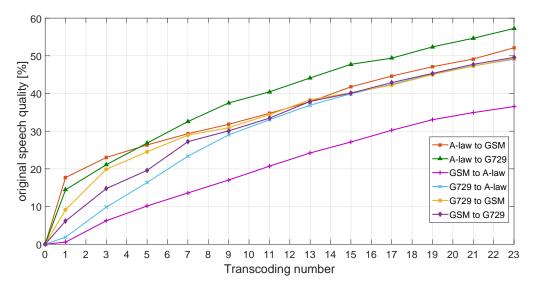
Fig. 7: Loose of original speech quality in percentage after transcoding.

ment ensures the degradation of the speech quality only from the codec transcoding perspective.

5.   Discussion

The results confirmed the premise that every transcoding has a significant influence on the final speech quality. The transcoding is one part of the VoIP network setting, which the VoIP providers are able to influence. G711 or GSM codecs were invented more than 20 years ago. Nowadays, there are totally different network conditions and more effective modern adaptive codecs such as Opus codec. The Opus codec is able to cover whole spectrum of audio record from narrowband, wideband, super-wideband to fullband audio spectrum. The new recommendations from telecommunication authorities consider Opus as the keystone of the future VoIP communications. Unification of the VoIP environment can rapidly increase quality of the VoIP services overall. The VoIP technology is one of the most related service on the network conditions because it ensures communication between the terminals in real time. The packet loss and delay are the most problematic network parameters, which influence the quality of VoIP service. In our future work, we are going to extend our measuring system for network parameter emulator. The end user, typically VoIP network administrator would be able to define network parameters and transcoding chain into the system. The system would emulate the VoIP call to get speech quality results between the end terminals in advance before implementation VoIP service into the network.

Acknowledgment

References

[1] PARTILA, P., M. KOHUT, M. VOZNAK, M. MIKULEC, J. SAFARIK and K. TOMALA. A Methodology for Measuring Voice Quality Using PESQ and Interactive Voice Response in the GSM Channel Designed by OpenBTS. *Advances in Electrical and Electronic Engineering.* 2013, vol. 11, no. 5, pp. 380–386. ISSN 1336-1376. DOI: 10.15598/aeee.v11i5.894.

[2] ROZHON, J. and M. VOZNAK. Development of a speech quality monitoring tool based on ITU-T P.862. In: *34th International Conference on Telecommunications and Signal Processing.* Budapest: IEEE, 2014, pp. 62–66. ISBN 978-1-4577-1411-5. DOI: 10.1109/TSP.2011.6043771.

[3] FAJKUS, M., M. MIKULEC, M. VOZNAK, M. TOMIS and P. FAZIO. Speech Quality Measurement of GSM Infrastructure Built on USRP N210 and OpenBTS Project. *Advances in Electrical and Electronic Engineering.* 2014, vol. 12, no. 4, pp. 341–346. ISSN 1336-1376. DOI: 10.15598/aeee.v12i4.1215.

[4] VOZNAK, M and J. ROZHON. Automated Speech Quality Monitoring Tool based on Perceptual Evaluation. In: *6th International Conference on Communications and Information Technology.* Wisconsin: World Scientific and Engineering Academy and Society (WSEAS), 2012, pp. 95–105. ISBN 978-1-61804-077-0.

[5] ITU-T Standard P.862. *Perceptual Evaluation of Speech Quality (PESQ): An Objective Method for End-to-end Speech Quality Assessment of Narrowband Telephone Networks and Speech Codecs.* Geneva: ITU-T, 2001.

[6] MAHDI, A. E. and D. PICOVICI. Advances in voice quality measurement in modern telecommunications. *Disital Signal Processing: A Review Journal.* 2009, vol. 19, iss. 1, pp. 79–103. ISSN 1051-2004. DOI: 10.1016/j.dsp.2007.11.006.

[7] MACURA, L., M. VOZNAK, K. TOMALA and J. SLACHTA. Embedded multiplatform SIP server solution. In: *35th International Conference on Telecommunications and Signal Processing (TSP).* Prague: IEEE, 2012, pp. 263–266. ISBN 978-1-4673-1118-2. DOI: 10.1109/TSP.2012.6256295.

[8] ITU-T Standard P.863. *Application guide for objective quality measurement based on recommendations P.862, P.862.1, and P.862.2.* Geneva: ITU-T, 2005.

About Authors

Martin MIKULEC received the M.Sc. degree from VSB–Technical University of Ostrava in 2011. Currently, he is working toward the Ph.D. degree at the Department of Telecommunications, (VSB–Technical University of Ostrava). His research is focused on advanced services in IP telephony and VoIP.

Marcel FAJKUS received his M.Sc. degree from VSB–Technical University of Ostrava in 2011. He is currently Ph.D. student, and he is currently working in the field of optical communications.

Miroslav VOZNAK is an Associate Professor with Department of Telecommunications, VSB–Technical University of Ostrava, Czech Republic. Topics of his research interests are Next Generation Networks, IP telephony, speech quality and network security.

Olle JOHANSSON is the CEO Edvina AB, Sweden and has more than 25 years of experience in the Unix and networking business, with ten years of VoIP experience. He is an Asterisk and Kamailio developer, trainer and consultant.