Performance Analysis of the Opus Codec in VoIP Environment Using QoE Evaluation

Péter Orosz, Tamás Skopkó, Zoltán Nagy, and Tamás Lukovics
Faculty of Informatics
University of Debrecen
Debrecen, Hungary
e-mail: oroszp@unideb.hu

Abstract—**VoIP has been a focus area of network communications for more than a decade now. The presence of VoIP traffic becomes more and more significant in the global Internet traffic. Although available access bandwidth is constantly increasing, higher capacity itself cannot guarantee higher quality of experience of the VoIP service. While QoE predicting methods are under active research, audio codecs also evolved greatly. The introduction and standardization of the Opus codec in 2012 is an important milestone in the voice codec evolution and Opus will probably be a royalty-free alternative for many VoIP applications in the near future. Past studies showed that audio quality of the Opus codec is superior when compared to almost every alternative. Mean Opinion Score (MOS) is a standardized scale for rating service quality. Our paper investigates the Opus codec in VoIP environment in terms of the relation between measured network QoS parameters and MOS value gained by subjective QoE assessments.**

Keywords—Opus codec; Speex codec; VoIP communication; speech quality; MOS; QoE evaluation.

I. Introduction

IP is a more and more preferred protocol for digital communication services and digital speech transport on IP (VoIP) is more and more significant in the global Internet traffic [1]. In the last decade, the evolution of real-time transport protocols and voice codecs resulted on an augmented user expectation in terms of service and speech quality. Ensuring high quality speech transmission is not a trivial task, since the service has to suit strict timing criteria. Voice communication is interactive and low latency (≤150 ms) is therefore a crucial requirement for the acceptable level of user experience. While mobile telecommunication companies operate dedicated infrastructure for speech transmission, provider independent VoIP sessions flow through heterogeneous public networks. It is more challenging to provide the sufficient level of service quality on a best effort infrastructure such as the Internet. Researches also point out that increasing bandwidth not necessarily ensures better subjective quality of the services.

Although the progression of VoIP technology is most visible on desktop platforms, an increasing number of users want to access these services using mobile devices. Lower computing performance and limited battery capacity make low code complexity a huge advantage for a voice codec. Simple PCM coding algorithms (like G.711 µLaw) were used for digital speech transmission from the beginnings. However, the increasing computing power in the last decade made possible to apply complex voice coding algorithms.

II. Evaluating Speech Quality

Codecs tolerate network transmission anomalies (i.e., jitter or packet loss) differently and their behaviors have a direct impact on the subjective Quality of Experience (QoE). Methods for predicting QoE are under active research and development. In these methods, measured flow level QoS parameters (delay, jitter, reordering, packet loss) are associated with metrics based on subjective quality evaluation of service users. After call termination, providers often ask their users about the quality of the recent call. The most popular form for the evaluation is the 5-score Mean Opinion Score (MOS) scale [2]. Of course, this simple scoring scheme makes no relation between the quality of experience and the effect of different network anomalies. More detailed assessment techniques can provide better correlation but in practice, users cannot be asked for detailed and reliable evaluation easily. Other methods are predicting subjective quality experience estimation applying mathematical statistical evaluation on the received audio samples. The International Telecommunication Union (ITU) has its own recommendation for standardized evaluation of speech quality: after many years of development (superseding PEAQ, PSQM, PESQ and PESQ-WB algorithms), POLQA (ITU-T P.863) is able to evaluate speech sampled up to 14 kHz using the MOS metrics [3].

QoE prediction methods and algorithms are based on the statistics of a large measurement dataset. Some methods require the original audio data, these are called Full-Reference (FR) designs, while methods not requiring the original material are No-Reference (NR) type ones. In practice, acquiring the audio flow is often not possible or not applicable (lack of full control of endpoints, storage capacity limitations, privacy restrictions), and therefore, constructing a reliable NR method is consequential.

The introduction of the Opus audio codec is a significant milestone in world of voice codecs. The codec standardized by the IETF in 2012 is derived from the combination of the previously existing SILK (focusing on speech transmission) and CELT (aiming low latency) codecs [4]. Like most advanced codecs, it supports constant as well as variable bitrates, and switching between rates with seamless transition. This feature makes possible to feedback altering network conditions (conjunction with Real-time Transport Protocol (RTP) [5] and RTP Control Protocol (RTCP) [6]). It is also effective for creating short audio clips because its algorithm does not need large code tables. Furthermore, it features advanced error correction. The correlation between audio frames can be adjusted, which controls how loss of audio frames affects voice quality. Also, optional Forward Error Correction (FEC) inserts redundant data (at the cost of some quality) to reduce the effect of packet loss. Opus is a new generation, universal audio codec, which is royalty-free in its every part. Its feature set, open source and industry support make presumably a popular audio codec for digital audio transmission over IP. In parallel with the standardization, a reference implementation of the codec library (i.e., encoder and decoder) is also developed and is freely available, which enables to evaluate the real-life performance of the Opus codec very effectively [7].

Although the audio quality was formerly evaluated (see Section III), the behavior of Opus codec under different network conditions has still to be investigated. We will sum up the works related to the Opus codec in Section III. In Section IV, a measurement setup for evaluating Opus in VoIP environment will be presented. The Opus codec had to perform on an emulated long distance network path implemented by our laboratory transport infrastructure. In Section V, a comparative performance analysis (set against its predecessor, the Speex codec) will be presented. Finally, Section VI concludes the presented work.

III. RELATED WORKS

Anssi Ramö and Henri Toukomaa evaluated Opus MDCT and LP modes with subjective listening tests and compared them with 3GPP AMR, AMR-WB and ITU-T G.718B, G.722.1C and G.719 codecs [8]. The paper keeps the codec a good alternative for the aforementioned codecs. The papers of C. Hoene et al. include different listening tests and compares the codec to Speex (both NB and WB), iLBC, G.722.1, G.722.1C, AMR-NB and AMR-WB [9][10][11][12]. They conclude that Opus performs better, though at lower rates, AMR-NB and AMR-WB still outperform the new codec. Jean-Marc Valin et al. present further improvements in the Opus encoder that help to minimize the impact of coding artifacts [13].

One of the motivational reason was that none of the researches above tested the Opus codec in VoIP environment. Moreover, the original or input audio signal is usually not available. Therefore, currently available FR-based methods cannot be applied by service providers. Our aim is to construct a NR method for predicting subjective QoE based upon measured QoS parameters.

IV. MEASUREMENT SETUP

An emulated long distance network path including two communication endpoints was constructed for the assessment of both codecs (Fig. 1). Endpoints feature generic multi-core x64-based architectures. They were equipped with Intel PRO/1000 NICs and interconnected with 2 m of industrial grade CAT6 cabling. Fedora Core 18 was installed to both hosts with unmodified Linux 3.8.1-x kernel (with a jiffy setting of 1000 Hz). We have chosen version 1.2.2 of the sflPhone VoIP application, since it supports Speex as well as Opus and its transmission parameters conformed the expected QoS performance (packet rate, uniform distribution of inter-arrival times and packet sizes) [14].

A carefully selected audio clip (easy to understand single channel of speech) was injected into the input of the softphone on Host A. JACK Audio Connection Kit is a general audio tool and is able to connect audio inputs and outputs of different applications and audio devices [15]. Current version of sflPhone can accept ALSA and PulseAudio datastream at its input. PulseAudio was selected since it can be directly connected with JACK. Since it has native output plugin (sink) for JACK, the audio clip was fed into JACK from an uncompressed PCM WAVE file with the GStreamer application. We carefully configured the applications not to perform unnecessary audio sample rate conversion throughout the digital audio path. The sflPhone application on Host B was configured to save the audio data into uncompressed PCM WAVE file for further QoE assessment. During the measurement we used the netem Linux kernel module, which was configured symmetrically on both directly connected interfaces to emulate a long distance path and produce various network anomalies that affect QoS (i.e., packet loss and variation of delay (jitter)). During the measurements we stored both the WAVE file from the receiver softphone and the PCAP files containing the received RTP stream [16]. The first 35 seconds of the original speech were used as input in all measurements. The network delay was set to 100 ms in each direction. The codecs were measured independently from each other, with the same series of parameters (Table 1). Netem network parameters were iterated using the following scheme:

TABLE I. MEASUREMENT PARAMETERS

Measurement series	**Opus**	**Speex**
Jitter (ms)	0, 1, 2, 3, ..., 20	
Packet loss (%)	1, 2, 3, ..., 40	
Combined: jitter (ms) and packet loss (%)	jitter: 1, 2, 3, ..., 10 loss: 1, 2, 3, …, 10	

The measurement sessions resulted in 160 audio clips per codec. As a reference of the evaluation, an initial measurement with zero jitter and no packet loss were run for both codecs.

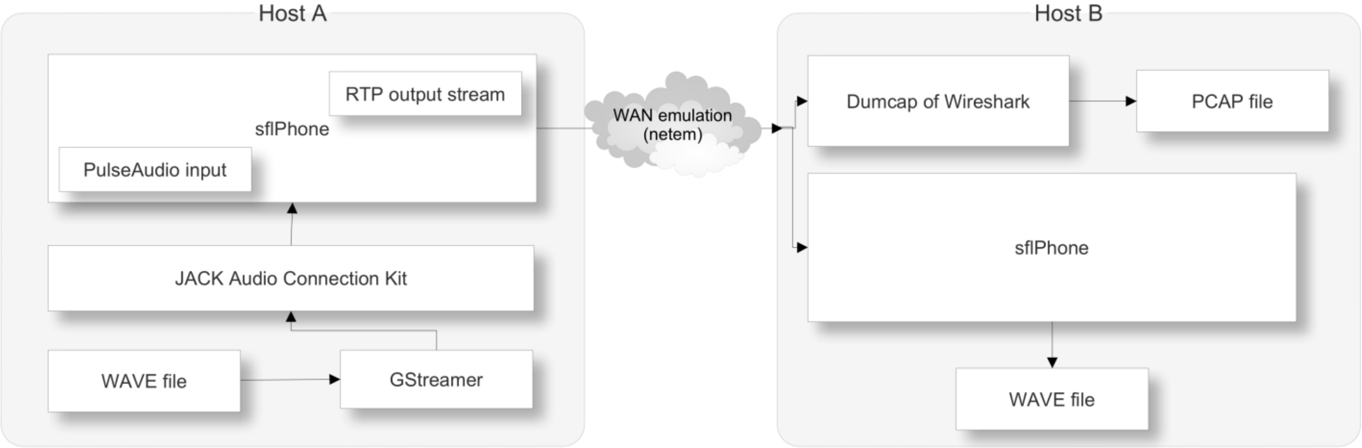

Figure 1. The measurement setup: audio is fed into the VoIP client on Host A and is transported through RTP to the other client on Host B.

V. EVALUATION OF THE MEASUREMENTS

We used 16 kHz sample rate for both codecs, since wideband (WB) operation mode is now a reasonable user claim. Both codecs were operated in variable bitrate (VBR) WB mode during the measurements. In case of the Opus testing, sflPhone generated 100 RTP packets per second, with variable packet size from 40 to 159 bytes that are sent out with a 8 ms (with a standard deviation of 500 µs) period. Opus was set to constraint VBR mode when the encoder assumes a transport with an average of the nominal bitrate and it creates one frame for the corresponding buffering delay (Fig. 2). The nominal bitrate was 64 kbps during the Opus measurements.

In case of Speex, 50 RTP packets were sent out per second, at an average period of 18 ms (with a standard deviation of 1 ms) and packet size was fixed to 124 bytes. The average bandwidth was 42 kbps (Fig. 3). Speex calls this setting "wideband". With a typical consumer access bandwidth, it is reasonable using such or even higher quality settings.

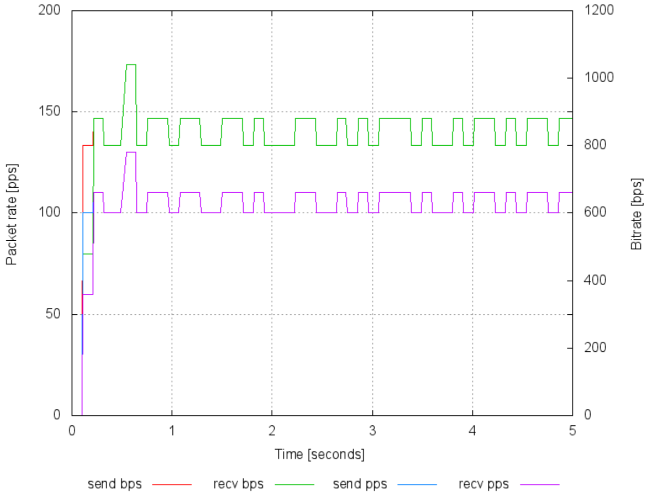

Figure 2. Opus: Packet rate and bandwidth during the voice transfer

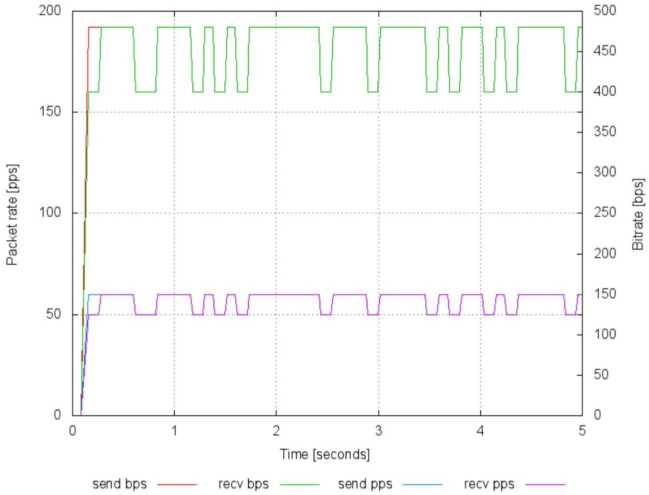

Figure 3. Speex: Packet rate and bandwidth during the voice transfer

All of the captured audio files were sequentially listened by users with sufficient amount of time for relax. The files were graded using the 5-point MOS scale.

A. *Jitter-sensitivity*

Under normal conditions, packets should arrive in a restricted time window to the decoder to maintain the real-time service. Variance of packet arrival are caused by infrastructural delay (routers can have queues with different priorities for forwarding) or by transient (longer burst than internal buffers allow) overload of the receiving endpoint. The jitter buffer of a real-time application is for holding the incoming packets and eliminating network jitter introduced by the infrastructure. The size of this buffer should be kept relatively small for the VoIP applications to achieve the required low latency performance. Packets arriving out of the expected time range are dropped.

Opus codec seems to be more sensitive to jitter but performs better than Speex at extreme conditions (see Fig. 4). Opus produced better voice quality at low jitter. Furthermore, even at 11 ms of jitter, the decoded voice was still more understandable than with Speex. None of the users gave 5 points for the Speex performance even at the smallest amount of jitter (1 ms) since it caused not annoying but clearly audible clicks. From the aspect of jitter, Speex gives average performance at a wider range but Opus provides higher voice quality under 4 ms of jitter and is easier to listen to under heavier network perturbation.

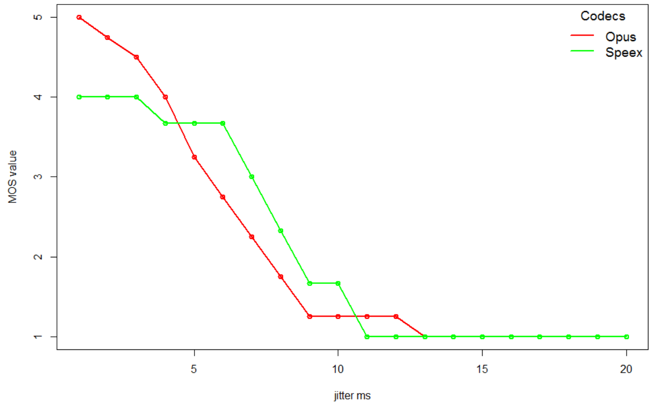

Figure 4. Correlation between jitter and subjective quality of experience expressed in MOS

B. *Loss-sensitivity*

Packets can be lost throughout the network path (e.g., inside a router) or at the endpoint itself. It is difficult to evaluate how efficient the codecs are in the compensation of information loss.

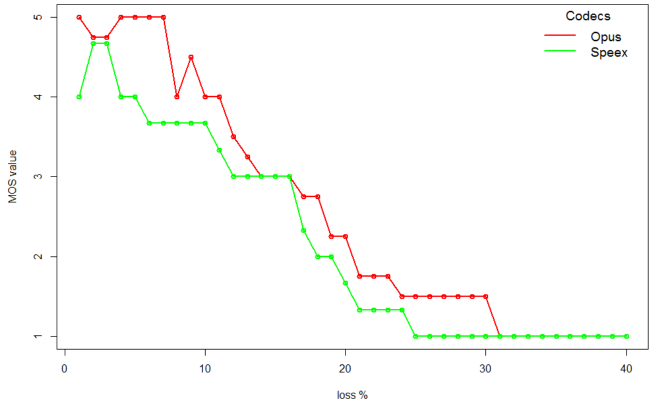

Figure 5. Correlation between packet loss ratio and subjective quality of experience expressed in MOS

As seen in Fig. 5, the Opus codec smoothes the effect of packet loss more efficiently. While Speex is gibberish even at 25% packet loss (using 802.11 access, it is not an unrealistic situation), Opus still gives acceptable result up to 30% of loss. Further observation is related to the split opinions at low packet loss: Opus codec performed better at 2% of loss that at 1%. This result may reflect the fact that a higher performance psychoacoustic model is working inside the codec. At a particular loss, quality of experience with Opus decreases less than with Speex.

C. *Sensitivity for multiple anomaly*

Anomalies detailed in the previous subsections are rare to occur alone. In reality, some combination of jitter and packet loss should be expected. Accordingly, we executed a complex measurement session to evaluate the audible effect of the presence of both jitter and packet loss.

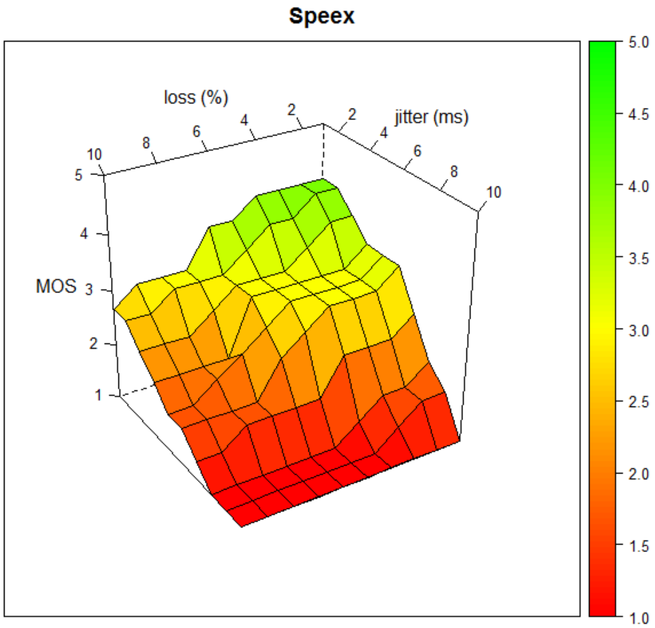

Figure 6. MOS values for the Speex codec under mixed network conditions

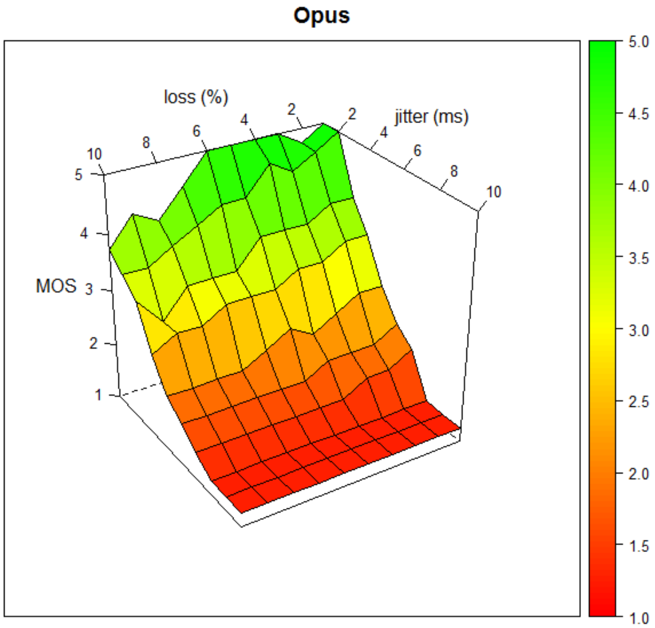

Figure 7. MOS values for the Opus codec under mixed network conditions

Figs. 6 and 8 show that Speex got 3 MOS points in a wide range of the investigated network parameters. In the MOS scale 3 points equivalent with the lower bound of the acceptable quality. It never reached 5 score even at small amount of anomaly. In contrast, Opus performs uniformly until its boundaries (see Fig. 7). Although jitter error affects its quality more drastically, it is more tolerant to loss than Speex. The QoS-QoE relationship of the Opus codec in terms of jitter is more close to linear than that of Speex (Fig. 7).

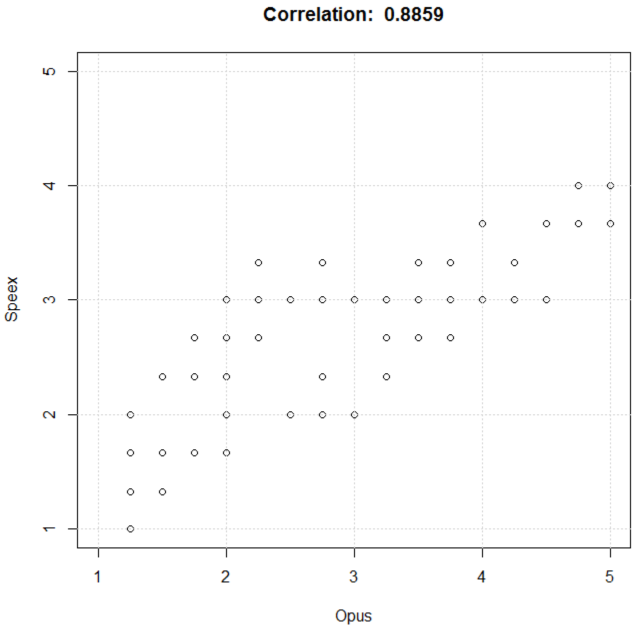

Figure 8. Correlation between Speex and Opus MOS scores for all of the combined measurement scenarios.

As presented earlier in this section, the QoE assessment assigned a MOS value for each measurement. According to our combined measurement series (both jitter and loss are present on the emulated network path) now we got 100 MOS values for both codecs. The correlation of the two MOS series, which is calculated from (1) is presented on Fig. 8.

$$\rho_{X,Y} = \frac{\mathrm{cov}(X,Y)}{\sigma_x \sigma_y} = \frac{E[(X-\mu_x)(Y-\mu_Y)]}{\sigma_x \sigma_y} \tag{1}$$

where μ is the expected value of the random variable and a σ is its standard deviation. Using two variable polynomial regression and Sum of Squares due to Error (SSE) goodness-of-fit statistics, we found that jitter and MOS values show a linear relation, while loss and MOS values suggest a quadratic relationship. In the near future, our goal is to construct a low order estimator function for calculating MOS values based on packet level QoS parameters (i.e., loss and jitter). This estimator function could be the basis of a NR-type objective QoE assessment method for Opus based VoIP conversations.

VI. CONCLUSION

In this paper, the fault tolerance of the royalty-free Opus and Speex VoIP codecs has been evaluated using laboratory QoS measurements and subjective QoE assessments. Although their roots are the same, under the investigated conditions Opus performs more uniform when multiple network anomalies of jitter and packet loss are present. Since there is no NR method available for speech quality prediction available, close-to-linear relationship between measured jitter and the gained subjective QoE values of Opus codec make possible to create a NR method to estimate QoE from the measured QoS parameters. We are actually moving this way on. We also note that Opus' Forward Error Correction option for transmitting redundant information is another important feature that has to be evaluated in a future work.

ACKNOWLEDGMENT

The work was supported by the TÁMOP 4.2.2.C-11/1/KONV-2012-0001 project. The project was implemented through the New Széchenyi Plan, co-financed by the European Social Fund.

This research was supported by the European Union and the State of Hungary, co-financed by the European Social Fund in the framework of TÁMOP 4.2.4.A/2-11-1-2012-0001 'National Excellence Program'.

REFERENCES

[1] Point Topic Ltd.,VoIP Statistics – Market Analysis, Q2 2012, October 2012

[2] ITU-T P.800: Methods for subjective determination of transmission quality, August 1996

[3] ITU-T P.863: Perceptual objective listening quality assessment, January 2011

[4] JM. Valin, K. Vos, and T. Terriberry, "IETF RFC 6716: Definition of the Opus Audio Codec", September 2012

[5] Real-time Transport Protocol, http://tools.ietf.org/html/rfc3550 [retrieved: September 2013]

[6] RTP Control Protocol, http://tools.ietf.org/html/rfc3550 [retrieved: September 2013]

[7] Opus Codec Downloads, http://www.opus-codec.org/downloads/, [retrieved: August, 2013]

[8] A. Ramö and H., "Voice Quality Characterization of IETF Opus Codec"INTERSPEECH, pp. 2541-2544. ISCA, 2011

[9] JM. Valin, K. Vos, and J. Skoglund, "Summary of Opus listening test results draft-valin-codec-results-00" June, 2011

[10] C. Hoene, Ed., JM. Valin, K. Vos, and J. Skoglund, "Summary of Opus listening test results draft-valin-codec-results-01", May, 2012

[11] C. Hoene, Ed., JM. Valin, K. Vos, and J. Skoglund, "Summary of Opus listening test results draft-valin-codec-results-02", May, 2012

[12] C. Hoene, Ed., JM. Valin, K. Vos, and J. Skoglund, "Summary of Opus listening test results draft-valin-codec-results-03", November, 2013

[13] JM Valin, G Maxwell, TB Terriberry, and K. Vos, "High-Quality, Low-Delay Music Coding in the Opus Codec, " 135th AES Convention , October 2013

[14] sflPhone, http://sflphone.org/, [retrieved: August, 2013]

[15] JACK Audio Connaction Kit, http://jackaudio.org/, [retrieved: August, 2013]

[16] J. Spittka, K. Vos, and JM. Valin, "RTP Payload for Opus Speech and Audio Codec draft-ietf-payload-rtp-opus-01", August 2013