



**HAL**  
open science

## Determination of Bit-Rate Adaptation Thresholds for the Opus Codec for VoIP Services

Yi Han, Damien Magoni, Patrick Mcdonagh, Liam Murphy

► **To cite this version:**

Yi Han, Damien Magoni, Patrick Mcdonagh, Liam Murphy. Determination of Bit-Rate Adaptation Thresholds for the Opus Codec for VoIP Services. IEEE International Symposium on Computers and Communications, Jun 2014, Madeira, Portugal. pp.1-7, 10.1109/ISCC.2014.6912585 . hal-01009398

**HAL Id: hal-01009398**

**<https://hal.science/hal-01009398>**

Submitted on 3 Nov 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Determination of Bit-Rate Adaptation Thresholds for the Opus Codec for VoIP Services

Yi Han\*, Damien Magoni<sup>†</sup>, Patrick McDonagh<sup>‡</sup> and Liam Murphy\*

\*School of Computer Science and Informatics, University College Dublin, Dublin, Ireland

Email: yi.han@ucdconnect.ie, liam.murphy@ucd.ie

<sup>†</sup>LaBRI - CNRS, University of Bordeaux, Talence, France

Email: magoni@labri.fr

<sup>‡</sup>Lero, Dublin City University, Dublin, Ireland

Email: patrick.mcdonagh@dcu.ie

**Abstract**—In this paper, we present an experimental evaluation of the recently standardized Opus codec used in a VoIP context. Opus operates in both narrow and wideband modes, similar to Adaptive Multi-Rate (AMR). Through the use of the Wideband Perceptual Evaluation of Speech Quality (WB-PESQ) metric, we have conducted an extensive set of experiments using multiple audio samples encoded at different bit-rates, to investigate the impact of packet loss on resulting speech quality. Using these results, fitting functions for each bit-rate were computed to provide a straightforward manner of evaluating speech quality when given a specified packet loss rate. Using *ns-2*, a simulation analysis was conducted to evaluate the effect of background traffic on transmitted Opus streams. We observed that, when using different levels of background traffic, the observed packet loss rates varied heavily depending on the stream bit-rate. By correlating this information with the fitting functions derived previously, we were able to define switching thresholds. These are points where the speech quality of a lower bit-rate stream is greater than that of a higher bit-rate stream for the same levels of link bandwidth saturation.

## I. INTRODUCTION

In VoIP systems the codec heavily influences the resulting speech quality. This is because it specifies components such as compression algorithms and encoding modes, allowable bit-rates, frame sizing, etc. Narrowband codecs operate on frequencies between 300 Hz and 3.4 kHz, while wideband codecs operate on a larger range of frequencies resulting in higher call quality.

Multiple wideband codecs have been released in an effort to improve the quality of encoded speech. Examples of these include G.722 [1], AMR-WB [2] and G.722.1 [3]. The Opus codec [4], standardized in September 2012, is designed to be an open codec, supporting many audio applications. Opus is comprised of a modified SILK codec [5] and the CELT codec [6][7]. Opus supports bit-rates from 6 kbps up to 510 kbps, with sampling rates from 8 kHz to 48 kHz, as detailed in [4], making it applicable for many usage scenarios. While limited research has been conducted using the Opus codec, it has been shown to be comparable and in some cases, it provides a better quality than many popular codecs for similar bit-rates [8]. Opus also allows for the use of a variety of frame sizes, from 2.5ms up to 60ms [4]. Using larger frame sizes reduces the number of packets transmitted per second, making more efficient use of bandwidth by minimizing the overhead of protocol (MAC/IP/UDP/RTP) headers. However, this increases

latency, as audio frames have an increased length for encoding and sensitivity to packet loss, as more audio data is contained per packet. Similarly to most codecs used for VoIP, a frame size of 20 ms is used throughout this paper.

An extensive set of experiments were conducted using the Opus codec in order to investigate the performance of different bit-rates under varying packet loss rates. The ITU-T standard wideband codec measurement tool WB-PESQ [9] was used to provide a Mean Opinion Score (MOS) for call quality for each test case. These were conducted using multiple audio samples and were run to a 99% confidence level with a 1% error (Section III-B). The reason for doing this was to allow us to derive a fitting function for each bit-rate which indicates the expected call quality as the packet loss rate increases, without requiring real-time WB-PESQ analysis (Section III-C). Thus, this provides a more straightforward method of assessing call quality for the Opus codec. Comparing packet loss with delay and jitter, our experiment indicated that, in a low delay and jitter scenario, with limited small buffer size, the call quality degradation, in terms of MOS, resulting from delay is limited to lower than 4%, while packet loss is the dominant factor in quality degradation.

Furthermore, a network simulation was carried out using *ns-2* to investigate the impact of increasing background traffic on a bandwidth-constrained network. The network environment with continuously increasing background traffic was simulated and the packet loss of an Opus VoIP call was monitored. Here, we only focus on the effect of packet loss for each bit-rate. From these experiments, it can be observed that different bit-rates experience varying levels of packet loss for a given level of congestion. We then combine these simulation results with the fitting functions from the PESQ analysis in order to find the relation between bit-rate, packet loss rate, and call quality. Furthermore, our results show that having a mapping from packet loss to call quality between different bit-rates allows for the determination of points where adapting the bit-rate results in a better call quality (Section IV-B).

## II. RELATED WORK

As network conditions can vary widely, particularly in wireless scenarios, the quality of VoIP services can be improved by adapting the bit-rate accordingly. This can be achieved through the use of an adaptive codec such as AMR or

Opus. With Opus' support for real-time adaptation of bit-rates, other, lower quality codecs are not required when adapting the bit-rate, thus noticeable silent gaps in the audio can be avoided [10]. Results comparing the quality and latency of Opus to other codecs are presented on the Opus website [8]. They show that the Opus codec MOS values outperform the other codecs for bit-rates higher than 16 kbps, and that they are similar to AMR and better than both iLBC and Speex for bit-rates between 16 kbps and 8 kbps. Several Opus listening tests [8] were undertaken by different groups and it was shown that for 11 kbps, 15 kbps and 20 kbps bit-rates, Opus achieved higher values for MOS than Speex (narrowband), iLBC, and Speex (wideband) respectively. In addition, the results also indicate that the encoding delay of Opus is lower than the ones of other popular codecs for bit-rates up to 80 kbps.

The authors of [11] investigate the quality of the 3 modes of the Opus codec: LP mode (SILK based), hybrid mode and MDCT mode (CELT based). According to their results, the LP mode of the Opus codec yields a moderate voice quality, comparable to AMR and AMR-WB for similar bit-rates. When using the hybrid mode, at 20kbps Opus achieves similar quality to the LP mode, while at 24kbps, it produces higher quality than G.722.1C [3]. They also demonstrated that the MDCT mode does not produce as high a quality as the hybrid mode, when used for voice applications. The authors of [12] investigate the performance of Opus for VoIP, however their analysis used only a single bit-rate configuration.

The WB-PESQ standard [9] is an extension of the PESQ standard [13]. This wideband extension predicts the subjective score that would be attributed for a wideband speech sample by a panel of listeners. The POLQA metric was not used as the toolkit was not directly available. However, a direct comparison between the MOS results of WB-PESQ and PESQ is not possible due to their different target applications and the fact that they are based on a subjective scale. According to the authors of the WB-PESQ standard [9], the MOS results for a given speech quality level can be impacted by the language used. Japanese and Korean languages [14] are known to yield higher MOS values when subjected to the same amount of signal distortion. In this paper, we thus only study American English samples, leaving the study of other languages for future work.

The survey described in [15] investigates application layer mechanisms for speech quality adaptation in VoIP, including bit-rate control, Forward Error Correction (FEC) control, play-out scheduling adaptation, etc. Moller *et al.* have shown in [16] that switching from one codec to another can improve the speech quality, if the former codec was suffering a high packet loss rate and if the latter codec significantly reduces the packet loss impact on the speech quality. Furthermore, in [17], the authors note that decreasing the transmission rate by increasing packet size is not an effective method of ensuring voice quality, possibly due to the fact that more voice data is lost when a packet is lost. To combat this, the authors proposed a similar switching method to [16]. However, in both of these papers, handover was investigated between two different codecs rather than one supporting real-time self-adaptation of its bit-rate. In this paper, we focus on the same idea applied to the Opus codec only, as it is able to provide a dynamically adaptable bit-rate.

### III. OPUS PACKET LOSS PERFORMANCE ANALYSIS

Assuming all other encoding parameters are maintained at the same value, a higher bit-rate results in better quality. When using a frame size of 20 ms, suggested bit-rate ranges for different applications are provided in [4]. In this paper, we investigate Opus with 20 ms frame size, which is the most common frame size with other popular codecs. In this way, the results in this paper can be compared to those from other codecs.

As existing wideband codecs such as AMR-WB support bit-rates up to 20 or 24 (limited use) kbps and 20 kbps is the highest bit-rate supported in Opus LP mode, 3 bit-rates from 8 to 20 kbps were chosen to be the focus for our experimental evaluation. Additionally, through our experiments, it was found that PESQ-WB gives more accurate results when all the test samples are in one single mode of the Opus codec. Inconsistent results were observed in our experimental evaluation when analyzing two samples that are encoded by two different modes, i.e. LP and Hybrid mode.

#### A. Experimental Setup

The aim of this experiment was to investigate Opus performance, under varying packet loss rates. The Opus development package (opus-tools v0.1.8 and libopus v1.1-rc3) from [8] was used for encoding and decoding voice samples using suggested samples from the ITU [18], along with a voice sample from the Opus website [8]. The sample from [8], (referred to as 'original') is uncompressed speech, consisting of a continuous speech by 2 males and 2 females, each for one sentence for a total duration of 10 seconds. The 4 samples selected from [18] are uncompressed speech in American English from 2 males and 2 females separately. Each clip is approximately 8 seconds long, each has two sentences separated by a 2 second period of silence. These are referred to as Male 1 & 2 and Female 1 & 2. This matches the recommended test sample duration of 8 to 12 seconds, with 2 to 4 sentences separated by silence periods [9]. All the samples had a 16 kHz sample rate and 16 bits per sample. The samples were encoded at 20 kbps, 14 kbps and 8 kbps from their original audio sample, using the OPUS\_APPLICATION\_VOIP flag.

Furthermore, to study the effect of transcoding (in which an intermediary node re-encodes the audio at a lower bit-rate), we also tested samples that were transcoded from 20 kbps encoding to 14 kbps and 8 kbps, and also from the 14 kbps encoding to 8 kbps. In a two person call, it would be likely that the source will adapt the bit-rate should this be necessary. However, in a multi-party call, the link qualities of participants may vary. Therefore, some callers may be able to support the high quality audio, while others can only support the medium or low quality audio. The use of transcoding enables this to be achieved without sacrificing the call quality for all participants. Thus, for the sake of completeness, we included this scenario in our experimental evaluation.

Figure 1 shows the experimental framework. Experiments were conducted using the Opus codec implementation available at [8]. A set of samples in PCM format were used as input to the Opus encoder. Opus encoded the samples using the specified bit-rate and generated an Opus format audio file as output. The Opus decoder then took the encoded Opus

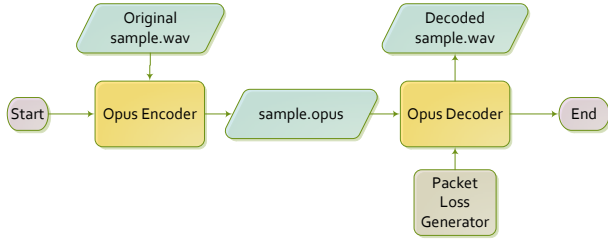


Fig. 1. Analysis Workflow

audio file and decoded it to a Wave file in PCM format for comparison with the original audio sample. We have modified the Opus decoder to introduce a time-based random packet loss generator that, when given a desired packet loss rate  $X$ , each packet that is packetized from the Opus format audio has a chance of  $X\%$  (i.i.d) of being dropped and replaced by an empty (null) payload packet. The decoded Opus audio, along with the original sample were input to PESQ-WB software and a call quality indicator for MOS was given.

It is not uncommon to have above 10% packet loss in a wireless connection [19], therefore we extended our analysis to a loss rate of up to 20% to have a better understanding of the audio quality degradation introduced by a larger range of packet loss rates. Packet loss rate experiments from 0% to 20% were conducted in 1% increments. For each audio sample and packet loss rate, experimental runs were repeated until the mean value over the set of experiments meets a confidence level of 99% with 1% margin of error. An initial set of 20 runs were completed before the first mean value is calculated, avoiding early termination of the experiment if the first few outcomes produce similar scores. After these 20 runs, the mean value is then calculated after every test run, until 99% of the output MOS values are in 1% error range of the mean value. Repeating the experiments for a large number of tests ensures a smooth and a more reliable indication of degradation over different packet loss rates.

### B. Experimental Results

The call qualities for each of 5 samples for each bit-rate at every packet loss rate point were calculated and shown in Figure 2. This figure not only shows impact of packet loss for each individual sample Figure 2(a-e), but also shows the MOS values averaged across all 5 samples Figure 2(f).

The results in Figure 2 have some variance from each other for different reasons. The ‘original’ sample results in (a) degrade faster than the others due to its different composition, in terms of more speech content contained in the sample. This is four sentences spoken by 2 males and 2 females in a single audio sample, leaving less silence, thus making it more sensitive to increasing packet loss. The results of the Female 1 sample in (b) indicate that it is more robust to packet loss, but the 8 kbps stream that is encoded from uncompressed audio drops faster than the stream transcoded from the already compressed 14 kbps and 20 kbps streams. This may be due to a combination of the speech pattern in the sample and the fact that in the directly encoded 8 kbps case, higher quality

audio is lost, leading to a larger degradation overall. This experiment was repeated to verify this behaviour. However, further investigation of distribution of loss for the Female 1 sample is required. Comparing the Female 2 sample (c) with the Male 2 sample (e), the Female 2 sample has lower MOS values than the Male 2 sample for similar packet loss rates. This is mainly because PESQ tool is known to underestimate the MOS of female audio samples [20]. However, this situation is not always true, if we compare the Female 2 sample with Male 1 (d), the results of these two samples are well correlated.

A large number of tests based on a single test sample will not always ensure a smooth curve and is not always applicable to other samples. However, using the average of 5 samples from different sources with both Male and Female voices reduces the sample dependency effect. Figure 2 (f) shows average results of 5 samples. This average plot gives a good estimation of how Opus codec performs over packet loss rates from 0% to 20% in a more general context, without the bias of a single sample. Additionally, since the transcoded streams already have the effect of quality degradation from a previous encoding (the output of which was then used as input for transcoding), the curve of transcoded stream for 20 kbps to 14 kbps is of a slightly lower quality than directly encoded 14 kbps stream, but still much higher than the three 8 kbps streams. For these three curves, the directly encoded 8kbps stream performs slightly better than the streams transcoded from 20 to 8kbps and 14 to 8kbps respectively, as expected.

Another observation from the results is that, on average, no matter where the stream is transcoded from, tests with higher bit-rate always have better quality than those with a lower bit-rate. Furthermore, the 8 kbps mode in Opus, which would be employed in low bit-rate transmission modes, has a significant degradation in quality from the 14 kbps mode, much greater than the difference between 20 kbps and 14 kbps. While the loss rate increases, the difference between MOS values decreases significantly. However, the difference between each bit-rate is still noticeable when the MOS is in an acceptable human perception range (MOS above 2.0).

### C. Opus MOS Fitting

Based on the results presented in the previous section, the performance of the Opus codec can be described using mathematic model, as shown in Figure 3. This can then be used in other situations to ascertain the call quality for a given packet loss / bit-rate combination. While the E-Model provides a method for estimating call quality when given network characteristics, such as delay, packet loss rate, jitter and the current codec, it is highly dependent on the codec used. The impairment introduced by the codec has to be ascertained and then standardized by a large experimental evaluation. Currently, there are only ITU standard narrowband codecs, along with a few non-ITU narrowband codecs that have been widely tested and their impairment factors standardized. Additionally, the wideband extension of E-Model [21] requires that the packet-loss robustness factor (Bpl) corresponds to the codec, packet size and packet loss concealment (PLC) used. For our purposes, as the PLC technique can vary in the Opus codec, we used the previously detailed method of inserting an empty packet. For a more straightforward estimation of the quality for a given bit-rate / packet loss rate combination and

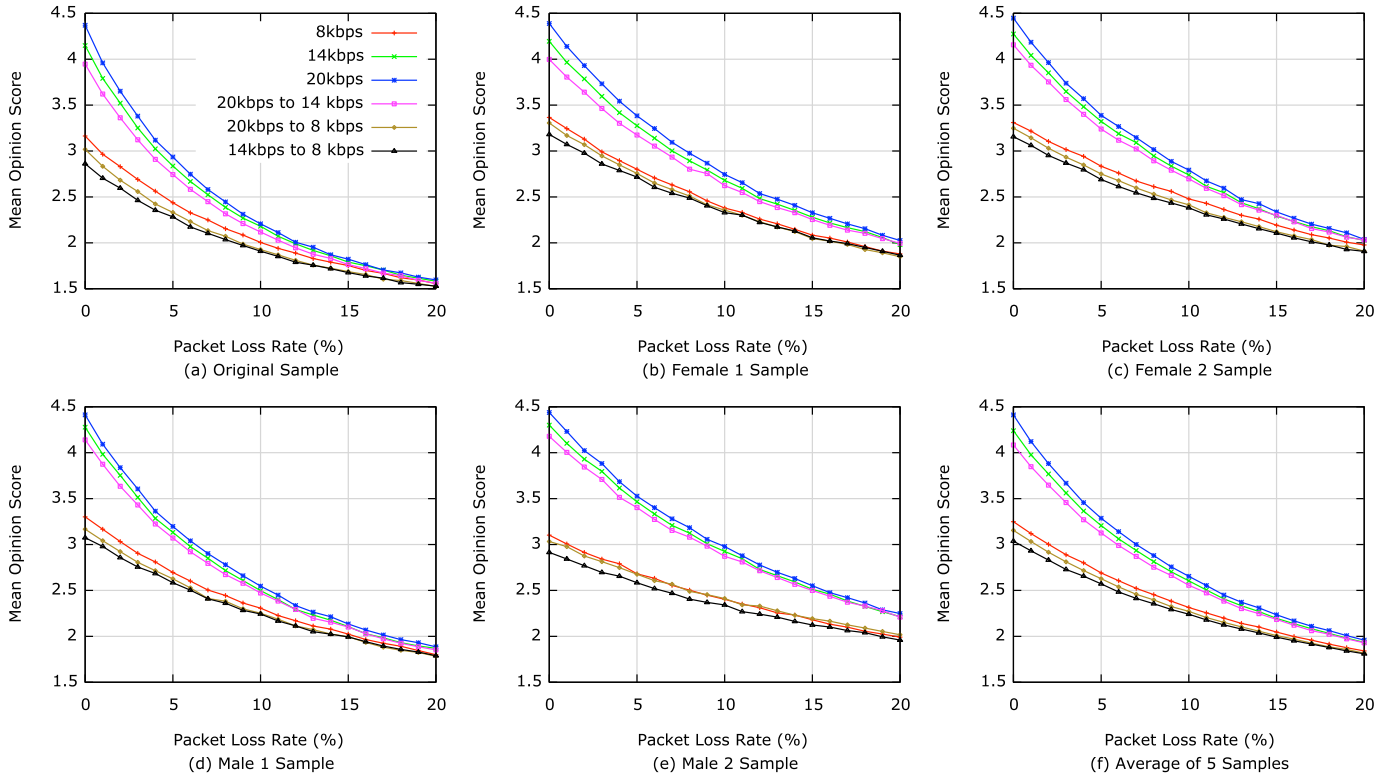


Fig. 2. MOS Results from PESQ-WB under Different Packet Loss Rates

assuming the same buffer size, a polynomial function is used to map packet loss rate to MOS from the previous WB-PESQ analysis for each tested Opus bit-rate. Each fitting function has a squared correlation coefficient above 0.99 to the results in Figure 2(f) which indicates a high quality fit.

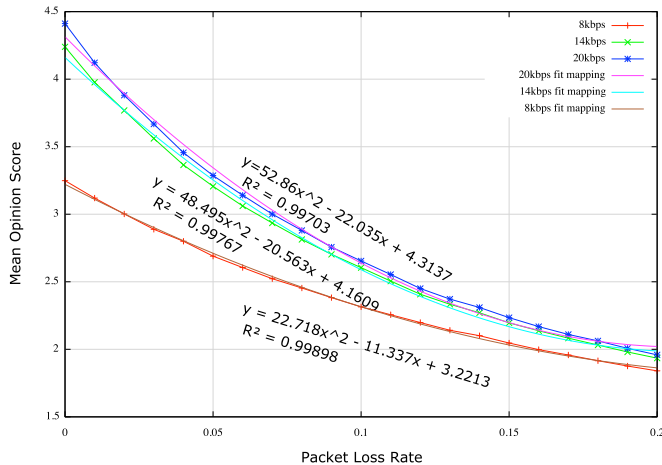


Fig. 3. Opus MOS Fitting Function

These mappings are then used in the following simulation experiment to further study the effect of packet losses and quality degradation in a congested network environment and to derive the appropriate bit-rate switching thresholds.

#### IV. OPUS CODEC SIMULATION-BASED EVALUATION

The purpose of this experiment was to further evaluate the performance of the Opus codec by employing it in a congested network scenario. Given a bandwidth-limited link, as background traffic increases, the network becomes congested and as a result all packets transmitted over that link will compete for transmission, leading to packet loss. By analysing the performance of the different bit-rates and understanding the relationship between the degradation curves, while introducing congestion, the benefits of bit-rate adaptation can be assessed. *ns-2*<sup>1</sup> was used to simulate the network topology, applications (VoIP traffic) and to provide monitoring output. The reason for doing this is to verify the assumption that a lower bit-rate Opus stream has a lower packet loss rate than a higher bit-rate Opus stream for a given background traffic level. This demonstrates that switching to a lower bit-rate results in better call quality, as well as extra bandwidth savings.

##### A. Simulation setup

In order to model a real VoIP call scenario, a single call using Opus was simulated. This is then perturbed by background traffic to saturate a bandwidth-limited link. This background traffic was simulated as a constant bit-rate application, with a fixed packet size of 62 bytes, where the transmission interval between packets decreases as the required background traffic level increases for each experiment. The 62 byte packet size represents 20kbps VoIP traffic, where the VoIP payload per packet is 50 bytes and an additional 12 bytes

<sup>1</sup><http://www.isi.edu/nsnam/ns/>

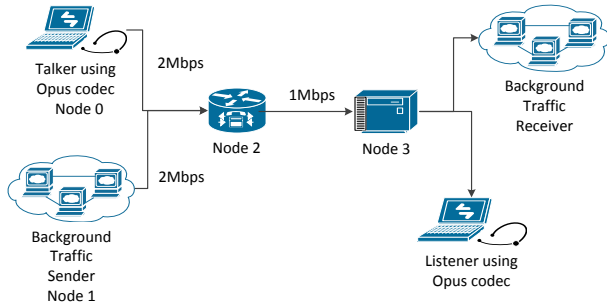


Fig. 4. Simulation Topology

are used to represent the RTP header in ns-2. As shown in Figure 4, Node 0 was used in this experiment to place the “monitored” call. Node 1 was used to simulate the source of background traffic. Node 0 and 1 send their traffic via Node 2 to Node 3. Node 0 and Node 1’s connections to Node 2 are modelled as 2Mbps links. Node 2 then forwards their packets onto the 1 Mbps bottleneck link connected to Node 3. All links have a queue size of 50 packets and a drop tail policy enforced. Node 3 has two receivers, one for receiving the background traffic from Node 1, and one receiver for the monitored Opus call from Node 0. Each received or dropped packet is recorded, including the reception or dropping time, packet size, flow id, etc. The VoIP flow from Node 0 to Node 3 through Node 2 was monitored and its packet loss rate was calculated. Measurements for delay were performed, however the maximum delay encountered was 25ms, which should only have a minor effect on call quality. A brief experiment was conducted using the E-Model with parameters for another codec and the effect of delay was minimal (causing between 1% to 4% degradation in MOS). Jitter is considered in loss by using a limited buffer length. This will lead to packet loss when the delay added by jitter exceeds the buffer length. The monitored VoIP flow started at 0 seconds and ran until the end of simulation at 50 seconds. Background traffic lasted for 30 seconds, starting at 10 seconds into the simulation, running to 40 seconds. The simulation was repeated for background traffic levels ranging from 0.97 Mbps to 0.999 Mbps in 0.5 kbps increments.

The packet loss was calculated for every newly received packet using a sliding average over the last 4 seconds from the current packet reception time. The authors of [22] found that the listener would not perceive the call quality degradation immediately if it drops instantaneously, but as time goes by the impairment would become noticeable and more annoying to the listener. There is a recency effect that a listener would remember call quality and start to notice the good-to-bad transition in about 5 seconds and 15 seconds for the bad-to-good transition. Thus, if remedial action can be made in the first 4 seconds to maintain the call quality, this will not alter the quality of user experience.

As discussed previously, the Opus tools collection were used for the previous experiment. The output of the encoder is a variable bit-rate audio file, with a mean bit-rate corresponding to a target bit-rate. The OpusRTP tool, which can be used for the transmission of Opus content over RTP, produces a list

of all the packets contained within the audio file, along with each packet size. Our simulation then generates packets of the exact size in the sample and then sends these packets every 20 ms (corresponding to the frame size). In this way, the simulated audio streams have the exact same parameters as the previously mentioned samples. Thus, the fitting described previously can then be used to calculate the corresponding call quality, given a packet loss rate for that bit-rate.

## B. Results Analysis

As the experiment was repeated for different levels of background traffic, each experiment generates a varying packet loss pattern like the one shown in Figure 5. In this figure, we have shown where the median value is located. (The label corresponds to the time of occurrence for the median value and its value, respectively). This figure shows the packet loss rate under a constant level 0.994 Mbps background VoIP traffic. Here, we can observe an increase in packet loss rate as the links and consequently, buffers in the network become congested, leading to packets being dropped. Figure 5 shows the performance for one particular background traffic level and clearly shows the packet loss behaviors of the 3 different bit-rates in our experiment.

As can be seen in Figure 5, the 20kbps stream begins to lose packets at 11.38 seconds and 14 kbps stream loses packets slightly later at 12.08 seconds. However, the 8 kbps stream maintains its 0% packet loss rate until 14.32 seconds (4.32 seconds after the background traffic begins). It can also be noted that the rate of increase for packet loss for the 8 kbps stream is much lower than that of the higher bit-rates. This means that it takes more time to reach to its peak packet loss rate for a given background traffic level.

Figure 6 shows the packet loss curve for increasing levels of background traffic up to 0.999 Mbps, just below the link bandwidth of 1 Mbps. These results were the average result of the 5 tested audio samples at each simulated background traffic level. For each simulation, the median value of the packet loss rates as shown in Figure 5 were used. The 20 kbps stream starts to lose packets at 0.9775 Mbps of background traffic. The 14 kbps stream begins to lose packets at 0.9825 Mbps, while 8 kbps stream remains at 0% loss rate until the background traffic reaches 0.9895 Mbps. Furthermore, the rate of increase for each stream is different. The 20 kbps stream is more sensitive to background traffic as shown by the higher packet loss rate. The sensitivity to background traffic decreases as the stream bit-rate is lowered, with the stream of 8 kbps being least sensitive.

As the average of 5 samples were used, this leads to some variability in the levels of packet loss for a given background traffic level, as a result of the variable bit-rate of each sample’s speech content. Our experiments also demonstrated that the ns-2 simulator sometimes produced a fixed pattern of packet losses. This is apparent in the significant drop for the 20 kbps stream (and less so, for 14 kbps) under 0.992 Mbps of background traffic. Upon further investigation, for background traffic with 62 bytes per packet and at a rate of 0.992 Mbps, it was found that these particular values produce exactly 2000 packets to be transmitted per second for the background traffic. This combination produces unusual results, in the case of

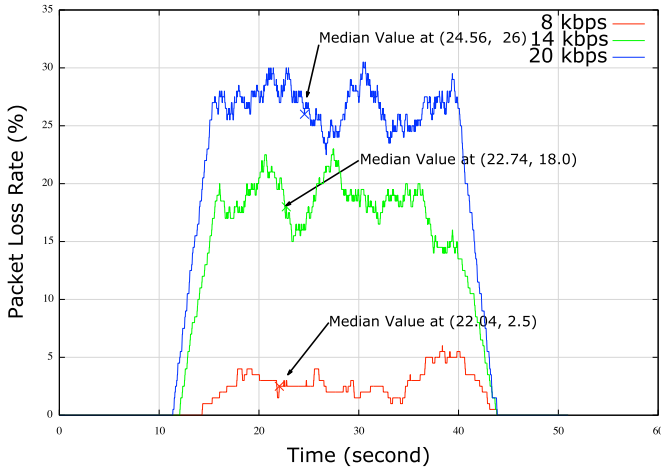


Fig. 5. Opus Stream Packet Loss under 0.994 Mbps Background Traffic

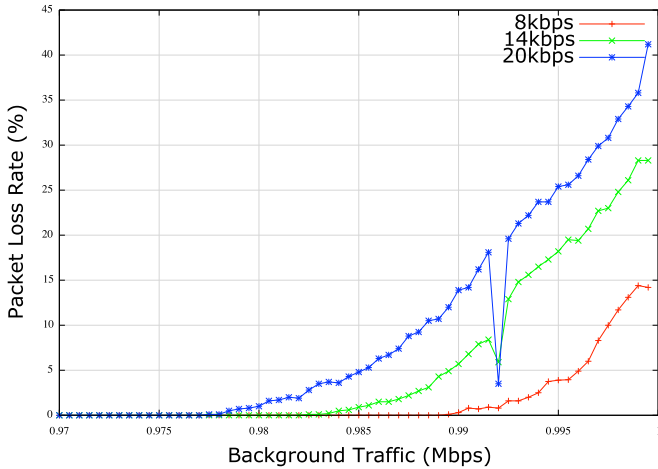


Fig. 6. Opus Packet Loss Trend under Increasing Background Traffic

20kbps which has approximately 3% packet loss, while for the 14 kbps the observed packet loss is around 6%, much below observed packet loss rates for similar levels of background traffic. However, for all other background traffic levels, the combination of packet size and background traffic level results in an amount of packets with a fractional part, which then produce coherent results. It would appear that this behavior is linked to some factor within the *ns-2* scheduling. To further validate our experimental approach, the background traffic packet size was incremented and decremented by 1 byte, which again produced correct results. For the purposes of our next experiment, the corresponding points at 0.992 Mbps were removed.

Most importantly, it can be observed in Figure 6 that for the same level of background traffic, streams with different bit-rates suffer different packet loss rates. This leads to thresholds where the call quality for a lower bit-rate stream would be better than that of a higher bit-rate stream. From Figure 6, a mapping from the packet loss rate of the current bit-rate to the corresponding packet loss rate of another bit-rate can be derived. These packet loss rates can then be compared in

terms of their corresponding speech quality (from Figure 3). This allows for the determination of thresholds where the call quality of one bit-rate exceeds that of another. In the figures below, we term 20 kbps, 14 kbps and 8 kbps as high, medium and low quality respectively. We summarize the methodology for threshold determination as follows:

- 1) Monitor the packet loss rate for the current bit-rate. And using Figure 6, compute the corresponding packet loss rate that would be observed by using a lower bit-rate.
- 2) Using the fitting provided in Figure 3, compute the MOS for the current and the lower bit-rate.
- 3) If the MOS for the lower bit-rate is higher than that for the current bit-rate, inform the encoder to switch to the lower bit-rate.

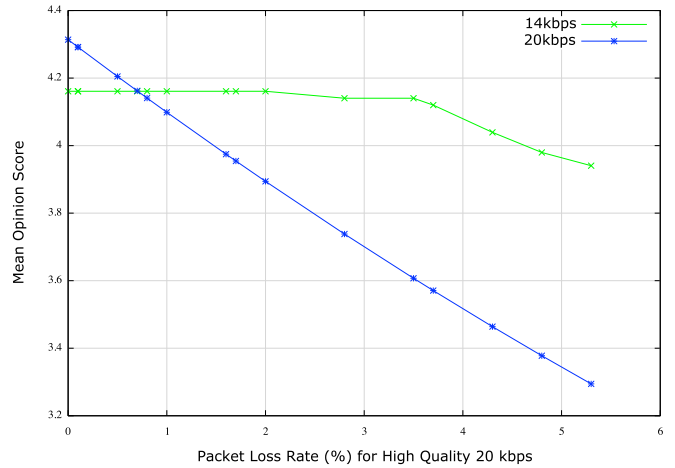


Fig. 7. Opus Stream Switching from High Quality to Medium Quality

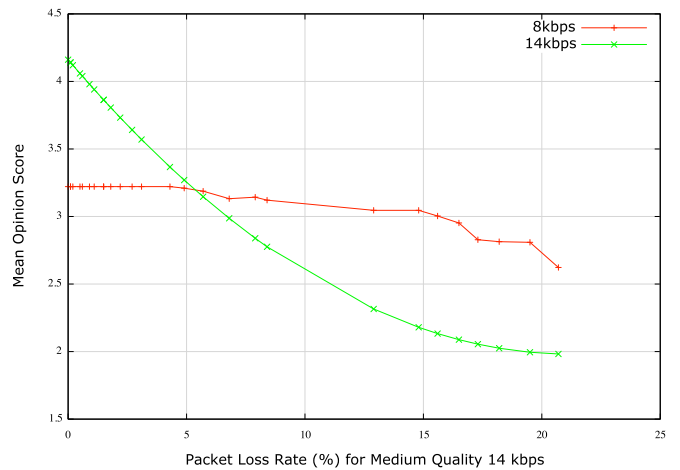


Fig. 8. Opus Stream Switching from Medium Quality to Low Quality

Figure 7 and 8 present the results for the above method, using the results obtained in our experiments. In both figures, the horizontal axes correspond to the packet loss rates experienced by 20 kbps and 14 kbps, respectively (i.e. these are the streams that the switching is made **from**). The corresponding packet loss rate for the lower quality stream being switched **to** (not

shown here) will be lower, as explained above. In Figure 7, we can observe that as the packet loss rate of the higher quality stream exceeds approximately 0.8%, the quality of the medium quality stream becomes greater than that of the high quality stream. This trend continues as packet loss rates increase. For the case of a transition between medium and low quality (Figure 8), once the packet loss rate of the medium quality stream exceeds approximately 5.7%, the quality of the lower quality stream becomes greater. A similar methodology can be used to enable switching back to a higher quality. However, this requires additional information regarding network congestion (as a 0% packet loss rate in lower bit-rates can correspond to multiple packet loss rates in higher bit-rate) and temporal logic in order to avoid repeated switching between lower and higher bitrates over a short period of time.

## V. CONCLUSION

This paper presented an analysis of the impact of packet loss on VoIP call quality for the Opus audio codec. From this analysis, a collection of fitting functions describing the relationship between packet loss rate and call quality for multiple bit-rates, were derived. Furthermore, this paper also investigated the performance of Opus in a congested network scenario, to understand the relationship between packet loss rates for different bit-rates (and the resultant call quality). The results of these experiments allowed us to determine that beyond approximately 0.8% packet loss rate for a 20 kbps stream, better quality can be achieved by switching to a 14 kbps stream. Similarly, when experiencing above approximately 5.7% packet loss for a 14 kbps stream, switching to a 8 kbps stream results in better quality.

To increase the applicability of these results, future work could include the analysis of the transcoded samples and samples from different languages. Focusing on Opus, future work could include the investigation of more bit-rates, frame sizes and the use of more complex packet loss concealment and forward error correction, as these were not used. While on the simulation environment, the use of different background traffic models could produce interesting results.

## ACKNOWLEDGMENT

This research was partially funded by Science Foundation Ireland via grant 08/SRC/I1403 FAME SRC. This work was also supported by the Irish Research Council for Science, Engineering and Technology, co-funded by Marie Curie Actions under FP7; and by the Telecommunications Graduate Initiative which is funded by the Higher Education Authority under the Programme for Research in Third-Level Institutions Cycle 5 and co-funded under the European Regional Development Fund. This work was also supported, in part, by Science Foundation Ireland grant 10/CE/I1855.

## REFERENCES

- [1] ITU, *Recommendation ITU-T G.722: 7 kHz audio-coding within 64 kbit/s - technical specification*. ITU-Telecommunication Standardization Sector, 2012. [Online]. Available: <http://www.itu.int/rec/T-REC-G.722-201209-1/en>
- [2] B. Bessette, R. Salami, R. Lefebvre, M. Jelinek, J. Rotola-Pukkila, J. Vainio, H. Mikkola, and K. Jarvinen, "The adaptive multirate wideband speech codec (amr-wb)," *Speech and Audio Processing, IEEE Transactions on*, vol. 10, no. 8, pp. 620–636, 2002.

- [3] ITU, *Recommendation ITU-T G.722.1: Low-complexity coding at 24 and 32 kbit/s for hands-free operation in systems with low frame loss*. ITU-Telecommunication Standardization Sector, 2005. [Online]. Available: <http://www.itu.int/rec/T-REC-G.722.1-200505-1/en>
- [4] J. Valin, K. Vos, T. Terriberry, and A. Moizard, "Definition of the Opus Audio Codec," <http://www.ietf.org/rfc/rfc6716.txt>, 2012.
- [5] K. Vos, S. Jensen, and K. Soerensen, "SILK Speech Codec," <http://tools.ietf.org/search/draft-vos-silk-01>, 2010.
- [6] J.-M. Valin, T. Terriberry, C. Montgomery, and G. Maxwell, "A high-quality speech and audio codec with less than 10-ms delay," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 18, no. 1, pp. 58–67, 2010.
- [7] J.-M. Valin, T. B. Terriberry, and G. Maxwell, "A full-bandwidth audio codec with low complexity and very low delay," in *17th EURASIP European Signal Processing Conference*, 2009, pp. 1254–1258.
- [8] opus codec.org, "Opus interactive audio codec," 2012. [Online]. Available: <http://www.opus-codec.org/>
- [9] ITU, *Recommendation ITU-T P. 862.2: Wideband Extension to Recommendation P. 862 for the Assessment of Wideband Telephone Networks and Speech Codecs*. ITU-Telecommunication Standardization Sector, 2007. [Online]. Available: <http://www.itu.int/rec/T-REC-P.862.2-200711-1/en>
- [10] H. Assem, M. Adel, B. Jennings, D. Malone, J. Dunne, and P. O'Sullivan, "A Generic Algorithm for Mid-call Audio Codec Switching," in *IFIP/IEEE International Workshop on Quality of Experience Centric Management*, 2013, p. 12761281.
- [11] A. Rämö and H. Toukoma, "Voice Quality Characterization of IETF Opus Codec," in *12th Annual Conference of the International Speech Communication Association (INTERSPEECH)*, 2011, pp. 2541–2544.
- [12] P. Orosz, T. Skopkó, Z. Nagy, and T. Lukovics, "Performance Analysis of the Opus Codec in VoIP Environment Using QoE Evaluation," in *ICSNC 2013, The Eighth International Conference on Systems and Networks Communications*, 2013, pp. 89–93.
- [13] ITU, *Recommendation ITU-T P.862 : Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs*. ITU-Telecommunication Standardization Sector, 2001. [Online]. Available: <http://www.itu.int/rec/T-REC-P.862-200102-1/en>
- [14] A. Takahashi, A. Kurashima, and H. Yoshino, "Objective Assessment Methodology for Estimating Conversational Quality in VoIP," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 14, no. 6, pp. 1984–1993, Nov 2006.
- [15] L. S. G. a. D. Carvalho and E. D. S. Mota, "Survey on application-layer mechanisms for speech quality adaptation in VoIP," vol. 45, no. 3, Jul. 2013, pp. 36:1–36:31. [Online]. Available: <http://doi.acm.org/10.1145/2480741.2480753>
- [16] S. Moller, M. Waltermann, B. Lewcio, N. Kirschnick, and P. Vidales, "Speech Quality While Roaming in Next Generation Networks," in *IEEE International Conference on Communications*, 2009, pp. 1–5.
- [17] T. Chakraborty, A. Mukhopadhyay, I. Misra, and S. Sanyal, "Active queue management with adaptive codec bit rate variation in WLAN for efficient VoIP performance," in *Computer and Information Technology (ICCIT), 2011 14th International Conference on*, 2011, pp. 179–184.
- [18] ITU, *Recommendation ITU-T P.834: Methodology for the derivation of equipment impairment factors from instrumental models*. ITU-Telecommunication Standardization Sector, 2002. [Online]. Available: <http://www.itu.int/rec/T-REC-P.834-200207-1/en>
- [19] H. Balakrishnan and R. H. Katz, "Explicit loss notification and wireless web performance," in *Proc. IEEE Globecom*, vol. 98, 1998.
- [20] N. Côté, V. Gautier-Turbin, A. Raake, and S. Möller, "Analysis of a quality prediction model for wideband speech quality, the WB-PESQ," in *Proceedings of the 2nd ISCA/DEGA Tutorial and Research Workshop on Perceptual Quality of Systems*, 2006, pp. 115–122.
- [21] S. Möller, N. Côté, V. Gautier-Turbin, N. Kitawaki, and A. Takahashi, "Instrumental estimation of E-model parameters for wideband speech codecs," vol. 2010. New York, NY, United States: Hindawi Publishing Corp., Jan. 2010, pp. 9:1–9:16. [Online]. Available: <http://dx.doi.org/10.1155/2010/782731>
- [22] A. D. Clark, "Modeling the effects of burst packet loss and redundancy on subjective voice quality," in *Internet Telephony Workshop*, 2001.