Quality Analysis of Streaming Audio over Mobile Networks

Sikha Bagui[1], Amitabh Mishra[1], John Compo[1], Debarghya Nandi[1] and Subhash Bagui[2]
[1]Department of Computer Science,
[2]Department of Mathematics and Statistics, The University of West Florida, Pensacola, FL 32514
bagui@uwf.edu; amishra@uwf.edu; johncompo@gmail.com, dnandi4@uic.edu

ABSTRACT

This paper utilizes open source software to analyze the quality of audio streamed over mobile cellular networks. Industry conventions have been developed to assess audio quality such as the Mean Opinion Score or MOS scale described in the International Telecommunications Union ITU P.800.1 document. The MOS scale is a subjective assessment based on the listener's experience. To eliminate the use of a trained audio listener we automate an estimated MOS calculation by measuring the packet loss, average latency and jitter over the network transport path. The network under test is a pilot network to replace the dedicated analog circuit from the broadcast center to a radio transmitter. We intend to automate the logging of an objective quality assessment using MOS, cellular router and decoder measurements with confirmation using automatically generated visual representations of sample audio received. These visual representations will aid in manual confirmation of poor MOS scores with audio samples available for more in-depth review.

Keywords: Audio Quality, Mean Opinion Score, Streaming Audio, Mobile Networks, Outages

1 Introduction

The production broadcast system text database stores safety and public information messages derived from weather forecast products. The broadcast center creates a broadcast script from these messages and sets up a play schedule. Text to speech software is used to convert these messages into a synthetic, digital voice message. The voice message is converted to an analog signal by a digital to analog audio converter that supplies radio transmitters via a dedicated analog circuit from the broadcast center. In addition to the broadcast voice, various frequency tones can be inserted into the message. Specific Area Message Encoding (SAME) tones are used to send an alert to receivers in a specific geographical location such as a county. Additional tones are used to remotely switch the transmitter from the primary to the secondary transmitter. The dedicated circuit is analogous to an audio cable from a personal computer to a set of speakers. The audio remains an analog signal from the broadcast center to the transmitter. Figure 1 presents a legacy broadcast center to transmitter configuration. In figure 1 we see the text to speech engine pulling text products from the text product database. Once the text is converted to an audio signal, it is transmitting over the analog circuit to the radio transmitter for broadcast.
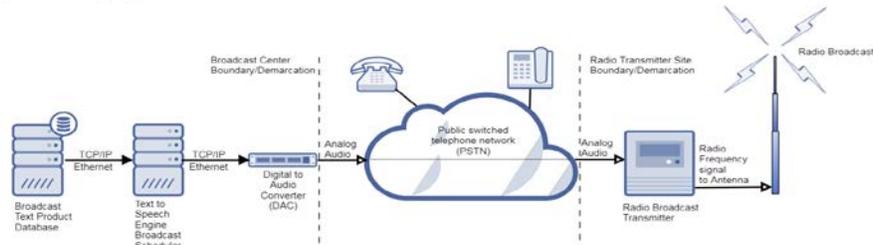
Figure 1. Legacy broadcast center to transmitter configuration - dedicated circuit (Analog).

Telephone communications continue to evolve from analog copper-wire circuits carrying analog signals between fixed geographical points to newer digital circuits using signals that can reach multiple geographical areas and locations through the Internet or private networks. This evolution is continuing, and analog circuits are becoming rarer and more expensive than digital circuits. In some markets these legacy analog circuits are no longer available.

The pilot system being tested uses cellular network technology commonly used with smart phones and smart devices. Within this cellular network, a Virtual Private Network (VPN) is created between the broadcast center and the radio transmitter to isolate the broadcast stream from the public internet. In Figure 2 the setup is like the previous analog circuit with the exception of the transport method between the broadcast center and transmitter boundaries. The analog audio is now encoded into a digital signal by an audio encoder once the signal is digital, the encoding router sends the Real Time Protocol (RTP) packets via 4G networks to the decoding router. Once received, the audio decoder converts the RTP packets back to an audio signal to supply the radio transmitter.
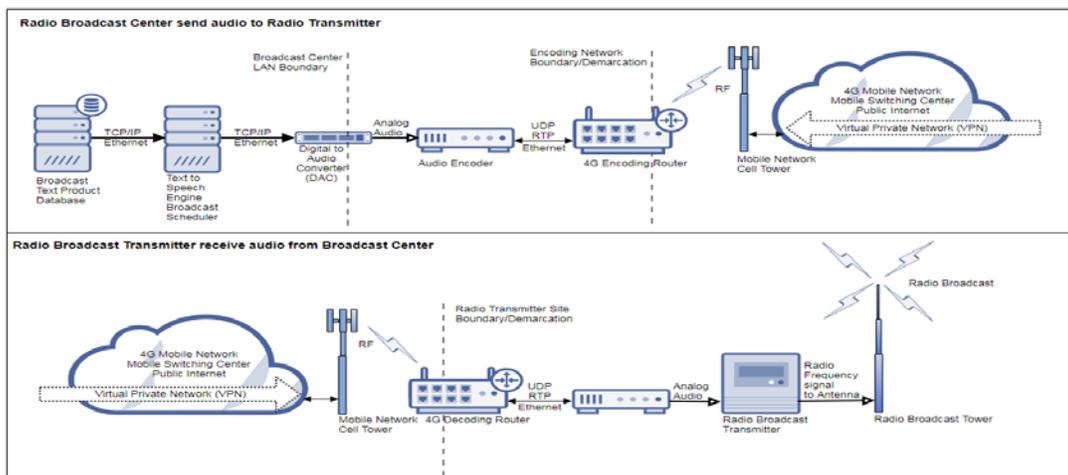

Figure 2. Pilot broadcast center to transmitter configuration - 4G mobile network with VPN (Digital).

The broadcast is a one-way audio stream and is transmitted over the VPN using the Real Time Protocol (RTP) G.711 codec as expressed in the document ITU [4] G.711. One distinct challenge with dedicated analog circuits and RTP is automating the objective assessment of audio quality. Transmitters that are close to the broadcast center can be monitored on a weather radio. Transmitters further away cannot be monitored for audio quality without dialing in to the transmitter and listening to a short sample. The broadcast center relies upon members of the public to report audio problems or outages. Their audio assessment may be subjective based on their experience, receiver setup, terrain blockage and the distance from the transmitter. Any prolonged interruption to streaming audio is noticed immediately by

the intended listener as silence or distortion. Silence could be caused by many factors, most likely an audio or network outage along the transmission path. Poor audio or strange noises such as echoes, or warbling tones would most likely be attributed to network jitter. In this paper we propose a setup using open source software to both automate the calculation of an estimated MOS along with detailed analysis about the audio quality. Our setup also provides automated visual references and sample audio for manual confirmation of poor MOS scores.

The information from the broadcast center to the radio transmitter includes current weather forecasts, watches, warnings, public information statements and emergency messages. Due to the critical nature of such information, a high level of service availability is required. These broadcasts don't have the dynamic range or fidelity requirements of an FM radio station; however, these transmissions should be clear, with limited noise and minimal outages. The transmitters broadcast on narrow band Frequency Modulation (FM). The designed area of coverage is 40 miles from the transmitter. This range should allow a received signal more than -89dBm which is more than adequate for the voice broadcast and more importantly high enough to trigger the receiver when SAME tones are sent.

In this work, we intend to automate the initial analysis of received audio. By using network statistics and software and tools at the radio transmitter received audio. We attempt to automate quality analysis and capture outages. The automated methods should correlate to waveform and spectrograph analysis to assure they are accurate. These visual representations will aid in manual confirmation of poor MOS scores with audio samples available for more in-depth review.

The rest of the paper is organized as follows. Section 2 presents the related works; section 3 presents the pilot setup; section 4 presents the methods used to assess audio quality; section 5 presents the data analysis; section 6 presents the predictions that can be made and finally section 7 presents the conclusions and findings.

2 Related works

[2] used structural similarity index to compare transmitted and received frequency structures. From this they derived a correlation coefficient. Their work was mainly with musical instruments sources and the dynamic range of their signals is much greater than our vocal and tone requirements.

To measure the quality of video streaming, [7] used correlation as well. While the title of their article was video streaming, portions of their setup are applicable to audio streaming as a component of that video. Their work is impacted by round trip time (RTT) and other elements prone to limitations of the TCP/IP protocol suite. One of the major impacts to their results was due to the TCP/IP retransmission or sliding windows flow where the transmit side delays the transmission to assure that the packets are in order. This would cause a serious effect on the quality of streaming audio. In our setup, the encoders and decoders communicate with each other using User Datagram Protocol over IP (UDP) which does not involve the complex handshaking used in TCP. In our application we continuously send audio regardless of packet losses, and there is no re-transmission.

The approach [10] used jitter distortion as a method for assessing audio quality. The focus of their paper was the delivery via cell phone, but they used a listener to provide a subjective assessment of the quality. Moreover, their application focused on sporadic burst transmission. On the contrary, our broadcast stream is continuous and employing a listener over extended periods of time is not feasible.

Authors for [6] specifically identified issues related to audio transmission over "lossy" networks such as the Internet. Their work focused on the comparison of quality to packet loss rates of the decoded signal. They encoded and decoded the audio with different rates to evaluate the quality. However, our system is a live system and our goal is to assess and improve the audio.

3   Setup

The pilot test setup consists of two live broadcast links and a test broadcast link. The live links can revert to their coexisting dedicated analog paths during intrusive testing. Intrusive testing involves transmitting one of the two known audio files over the link. Without switching the site back to the legacy dedicated circuit, the radio transmitter broadcasts the test message. Since tones are included in the test messages, there is an additional risk of causing false alerts on the receivers.

The first pilot 4G link is from a broadcast center in Texas to a simulated transmitter in Vermont, this link is approximately 1,800 miles. This simulated transmitter is the logging PC which also records the audio. The second pilot link is a broadcast from the Vermont broadcast center to a transmitter in Vermont approximately 16 miles away. A third local test setup 4G pilot link for comparison uses collocated routers. This local setup has the two routers separated by a short distance of 25 feet. All three links use mobile network 4G LTE as the primary network technology. The links are named T1800V, V16V and V0V to denote the origin, distance and destination. All three links use commercially available audio encoders and audio decoders in pairs. In addition to the encoders and decoders, single board computers (SBCs) are used at the origin and destination. The originating SBC is used to inject a known audio signal into the encoder during intrusive testing. This audio is either a 30 second loop of a 440 hertz concert A (A4) sine wave tone or a known one-minute loop of synthetic voice reciting the Harvard Sentences [3] List # 11 with some 400Hz and 1 kHz tones. Examples of the transmitted waveform and spectrographs viewed in Audacity [1] are presented in figures 3 to 5.

Figure 3. Waveform 440 Hz

Figure 4. Spectrograph 440 Hz

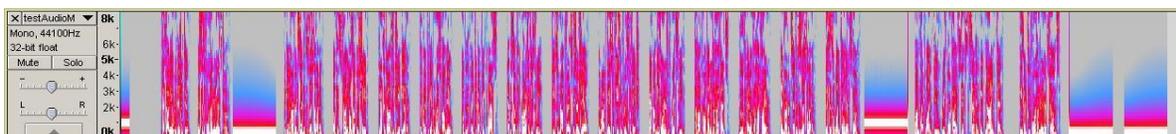**Figure 5. Harvard Sentences List 11 and tones.**

Figure 6. Spectrograph - Harvard Sentences List 11 and tones.

In Figure 6 we see a spectrograph representation of the test audio file. The test audio file was constructed using the Harvard Sentences [3] #11 and a 440Hz at specific locations as marker. This file is of a fixed duration and content will aid in the quality analysis of the received signal. The voice and tone loop or constant tone provide a reference waveform to compare against the received audio. These known audio signals help in detailed manual analysis of the received audio if necessary. The live broadcast varies greatly and is difficult to discern the quality visually. At the receive end an SBC is used to record the audio from the decoder as well as perform the network tests to estimate a MOS Score.

The system setup we designed produces this estimated MOS score based on the network packet loss, average latency and jitter from the decoder to the encoder. A poor estimated MOS provides a timestamp to check in cases where further analysis is required. Our system also logs the encoding and decoding router and decoder data. It records 15 second samples of audio 2-3 times per minute. From this audio, our system automatically logs the audio sample's statistics and automatically generates the waveform, FFT and spectrograph plots using the Matlibplot [5] library in Python [8].

4   Methods

To assess the audio quality one can simply listen to the audio stream and determine if it can be clearly heard and with ease. To establish a scale of that quality, the Mean Opinion Score (MOS) was developed by the International Telecommunication Union (ITU) [4]. The MOS scale is subjective; the listener should be a trained audio professional. Employing a human listener for a continuous MOS scale assessment is not always practical. Table 1 presents the scales used by the ITU [4] and G.711. Table 1 shows the MOS scale ranging from a listener's estimated experience ranging from bad to excellent.

Table 1: MOS scale published by International Telecommunication Union (ITU) for G.711 CODEC.

MOS	ITU-T P.800.2
5.0	Excellent
4.0	Good
3.0	Fair
2.0	Poor
1.0	Bad

Source: ITU-T P.800.2 07/2016

The codec used for our testing is the G.711 Real Time Protocol (RTP) codec defined in ITU-T [4] Recommendation [2, 3] G.711. It is a commonly used codec for Audio over IP (AoIP) applications, available for free. The audio encoding for the encoder and decoder uses μ-law 8-bit Pulse Code Modulation (PCM). One of the benefits of the μ-law is the dynamic range available and reduced bandwidth. It should be noted that a maximum MOS score for the G.711 codec is expressed as 4.4 in the MOS scale table. This maximum theoretical MOS score varies from 4.1 - 4.5 depending on the reference source. Any value above 4.0 is acceptable as the signal is in the good range or satisfied level of quality depending on the chart selection. A subjective good or satisfied MOS threshold can be identified by listening to the received audio and cross referencing against the estimated MOS to establish an acceptable baseline for audio quality if necessary. MOS subsets are defined as MOS-LQE, MOS-LQO and MOS-LQS, where L=listening, Q=quality and E=estimated. O=Objective and S=Subjective were not used.

To test the three AoIP links, we have used a passive approach for the live broadcasts including the estimated MOS, detailed logging as previously mentioned and audio recording. For the intrusive test we inject one of our known audio files into the broadcast stream. For the pure tone injection, we additionally have an automated audio metric log which logs amplitude, norms and deltas of that audio signal.

The broadcast center to radio transmitter sites and distances are identified in table 2.

Table 2: Site to site locations and payload type.

#	Transmit	Receive	Distance (mi)	Type	Payload	MOS Type
1	Texas	Vermont	1,800	Encoder to Decoder	Audio and/or Tones	MOS-LQS/LQO
2	Vermont	Vermont	16	Encoder to Decoder	Audio and/or Tones	MOS-LQS/LQO
3	Vermont	Vermont	0	Encoder to Decoder	Audio and/or Tones	MOS-LQS/LQO

There are many factors that contribute to the audio quality of the broadcast. The audio encoding quality was assumed to be good and constant since we injected a known reference audio signal into all three encoders. We compared this audio file directly against the received audio samples.

There are several possible causes for audio degradation during transmission. The encoder, encoding router (transmit), decoding router (receive), decoder could be suspect. The routers could suffer from poor signal strength or signal to noise ratio. If the routers do not have a strong signal, the attachment to the mobile network will be prone to errors. The audio encoder and decoder pair could have problem exasperated by the network connection/conditions between them. These could cause problems with RTP latency, drops, duplicates and errors. In addition, the audio encoder decoder pair has to sync with each other, this takes approximately 2-3 seconds. Any network disconnects are amplified by this additional time to come back online and sync.

Network transport is another common cause of audio degradation. The situation with network transport is analogous to two tin cans and a string where the string is slack or over-tight or is struck with an object during transmission. Two of the important measurements that should be logged for judging the mobile network connection quality are Received Signal Strength Indicator (RSSI) and the Signal to Interference Noise Ratio (SINR). In our setup, whenever SINR on the router was not available for logging, we logged the Received Signal Received Quality (RSRQ) values.

There are additional parameters related to the decoders that we logged - the most informative of these are the Real Time Protocol attributes for latency, loss, drops duplicates. The decoder buffer, soft errors and reconnects are also logged.

In addition to recording the above data, we used a raw socket Python [8] ping to determine decoder to encoder packet loss, average latency and jitter to calculate an estimated MOS for the transmissions. We also captured 15 second audio samples approximately two to three times per minute. From these samples we plotted the first few cycles of the 440 Hz tone to analyze the sine wave. We also plotted the spectrum of the broadcast a constant tone can be analyzed quickly with a Fast Fourier Transform (FFT). Any frequency deviation of the constant tone can be easily identified. In the image below, the transmitted or reference audio signal is plotted together with the received audio signal. It may be noted that the waveform plot and FFT are plotted for a small fraction of the 15 second sample in figure 7.
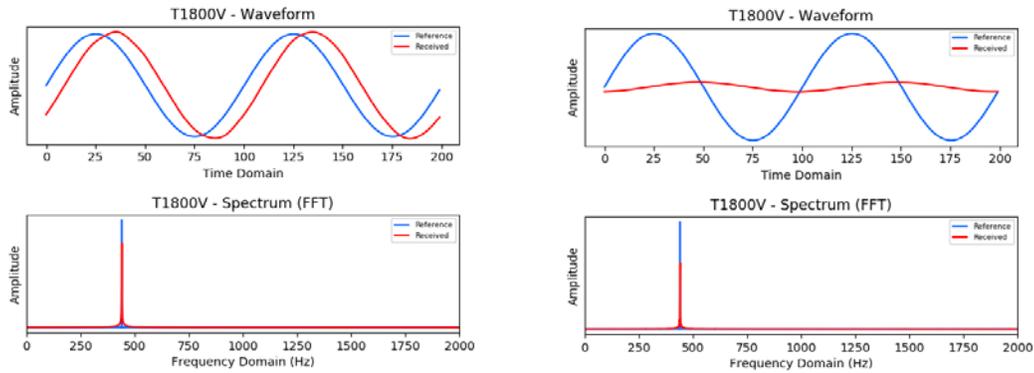
Figure 7: Waveform and FFT plot.

In the above waveform diagram, we analyze the sine wave characteristics of a sample; the phase difference varies based on the transmission delay. The FFT plot shows if the reference and received signals are on same frequency, or if there is a drift. The amplitudes of the transmitted and the received signals are plotted to check just for deviation and may not necessarily be relative to each other. Volume adjustments on the encoder, decoder and SBC can also be the reasons for changes in the amplitude. In the two waveform samples, it is evident that there was signal loss or fade. The entire 15-second sample can be opened in Audacity [1] for detailed information on the complete waveform or to listen to the sample.

The complete 15 second sample is plotted below with a spectrograph. The outages and deviations from the primary frequency are more easily discerned in the spectrograph output as shown in Figure 8.
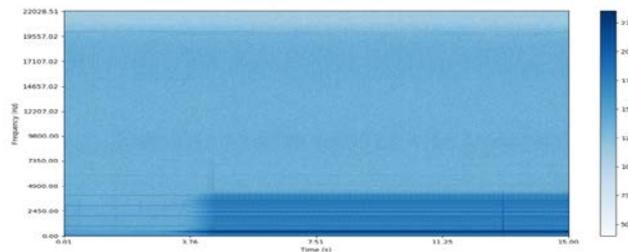

Figure 8: Spectrograph of constant frequency tone.

From figure 8, we can see that, in the first quarter there appears to be no signal in the spectrograph output. The dark blue, highest amplitude component in the plot is the received 440 Hz signal. The signal returns to 440 Hz just before the 3.76 second mark. Once again, we can view the entire 15 second waveform with a tool like Audacity and listen to the audio.

The estimated MOS score and the decoder RTP attributes are the main indices to research further into the spectrographs and recorded audio samples.

5   Data Analysis

Our setup and the experimental process to find bad or missing audio were used at all three sites T1800V, V16V and V0V. The MOS, encoding router, decoding router and decoder log values were plotted to look for trends. Of interest were the MOS scores, routers' signal levels and the decoder buffer, and RTP latency, drops, loss and duplicates.

In all three sites, the encoding and decoding router had a strong received signal and good signal to noise values. The 15 second audio sample log was also plotted to look for trends. From the experiment results, we found that of importance in this log are the root mean squared amplitude and the amplitude mean norm values. The spectrograph and waveform plots were compared against low MOS scores and low values in the root mean squared amplitude and the amplitude mean norm.

The MOS score below is of a live broadcast voice from the Texas site to Vermont T1800V. The MOS score is estimated continuously for all the broadcasts. The MOS score below is 1.93 which is in the 'Poor' category on the ITU [4] MOS scale. In Figure 9 we can see the variance ranging from Poor to Good MOS scores.
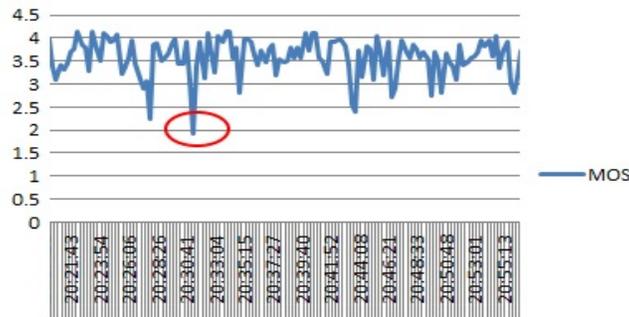

Figure 9: Automated MOS score example Texas to Vermont T1800V link.

Using this low MOS score as the date time reference we selected the correct spectrograph for this audio sample. It may be noted that the darker the shade of blue the higher the amplitude of the signal. Figure 10 is small sample spectrograph of the audio associated with the Poor MOS score in Figure 9.
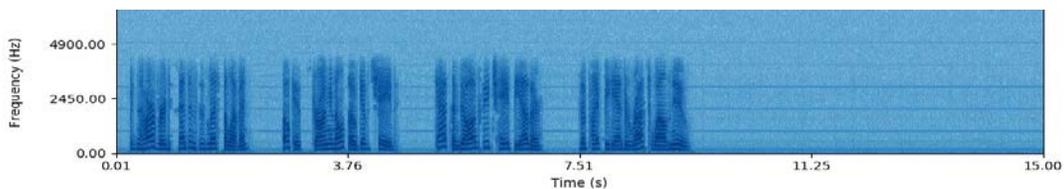

Figure 10: Spectrograph of 15 second sample correlating to MOS score in Figure 9.

The complete 15 second sample of the audio was viewed and listened to in Audacity [1] (figure 10). When we loaded the waveform, the wave signature matched the spectrograph. When we listened to the sample audio, it was confirmed that the audio was missing.

Figure 11: Suspect audio waveform, same sample as figure 10.

The sample below is the voice tone loop injected into the encoder at the Vermont to Vermont site V16V. This site MOS scores ranged from 3.2 "Fair" to 4.29 "Good" on the ITU [4] MOS scale for the total sample

period. The MOS score for this sample was 3.5 which is between "Fair" and "Good" on the ITU [4] MOS scale. We can infer from the plot below that the MOS score was dropping prior to this low value. The score in question is circled in red on figure 12.
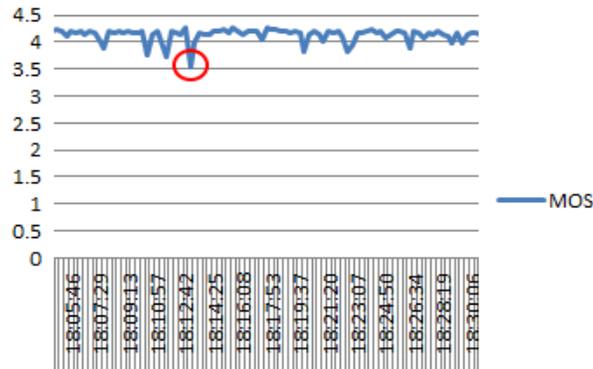

Figure12: Automated MOS score example Vermont to Vermont V16V link.

When we referenced the spectrograph using the time and date from the log we found that the signal was missing and then faded back in. One of the embedded test tones can be seen to the right of the spectrograph represented by the dark horizontal line.
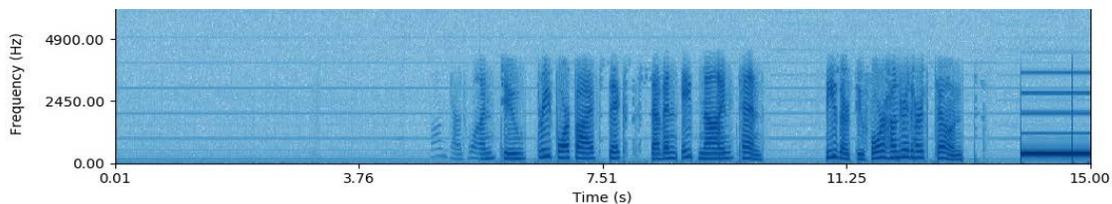

Figure 13: Spectrograph of 15 second sample correlating to MOS score in Figure 12.

From the spectrograph above in Figure 13 it appeared the signal was missing and then faded in. We confirmed this with the aid of waveform representation of the audio in Figure 14 and by listening to the sample.
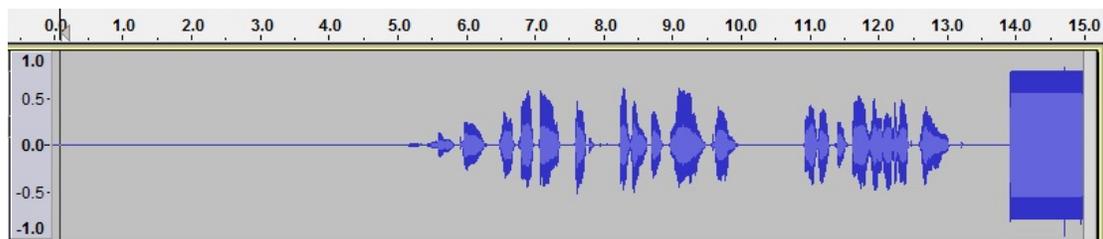

Figure 14: Suspect audio waveform, same sample as figure 13.

The previous samples were from a live broadcast or the test voice and tone loop. Another technique we used to test for audio quality was to inject a constant 440 Hz tone loop into the encoder. The following are examples of audio problems discovered using the MOS method referenced above. In addition, since we are using a constant tone, we can now use audio statistics of the recorded 15 second audio as a reference as well. The reference audio file was tested with the same SoX [9] utility used to generate the logged values as a baseline. The statistics for the test 440 Hz file are below. We expected the log audio

statistics to be different based on audio volume adjustments throughout the audio link; however, the relationships should be similar. In Table 3, we can see the statistics of the 440Hz tone.

Table 3: Sound Exchange Utility (SoX) audio statistics.

Maximum amplitude:	0.692017
Minimum amplitude:	-0.691986
Midline amplitude:	0.000015
Mean norm:	0.44042
Mean amplitude:	0
RMS amplitude:	0.489183
Maximum delta:	0.043854
Minimum delta:	0
Mean delta:	0.027605
RMS delta:	0.030662

Shown below in Figure 15 is a plot of the MOS scores from the T1800V site with an injected tone. Of interest is the negative spike well below 2.5 circled in red (as shown in figure 15).

Figure 15: Automated MOS score example Texas to Vermont T1800V link, constant 440Hz tone.

Figure 16 is an example of how the audio statistics can now be used to indicate an outage. The negative spikes seen in the plot below are of special interest. We had high success in correlating these negative spikes to audio problems. For example, the spike circled in red matches the date time of the MOS score negative spike above.
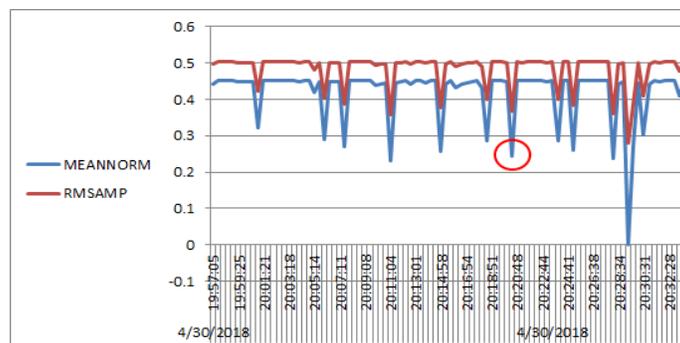

Figure 16: SoX audio statistics correlate with MOS score to indicate an outage.

The outage is clearly seen in the spectrograph below (figure 17) as well as in the waveform image that follows (figure 18).
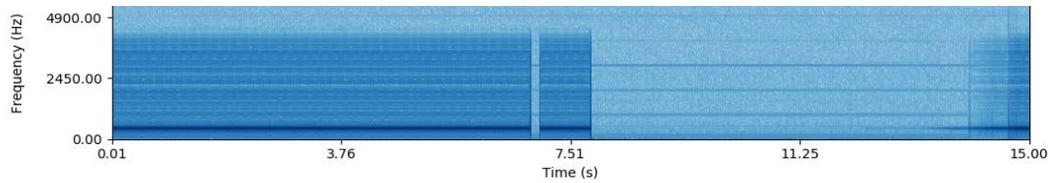
Figure 17: Spectrograph of 15 second sample correlating to MOS score in Figure 15 and SoX audio statistics in Figure 16.
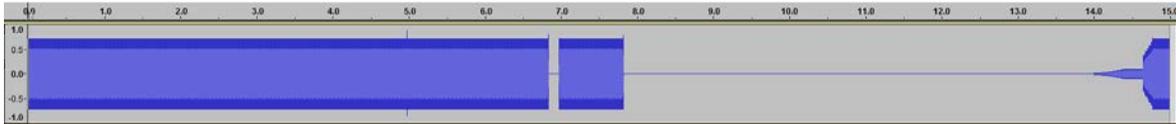

Figure 18: Waveform of 15 second sample correlating to MOS score in Figure 15 and SoX audio statistics in Figure 16.

In the next example, we bypass the MOS score as a reference and utilize the audio statistics alone to find bad audio. Note how defined the audio statistics plot is, each negative spike was correlated to an audio problem. While the MOS score is a useful parameter in rating audio quality, it doesn't always catch all the problems due to the decoder's' buffer. This buffer helps smoothen out network variances in latency and jitter, but the smoothing is not always successful as we observed that bad audio correlated to a low MOS score as well. The buffers effectiveness degrades with lower MOS scores. The audio statistic plot below shows an audio outage which is further represented by the spectrograph, waveform and confirmed by listening to the sample.
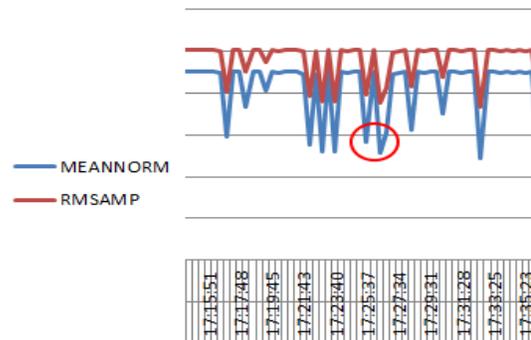

Figure 19: SoX audio statistics as a stand-alone indicator of audio quality problems (constant tone).
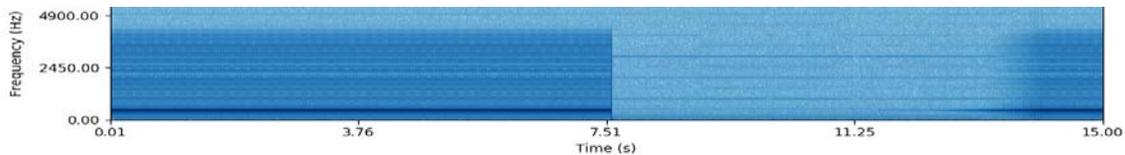

Figure 20: Spectrograph of 15 second audio sample from Figure 19.

Figure 21: Waveform of 15 second audio sample from Figure 19

6   Data Prediction

Using the data collected, we were able to rate audio quality. The MOS score remains the best indicator in our configuration, and we could safely rely upon it to forecast future performance. The audio quality statistics worked well for a constant tone but not for the variable frequency voice of the live broadcast of the voice and tone test loop. We counted the number of MOS scores over a 24-hour period to establish how many were "Good", "Fair", "Poor", "Bad" or "Unusable". The data was compared using the ITU-T rating scale [4] for MOS, and tables 4 to 6 present the results.

Table 4: T1800V

Total Events	5279	100%	
4.0-4.5	2935	56%	Good
3.0-3.9	2215	42%	Fair
2.0-2.9	128	2%	Poor
1.0-1.9	1	0%	Bad
<1.0	0	0%	Unusable

Table 5: V16V

Total Events	5012	100%	
4.0-4.5	4664	93%	Good
3.0-3.9	342	7%	Fair
2.0-2.9	0	0%	Poor
1.0-1.9	0	0%	Bad
<1.0	0	0%	Unusable

Table 6: V0V

Total Events	5402	100%	
4.0-4.5	5238	97%	Good
3.0-3.9	149	3%	Fair
2.0-2.9	0	0%	Poor
1.0-1.9	0	0%	Bad
<1.0	0	0%	Unusable

The T1800V audio link was found to be the link with the poorest quality. Only 56% of the time did it provide "Good" audio. Nearly half the time, 44% it provided "Fair" to "Poor" audio. The V16V and V0V audio links met our goal of "Good" audio 93% and 97% respectively. As the other data received and analyzed by us showed similar, corroborating characteristics, there is no reason to assume the T1800V audio link will improve or the V16V and V0V will degrade below their current levels.

7   Conclusion and Findings

The MOS scores provided the expected date time indicator with variances to research further into the audio outages. The network latency and jitter were negated to some degree by the buffer in the decoder. The buffer helped reduce the impact of a poor MOS score on audio quality to a degree. The T1800V audio link was "Fair" for nearly half the time. As a listener, we found it unacceptable and difficult to listen to. The link had frequent drops of varying durations evident in the spectrographs. This was true for the live

broadcast audio as well as the voice and tone and constant tone loop files. The V16V and V0V links were "Good" most of the time confirmed by the spectrographs and listening to samples.

In our tests it was easiest to locate an audio problem using the audio statistics log for the 440 Hz tone signal. With the constant signal and constant frequency, we were able to use the audio statistics as a further indicator of where to look at the audio. The root mean squared amplitude and mean norm amplitude provided excellent results most of the time. The audio statistics logging was only useful for the tone loop, it didn't perform well with the live broadcast or voice and tone loop. The MOS scores however applied to all audio sources.

To improve on our technique, the SoX [9] statistics could be set up to monitor the live broadcast audio. This may help to provide a baseline for the live broadcast or test voice and tone audio. We could compare this baseline to the live broadcast audio stream. If successful, this would provide a high frequency of audio problem detection, like the results we achieved with the constant tone loop. The logging for the MOS, encoding router, decoding router and decoder could be combined with the audio statistics so that the date times are more in sync. These methods could be refined and applied to live broadcast audio to further automate the process for continuous monitoring of streamed audio to radio broadcast transmitters.

ACKNOWLEDGEMENT

REFERENCES

[1]   Audacity Team (2017). Audacity(R): Free Audio Editor and Recorder [Computer application]. Version 2.2.0 retrieved February 5th 2018 from https://audacityteam.org/

[2]   Gan, C., Wang, X., Zhu, M., & Yu, X. (2011). Audio quality evaluation using frequency structural similarity measure. IET International Communication Conference on Wireless Mobile and Computing (CCWMC 2011). doi:10.1049/cp.2011.0896

[3]   IEEE recommended practice for speech quality measurements (No. 297). (1969). New York, NY: Institute of Electrical and Electronics Engineers.

[4]   ITU-T. (1993). G.711 : Pulse code modulation (PCM) of voice frequencies. Retrieved from https://www.itu.int/rec/T-REC-G.711

[5]   J. D. Hunter, "Matplotlib: A 2D Graphics Environment," in Computing in Science & Engineering, vol. 9, no. 3,pp. 90-95, May-June 2007. doi: 10.1109/MCSE.2007.55

[6]   Khalifeh, A. F., Al-Tamimi, A., & Darabkh, K. A. (2017). Perceptual evaluation of audio quality under lossy networks. 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET). doi:10.1109/wispnet.2017.8299900

[7]   Mok, R. K., Chan, E. W., & Chang, R. K. (2011). Measuring the quality of experience of HTTP video streaming. 12th IFIP/IEEE International Symposium on Integrated Network Management (IM 2011) and Workshops. doi:10.1109/inm.2011.5990550

[8]   Python Software Foundation. Python Language Reference, version 2.7. Available at <http://www.python.org/>

[9]   SoX (2018). SoX - Sound eXchange: Free Audio Converter and Processor [Computer application]. Version 14.4.2 retrieved February 5th 2018 from http://sox.sourceforge.net/

[10]   Yuhong, Y., Hongjiang, Y., Ruimin, H., Song, W., & Songbo, X. (2013). A new mobile audio quality assessment using Jitter Distortion Measure approach. 2013 Fifth International Workshop on Quality of Multimedia Experience (QoMEX). doi:10.1109/qomex.2013.6603234