Wainhouse Evaluation- Enterprise Calling (Q4 2020)

Wainhouse quality evaluation of five market-leading enterprise calling solutions – methodology and top line results.

Bill Haskins, Bryan Hellard

In Q4 2020, Wainhouse evaluated market-leading enterprise calling platforms from 8x8, Microsoft, RingCentral, Cisco, and Zoom. Our primary objective was to quantify and compare audio quality between each platform, using established and accepted industry standards to remove subjective user or tester influence.

We performed this evaluation in our tightly controlled lab environment following a repeatable process adapted from our existing methodologies and documented guidance for the metrics we employed during testing.

This brief summarizes the calling component of this evaluation, including details on our methodology, the solutions we tested, and key findings.

Note: this evaluation was commissioned by Zoom. However, the findings provided in this report are unbiased and represent Wainhouse testing results and related perspectives on the topic. All these platforms are under constant improvement, and it should be recognized that what Wainhouse tested in Q4 2020, is likley to be different today. The reader should evaluate this, and all other evaluations against their own unique environment and enterprise requirements.

Wainhouse Quality Evaluation – Enterprise Calling

In Q4 2020, evaluated market-leading enterprise calling platforms from 8x8, Microsoft, RingCentral, Cisco, and Zoom. Our primary objective was to quantify and compare audio quality between each platform, using established and accepted industry standards to remove subjective user or tester influence. We performed this evaluation in our tightly controlled lab environment following a repeatable process adapted from our existing methodologies and documented guidance for the metrics we employed during testing.

This brief summarizes the calling component of this evaluation, including details on our methodology, the solutions we tested, and key findings.

Executive Summary

Evaluating and comparing quality across solutions has always been a tricky business. The word *'Quality'*, when associated with a meeting or calling solution, is generally interpreted as *'the user experience'* – general, generic, broad, and personal. Many of the variables that make a user experience 'great' are, of course, subjective – experience, familiarity, design, and how a solution 'fits' within a user's workflow. Each enterprise also has a unique set of criteria by which its compares meeting solutions: price, security, manageability, and so on.

But at the base of this conversation, we find an objective layer that serves as the foundation of the user experience – audio quality, video quality, and baseline performance. If a solution does not deliver at this layer, it is unlikely to make its way up the subjective ladder – users are unlikely to adopt low quality solutions, and the enterprise is unlikely to deploy them.

Enterprise Calling and Meetings - Quality Attributes
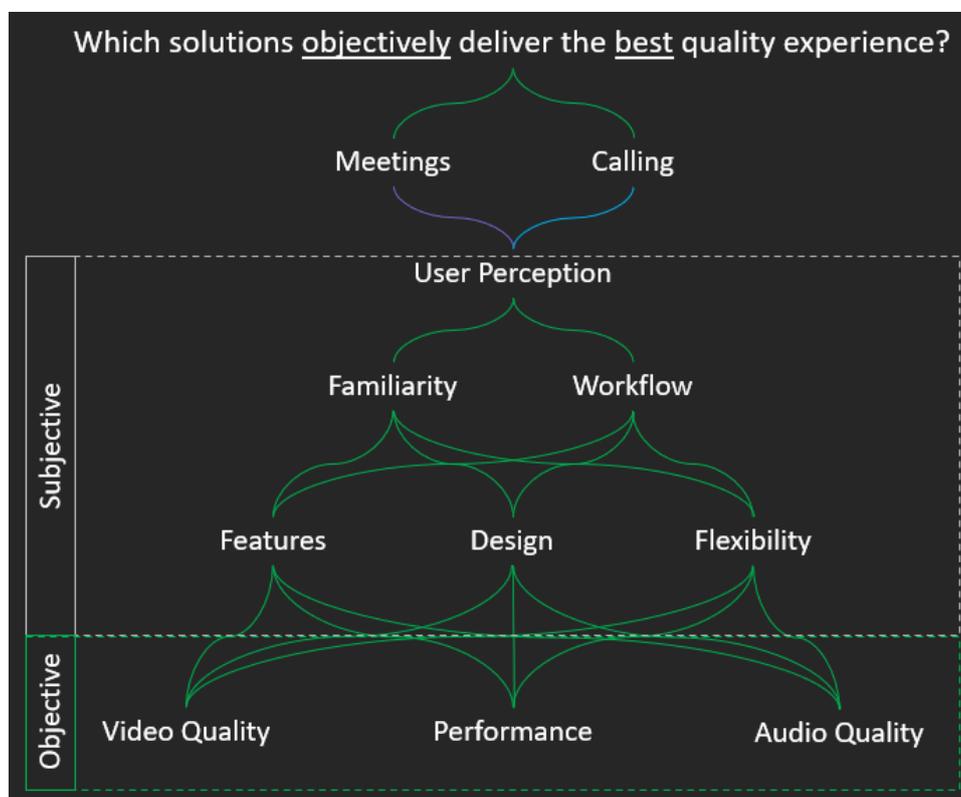
Source: Wainhouse Evaluation Lab, Q1 2021

The objective foundation can be quantified and compared using existing and accepted industry standards. We focused on measuring the user experience at this objective layer, with a metrics-based focus on Audio and Video quality, applying industry standards and established testing procedures. Our evaluation included baseline, packet-loss, and latency tests, with each being repeated multiple times to validate results.

In Q4 2020, we applied this methodology to evaluate solutions from the market-leading enterprise cloud vendors: 8x8, Cisco, Google, Microsoft, RingCentral, and Zoom. We performed two separate-but-related evaluations, one focused on Enterprise Meeting solutions, and one focused on Enterprise Calling solutions. This brief outlines our methodology and a few top-level results from our **Calling evaluation** with a focus on **VoIP audio quality**.

Evaluation Timeframe
This evaluation ran from November through December 2020. However, the bulk of this effort consisted of environmental configuration, pre-testing, and control validation. The final results summarized in this report are from tests conducted in **the first week of December 2020.**

Evaluation Methodology
Wainhouse employs a repeatable evaluation methodology following a detailed test script that is applied to each solution. Core elements include:

Lab Environment
The WH lab consists of a set of control endpoints, distributed across multiple locations – West Chester, Ohio and Boulder, Colorado were the primary locations for this evaluation.

The lab employs a set of endpoints intended to reflect a common enterprise environment – for this evaluation, WH primarily used a relatively new (12-months-old) desktop, a relatively new laptop, and an aging desktop for quality- and performance-based tests. Additional endpoints were used when required (e.g., audio/video file capture), falling within a similar hardware / aging footprint as described in the table below.

Wainhouse Evaluation Lab – Endpoint Details

Lab Endpoint - Description	CPU	Memory	GPU	OS	WAN	Location
1 yr-old Desktop - discrete GPU	AMD Ryzen 5 2400G	16 GB	AMD Radeon RX 580	Win10 Home (19041)	Cable 500MB down 20MB up	West Chester, Ohio
1 yr-old Laptop - Integrated GPU	i7-1065G7 @1.3GHz (4 Cores)	16 GB	Intel Iris Plus	Win10 Pro (19041)	Cable 1GB down 35MB up	Boulder, Colorado
5 yr-old Desktop - discreet GPU	i7-5820K @3.3GHz (6 cores)	32 GB	NVIDIA GTX 1070	Win10 Pro (19041)	Cable 1GB down 35MB up	Boulder, Colorado

Source: Wainhouse Evaluation Lab, Q4 2020

Environmental Controls

The lab environment is intended to reflect an average end-user's environment – accessing cloud calling services via the public internet, antivirus software enabled, etc. However, lab-specific controls were implemented to ensure a fair and consistent set of results across each solution, including:

Windows Resources: each endpoint was configured to reduce process and resource conflicts between tests: disabling shared storage / sync services and disabling indexing and antivirus for any folders sourcing or capturing video files. Note antivirus and indexing services were running through each test, but relevant folders were excluded from each. Endpoints were rebooted in between each test.

Time-of-day: all tests requiring public internet were conducted outside of business hours, after 5 pm ET and on weekends.

Network: both send and receive / capture clients were in the same physical location, on the same network and subnet for each test. We did not enable QoS within this network for these tests.

Standards

We leveraged **ViSQOL** (Virtual Speech Quality Objective Listener) as our primary audio quality metric in this evaluation. ViSQOL was specifically for measuring quality issues associated with Voice over IP (VoIP) solutions and delivers an estimation of subjective listening quality using an objective measurement technique.

Process Detail

We applied the following high-level test flow for this evaluation:

1) Reference Files: We adapted two 16 kHz ITU reference files for testing VoIP solutions (one male voice, one female) – modifications were minimal and required to deliver consistent results within our evaluation environment.

2) Virtual Microphone: Each reference file was played at its original sample rate through a virtual audio driver. Sample rates were adjusted for all virtual drivers after multiple rounds of pre-testing to validate the highest and most consistent results before the final evaluation tests were conducted.

3) Endpoint Control: We cleaned (rebooted) each test PC and checked each calling application for updates prior to each test cycle, making sure only the target test application was running at the time. We muted video and disabled microphone gain and any active noise suppression in those applications that provided related controls.

4) Network Conditions: We start each test under 'normal' network conditions (Baseline), and then adjust network conditions across four packet-loss and latency steps – two files, 10 baseline iterations and five iterations per degradation step – 50 tests in total for each solution.

5) Capture Client: We mute the receive client's microphone and play each audio file while monitoring the microphone gain in our capture software – adjusting driver gains to align captured audio within a target dB range.

6) Capture software: Each audio file is recorded at the same sample rate and bit-depth as the original reference files.

7) Scoring: Each 'degraded' test file is compared to the original 'reference' video using specialized scoring software – producing a ViSQOL score for each recording. Iteration scores are then averaged to deliver a final score for each test category.

Audio Quality – Test Methodology
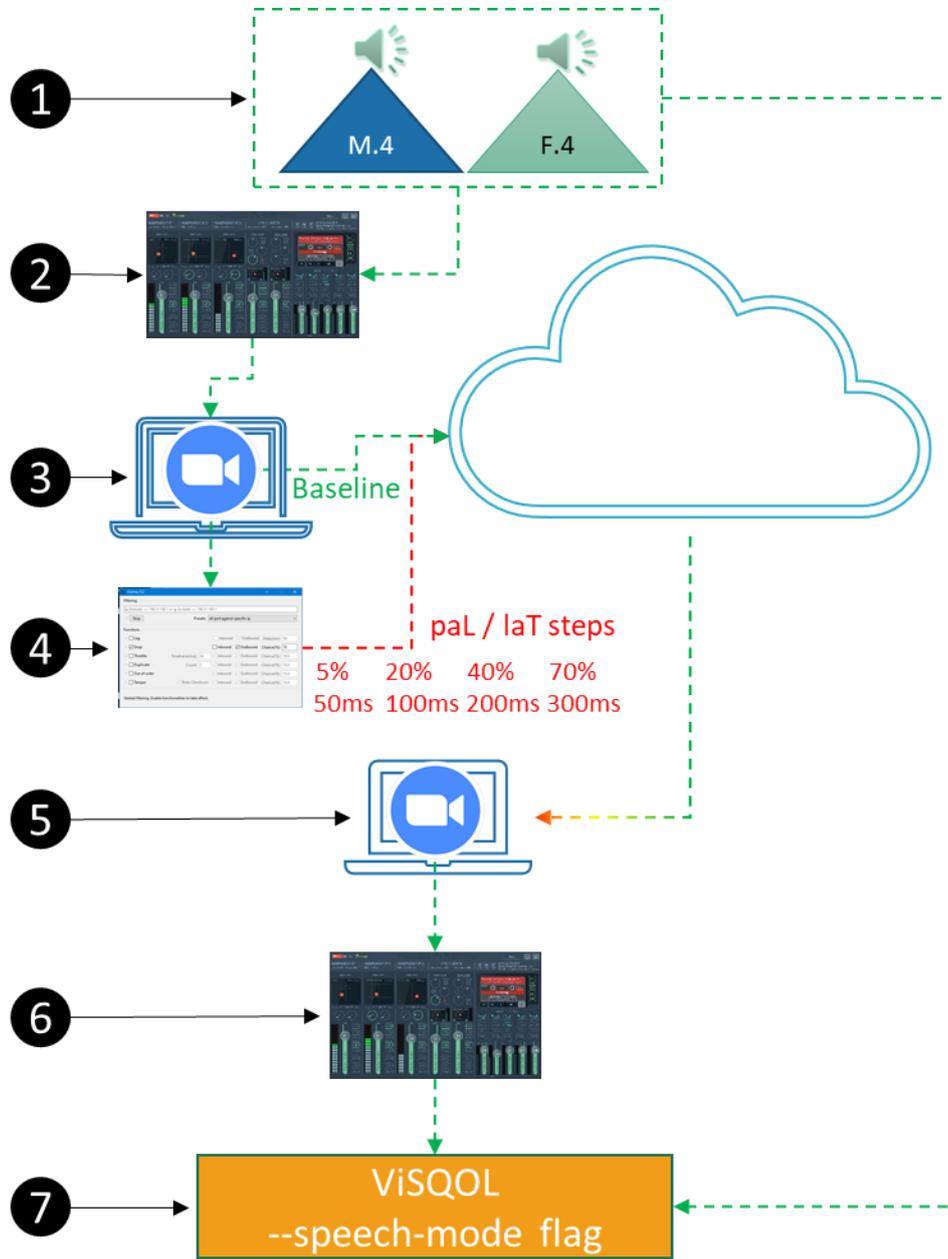
Source: Wainhouse Evaluation Lab, Q4 2020

Evaluated Solutions

WH evaluated five market leading Cloud Calling solutions: 8x8 Express, Microsoft Office 365 Phone System, RingCentral Office, Cisco Webex Calling, and Zoom Phone, as detailed in the following table. We conducted this evaluation using each solution's standard calling application with the exception of Cisco Webex Calling – we evaluated Cisco's calling experience across its Webex Teams and Webex Calling (BroadSoft's UC-One rebranded) applications.

Note we evaluated solutions across a range of product tiers. The mix of product tiers was primarily a function of existing in-place Wainhouse tenants and purchasing / procurement requirements. In each case, Wainhouse evaluated each procured solution's feature-set to ensure it aligned with the core Quality-focused requirements for this evaluation – primarily support for HD audio, video, and PSTN calling. Each solution was marketed and positioned as supporting the core required feature set, and Wainhouse therefore expects no change in related quality scoring based on the selected service tiers.

Calling Evaluation – Service Description

CALLING EVALUATION	8x8	MSFT Teams	RingCentral	Webex Teams	Webex UC-One	Zoom Phone
Evaluated Plan	8x8 Express	Microsoft O365 E5 + Phone	RC Office Essentials	Cisco Webex Calling		Zoom Pro + Phone
Evaluated Solution	8x8 Work 7.3.5-3	v.1.3.00.28779 (64-bit)	RC Desktop 20.4.22	v.40.11.0.17133	v.22.9.18	v.5.3.2
VoIP Codecs	G.711, G.722	Satin, Silk Wide, G.722	Opus	Opus, ILBC	Opus	Opus

Note: capabilities reflect the plans and services used at the time of this evaluation.
Source: Wainhouse Evaluation Lab, Q4 2020

Summary Results

Baseline Scores

Description: This is a matrix chart that serves as our primary format to visualize quality-related results, comparing three related results within the same segment or test category. Here we are comparing each solution's average baseline ViSQOL Score on the X-axis, the average packet-loss ViSQOL Score on the Y-axis, & the Average Latency ViSQOL Score on the Z-axis (circle size). Rankings are based on average ViSQOL scores across all tests for each solution.

Note the color pattern within this chart – these are a visual representation of the established quality range for each metric. These ranges follow a 1 (bad) to 5 (excellent) Absolute Category Ranking scale for each metric. Each metric includes a standard quality range – however, Wainhouse adapted and extended each range based on complexities and limitations that may have reduced or otherwise negatively impacted scores (e.g., virtual audio drivers, sampling, and encoding processes). That said, all solutions were scored within the same controlled environment – we believe the scores provide a relevant view of each platform's relative relationship.

Analysis: This metric-based quality evaluation highlights just how competitive the enterprise calling market is: each solution delivered at least 'good' scores, with half the field delivering average scores above the 4.0 'excellent' threshold during our baseline tests (positioned left-to-right).

By plotting average baseline quality and loss-handling ViSQOL scores we position each solution into one of four quality quadrants, with the **upper-right quadrant** representing above-average quality during baseline *and* packet-loss tests. Here we find only Zoom Phone, the clear leader in this Enterprise Calling quality evaluation. Note, however, two solutions scored above Zoom in Baseline tests (again, left-to-right) – you will see Cisco Webex UC-One and RingCentral solutions with the top

baseline ViSQOL scores. However, these solutions did not match Zoom's loss-handling capabilities, and those lower scores dropped their average totals accordingly.

The **bottom-right quadrant** represents excellent VoIP quality, but below-average loss handling. Cisco Webex UC-One and RingCentral solutions both receive an analyst hat-tip for their excellent (and consistent) baseline VoIP scores here – both solutions delivered rock-solid, best-in-class quality during baseline tests. Unfortunately, this feat was negated by poor loss-handling, with Cisco Webex UC-One specifically standing out for its inability to handle dropped packets.

The **upper left quadrant** represents above-average loss-handling, but below-average audio quality – here we find Microsoft and WebEx Teams solutions. These solutions, along with Zoom, share a few things in common: a combination of established IP, experience, and (most importantly) solid development teams.

And the **bottom-left quadrant**, affectionately termed the Basement Quadrant, represents below-average quality across baseline and packet-loss tests. Here we find 8x8 with the lowest average score across all VoIP tests. Overall, we were a little surprised at this solution's lower baseline ViSQOL scores considering the vendor has a long tenure in the cloud calling market; of course, early success can make it difficult to keep up with the pace of innovation.

You may also be wondering why we did not focus on the latency results in this evaluation. The honest explanation: our test flow resulted in latency scores within the expected margin of error compared to baseline results – compare each solution's average baseline and latency scores in the table below and you will see they are exact or within a few hundredths of a point. We have received feedback from a number of smart development teams that testing jitter-handling (where voice packets are received out of order) is more valuable than latency tests – we agree and are working to add this component to future audio quality evaluations.

ViSQOL Results – Average Scores
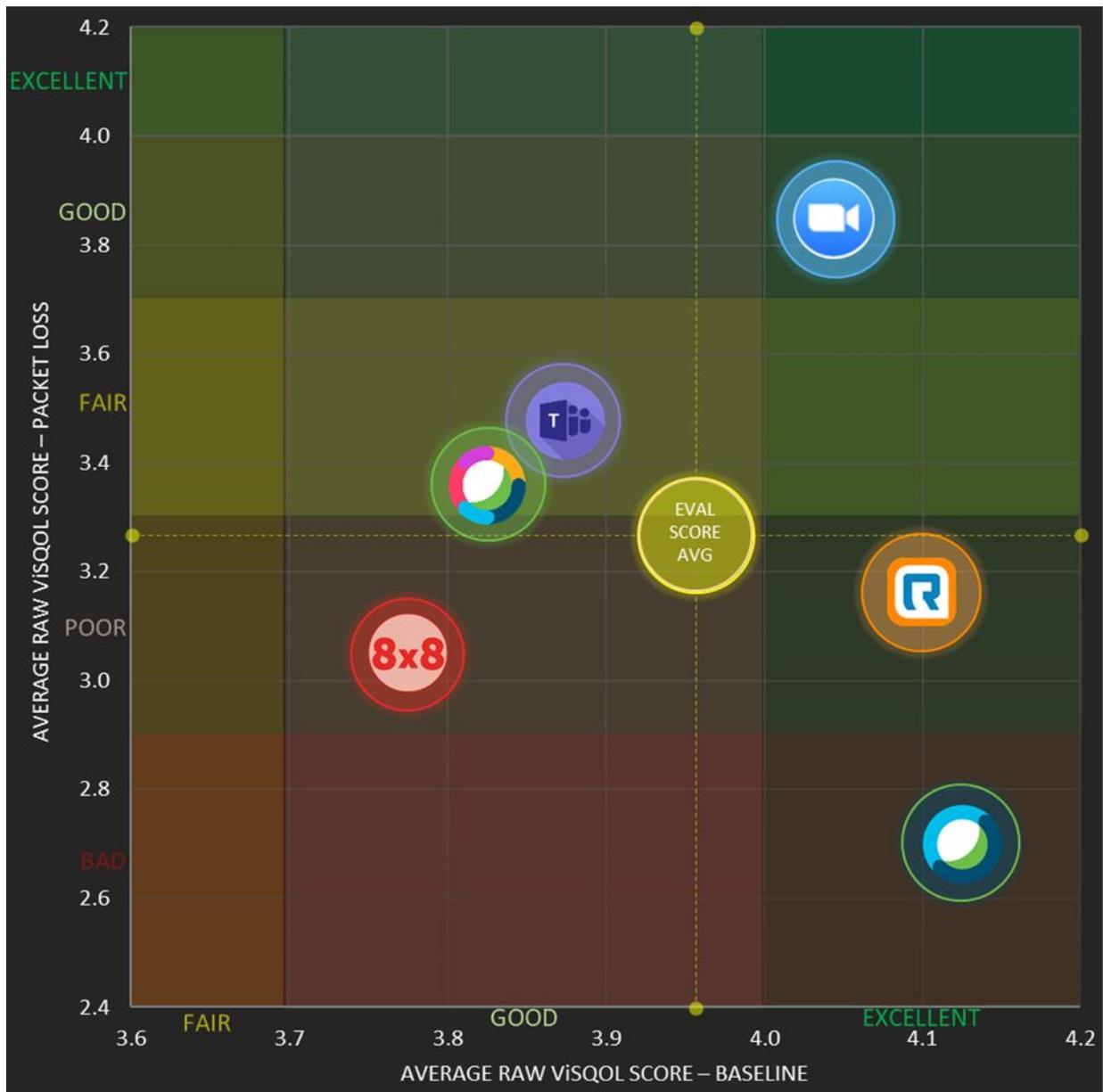
Source: Wainhouse Evaluation Lab, Q4 2020

ViSQOL Results – Average Scores

SECTION	Zoom Phone	RingCentral	Microsoft Teams	Webex Teams	Webex UC-One	8x8	AVERAGE
Baseline	4.05	4.10	3.87	3.83	4.12	3.77	**3.96**
Packet Loss	3.85	3.16	3.48	3.36	2.70	3.05	**3.27**
Latency	4.02	4.12	3.87	3.83	4.12	3.76	**3.95**

Source: Wainhouse Evaluation Lab, Q4 2020

Network Degradation

Description: This 'Teardrop Chart' plots each solution's average ViSQOL scores for all baseline and packet-loss tests: 5%, 20%, 40%, and 70% packet-loss steps. Note the X-axis here positions each solution into its own vertical swim lane, with the field's average scores on the far left and rankings based on average ViSQOL scores across all tests for each solution.

Analysis: This chart highlight's Zoom's technical strength in the loss-handling category, relative to the competition – these average ViSQOL scores show Zoom Phones quality declines only 10% between baseline and 70% packet Loss. Pull Zoom's scores out of the mix, and you can compare its 10% decline to an average 46% decline across the rest of the field. Compare this to Cisco Webex UC-One and RingCentral, the two solutions delivering the highest average baseline scores – these two declined a respective 57% and 48% between baseline and 70% packet-loss.

Microsoft represents Zoom's closest competition in terms of packet-loss handling, holding even with Zoom through 20% loss – but after that it falls off, ultimately dropping 34% between baseline and 70% packet-loss tests.

ViSQOL Results – Baseline and Packet-Loss Average Scores
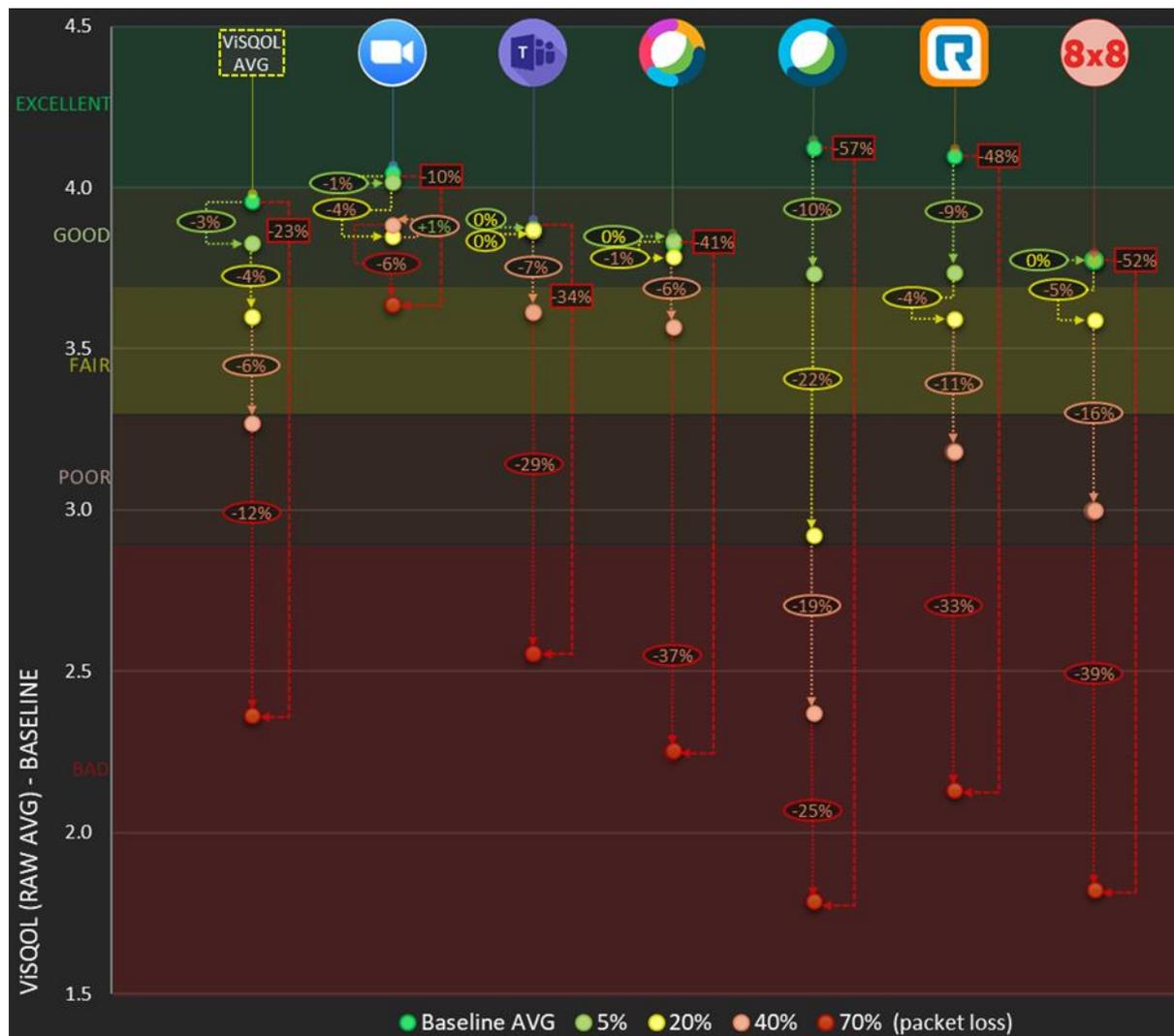
Source: Wainhouse Evaluation Lab, Q4 2020

ViSQOL Results – Baseline and Packet-Loss Average Scores

SECTION	Zoom Phone	RingCentral	MS Teams	Webex Teams	Webex UC-One	8x8	AVERAGE
Baseline	4.05	4.10	3.87	3.83	4.12	3.77	**3.96**
Packet Loss							
5%	4.02	3.74	3.87	3.83	3.73	3.78	**3.83**
20%	3.85	3.59	3.87	3.79	2.92	3.59	**3.60**
40%	3.89	3.18	3.61	3.57	2.37	3.00	**3.27**
70%	3.64	2.13	2.56	2.26	1.79	1.82	**2.37**
Latency							
50ms	4.02	4.08	3.87	3.83	4.10	3.77	**3.94**
100ms	4.05	4.11	3.89	3.83	4.13	3.77	**3.96**
200ms	4.01	4.14	3.88	3.83	4.13	3.73	**3.95**
300ms	4.03	4.14	3.86	3.82	4.11	3.76	**3.95**

Source: Wainhouse Evaluation Lab, Q4 2020

Summary

Our goal in this brief is to provide transparency into our process and methodologies, and a few examples of the output our approach delivers. Our objective is to maintain objectivity and to continually improve our process – we welcome related questions and feedback on this note.

Keep in mind this evaluation was conducted in a controlled lab environment – we think the results are useful from a comparative point of view, but individual experiences will vary based on each unique environment. Also note, this evaluation was conducted at a point in time, and each vendor has been busy updating, iterating, and enhancing their platforms. Wainhouse will continue this series over time, keeping tabs on each solution as they evolve, retesting, re-evaluating, and reporting our findings as we move forward.

About Us

About the Authors

Bill Haskins is a Senior Analyst at Wainhouse Research with a strategic focus on unified communications products and services. Bill has over 15 years of experience supporting, delivering, and designing converged Collaboration services in a global communications environment. He has authored multiple white papers and articles detailing the keys to a successful UCC implementation and delivered various UCC presentations, highlighting his experience integrating Collaboration solutions into business process and enterprise applications.

Bryan L Hellard is a Researcher at Wainhouse where his primary focus is product evaluation and testing. He has 20 years of experience in the industry across several roles, including product engineering and management, R&D, and end user consulting. Prior to Wainhouse Research, he was President of True View Video where he developed video conferencing related products and consulted with end users on best practices for collaboration. Bryan has also been a consultant to video collaboration vendors providing product design services. He lives in the Cincinnati, Ohio area.

About Wainhouse Research

Wainhouse Research is an independent analyst firm that focuses on critical issues in the unified communications and collaboration market. The company provides 6 different vendor subscriptions covering unified communications, enterprise video, meeting room collaboration, personal & web-based collaboration, and audio conferencing, as well as a single all-inclusive subscription for enterprise users. The company acts as a trusted advisor providing strategic advice and direction for both the UC&C industry and its enterprise users. For further details contact sales@wainhouse.com or see http://www.wainhouse.com.

Notices