IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 15 April 2023, accepted 21 April 2023, date of publication 28 April 2023, date of current version 17 May 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3271525

RESEARCH ARTICLE

A Perceptual Evaluation of Music Real-Time Communication Applications

DANA KEMACK GOOT, HARSHIT CHAUBEY, TIMOTHY Y. HSU, AND WILLIAM SCOTT DEAL
Purdue School of Engineering and Technology, Indiana University–Purdue University Indianapolis, Indianapolis, IN 46202, USA

Corresponding author: Dana Kemack Goot (dgoot@iu.edu)

ABSTRACT Music Real-time Communication applications (M-RTC) enable music making (musiking) for musicians simultaneously across geographic distance. When used for musiking, M-RTC such as Zoom and JackTrip, require satisfactorily received acoustical perception of the transmitted music to the end user; however, degradation of audio can be a deterrent to using M-RTC for the musician. Specific to the audio quality of M-RTC, we evaluate the quality of the audio, or the Quality of Experience (QoE), of five network music conferencing applications through quantitative perceptual analysis to determine if the results are commensurate with data analysis. The ITU-R BS.1534-3 MUlti Stimulus test with Hidden Reference and Anchor (MUSHRA) analysis is used to evaluate the perceived audio quality of the transmitted audio files in our study and to detect differences between the transmitted audio files and the hidden reference file. A comparison of the signal-to-noise ratio (SNR) and total harmonic distortion (THD) analysis to the MUSHRA analysis shows that the objective metrics may indicate that SNR and THD are factors in perceptual evaluation and may play a role in perceived audio quality; however, the SNR and THD scores do not directly correspond to the MUSHRA analysis and do not adequately represent the preferences of the individual listener. Since the benefits of improved M-RTC continue to be face-to-face communication, face-to-face musiking, reduction in travel costs, and depletion of travel time, further testing with statistical analysis of a larger sample size can provide the additional statistical power necessary to make conclusions to that end.

INDEX TERMS MUSHRA, music real-time communications (M-RTC), networked music, perceived audio quality, perceptual evaluation, quality of experience (QoE), signal to noise ratio (SNR), telematic, total harmonic distortion (THD), web RTC.

I. INTRODUCTION

Music Real-time Communication applications (M-RTC) enable music making (musiking) for all levels of musicians simultaneously across geographic distance. Real-time Communications (RTC) [1] allows voice and video communication of individuals and groups over the Internet. Network Music Performance software (NMP), as described by the authors in [2], are software applications designed specifically for real-time communications which allow musiking between participants in different physical locations via the Internet [2]. However, M-RTC can include any conferencing application that is capable of transmitting audio over the Internet. M-RTC has seen an increase in demand due to social distancing and its use is perpetuated due to the value of face-to-face musiking and reduced travel cost and travel time. When used for musiking, M-RTC such as Zoom and JackTrip, require satisfactorily received acoustical perception of the transmitted music to the end user; however, degradation of audio can be a deterrent to using M-RTC for the trained musician. While latency of real-time communication is also a concern for musicians, it has been previously analyzed [3], [4], [5].

The associate editor coordinating the review of this manuscript and approving it for publication was Guangjie Han.

II. LITERATURE REVIEW

When experiencing face-to-face musiking over the Internet in real-time, the musician distinguishes the musical timbre of pitch, the duration of notes, and the loudness of music from other transmitted sounds. The transmission and audibility of our research is specific to the audio quality of M-RTC.

We evaluate the quality of the audio, or the Quality of Experience (QoE), of five network music conferencing applications through quantitative perceptual analysis with live human listeners to determine if the results are commensurate with data analysis. Unrelated musical and communicative sounds may be perceived negatively by the musician and may deter from the use of M-RTC. The transmission of audio from M-RTC applications can be evaluated and can provide for more informed choices of M-RTC technology for the musician.

A. EVALUATION OF QUALITY OF EXPERIENCE (QOE)

There are several methods of audio evaluation, including Perceptual Speech Quality Measure (PSQM) [6], which grade the subjective assessment of audio of telecommunications methods based on human perceptibility of the differences between a reference audio file and degraded audio signals. Practical limitations of testing networks led to Perceptual Evaluations of Speech Quality (PESQ) 2001 [7], [8] which defines objective methodology in which algorithms compare reference signals to listener signals for the prediction of the subjective quality of 3.1 kHz (narrow-band) handset telephony and narrow-band speech codecs. A third method for objectively measuring audio quality is Perceptual Evaluation of Audio Quality (PEAQ), a recommendation of the International Telecommunications Union in 1998 and updated in 2001, which codes perceptual properties of the human ear and then integrates multiple model output variables into a single metric [9]. Each method seeks to evaluate audio based on specific properties of the auditory system that cannot be assessed through traditional physical metrics.

The authors of [10], [11] argue that PEAQ is not the superior assessment indicator for audio streaming applications over lossy networks. They used an experimental evaluation method to analyze the PEAQ of a 30 second audio file under different packet loss rates. The objective difference test was used to compare the reference file (the undecoded file) against a degraded file (the decoded file), with a resulting score called the Objective Difference Grade (ODG). The ODG values shown range between −4.0 (very annoying impairment) to 0.0 (imperceptible impairment). The resulting ODG scores did not correspond to the amount of packet loss randomly assigned through MATLAB and state that PEAQ fails even when small external impairment such as packet loss is introduced to the decoded audio signal regardless of the encoding rates and error concealment methods. They conclude that the ODG analysis is not a superior method of measurement of PEAQ.

The authors of [12] examine the effect of impaired quality of audio codecs typically used in current digital audio broadcasting (DAB) systems and web-casting applications as perceived by the end user. A subjective listening test was given to participants based on ITU-R Rec. BS.1116-1 with ratings ranging from 1 (very annoying) to 5 (imperceptible) for audio files including a degraded sample and reference file. There were six different sample files, each recorded at two different bit rates (kbps) per sample for a total of 72 test items. Higher bit rates resulted in improved performance with negligible differences in quality between the codecs. The degradation introduced by the lossy audio codecs at the lowest investigated bit rates result in a largely negative impact on perceived audio quality. The perceived audio quality of the investigated codecs at the highest bit rates resulted in negligible differences between them and transparent audio quality amongst them. Assessments are within expectations in comparison to predicted assessments, although the perceptual Mean Opinion Score (MOS) values are lower than Perceptual Objective Listening Quality Assessment (POLQA) predicted MOS values. POLQA, Recommendation ITU-T P.863, is a standard that covers a model to predict speech quality by means of analyzing digital speech signals [13]. The authors in [14] indicate that the low MOS of 2.96 compared to the predicted MOS of 3.90 in instrumental music indicates that POLQA is not appropriate in instrumental test cases.

Additionally, audio evaluation by the authors in [15] assess the performance of codecs Opus and Enhanced Voice Services (EVS), which were tested through vocal, instrumental, and mixed music signals using POLQA [14]. The effect of lower bitrates (16.4 kbit/s and 20 kbit/s respectively for EVS and OPUS), indicate high degradation. In general, Opus codec correlated to the POLQA results, indicating more consistent assessments than EVS over all bitrates and music pieces tested than EVS.

The authors of [14] also focus on the audio and speech quality assessment of the Opus codec within the real-time communication mode. As in the work by the authors of [15], digital signals are assessed through POLQA, and additionally, through the non-standardized Audio Quality Analyzer (AQuA) from Sevana company. The authors state that the prediction of the instrumental measures should closely correspond to quality scores from a human listening test, considered a subjective test. Test files consisted of natural read speech examples (full band, studio recording), wideband speech with emotional and neutral speech, and full band music and singing examples. The testing of prototypical cases of Opus coding via a WebRTC framework found that the instrument assessment (POLQA) achieves a similar MOS (up to 4.73) as the standalone coding (4.39). A comparison is warranted for further assessment of WebRTC applications using the Opus codec.

Evaluation of the quality of network and protocols of Real-Time Communication (RTC) applications have been undertaken by [1] in identifying and collecting network traffic packet traces for RTC applications under different conditions. Their evaluations find that most of these applications use the Real-time Transport (RTP) protocol in combination with STUN/TURN, but each has its own peculiarities, such as the sending of redundant data or FEC (Forward Error Correction). Assessments have also been made on latency, which is not the subject of this paper. However, latency is often found to be rooted at the access layer procedures [3], and due to indirect packet routing, which is proposed to be reducible through SpaceRTC [4], [5].

B. THE NEXT STEPS IN AUDIO EVALUATION

Increased use and continued development of RTC and network music conferencing applications demands further evaluation of its QoE based on bandwidth control and congestion control [16]. While analyses have been made on audio codecs as perceived through perceptual assessment, a gap exists in analysis of commonly used network conferencing applications or RTC specifically utilized in music rehearsal and performance settings, which we have designated as M-RTC. There is also a lack of ability to consider the human auditory experience and human preferences. The QoE of M-RTC audio is of importance to the musician that may not wish to experience echoes and artifacts while attempting to communicate and collaborate with a fellow musician online [17]. Protocols, bandwidth, and Internet traffic can contribute to degradation of audio quality [18] and, as discussed, discrepancies exist among the evaluation tools for audio QoE. While attempting to minimize Internet transmission contributors to degradation, a determination is sought on whether the QoE of the perceived transmitted audio of M-RTC applications by the end user is commensurate with computer analysis.

III. MATERIALS AND METHODS

A. PURPOSE

The purpose of this research is to determine if the perceived audio quality of M-RTC is analogous to its analysis of signal-to-noise ratio (SNR) and total harmonic distortion (THD) analysis. QoE of RTC has been previously analyzed based on the network and protocols [1], latency [3], and audio [10], [11]. However, there is discrepancy as to which methods of audio evaluation are superior [1], [15]. The ITU-R BS.1534-3 MUlti Stimulus test with Hidden Reference and Anchor (MUSHRA) has been designed as a subjective measurement of intermediate quality level of audio systems and has been used to evaluate technology such as headsets, speech codecs for telecommunication, and audio codecs [19]. MUSHRA analysis will be used to evaluate the perceived audio quality of the transmitted audio files in our study and detect differences between the transmitted audio files and the hidden reference file. A comparison of the SNR and THD analysis to the MUSHRA analysis will describe the commensurate findings between computerized test results and perceptual analysis.

B. CONCEPTUAL FRAMEWORK DESCRIPTION

Through quantitative analysis of the independent variable, the perceived audio quality utilizing MUSHRA, audio transmitted through five M-RTC applications; Deck 10, JackTrip, JamKazam, SonoBus, and Zoom will be evaluated as shown in (fig. 1). The measurement of THD and SNR of the five M-RTC applications, the dependent variable (DV), will be compared to the independent variable (IV), the quantitative

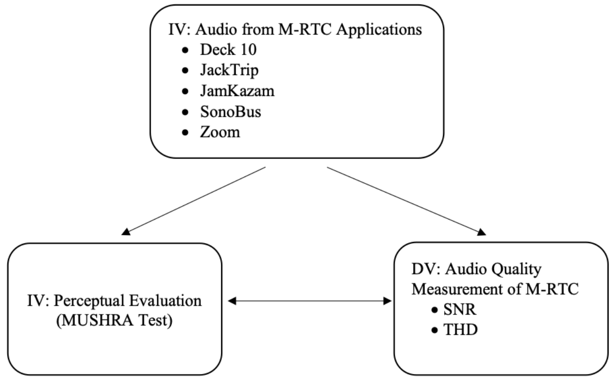

FIGURE 1. The conceptual framework of the audio evaluation of the perceived audio quality of five M-RTC applications.

perceptual evaluation [19]. The replication of audio transmission that typically occurs in rehearsal and performance situations will be analyzed.

C. PARTICIPANTS

The 15 respondents (two female and 13 male) of the evaluation are 30-61 years, with a mean age of 44.77, and were recruited through non-professional and professional musical organizations through social media and email. Three respondents indicate being professional musicians, two respondents instruct music at the university level, and one respondent is a PhD student. The remaining respondents work in professional fields such as engineering, software development, and broadcast engineering. Nine of the respondents consider their level of musicianship as professional. Three respondents rate themselves as advanced musicians, and the remainder rate themselves as intermediate level, with no respondents rating themselves as beginners. The respondents are made up of 26% percussionists and 21% string players. Other instrument families played by the respondents include brass, wind, keyboard, and voice. These percentages indicated that some respondents play more than one instrument. M-RTC has been used by 94% of the respondents, with 26% having used Zoom, 10% having used Jamulus and SonoBus each, 16% having used JackTrip, and 6% having used JamKazam and Jamulus each. Respondents' frequency of use of M-RTC range from having used it once at 13.33% to weekly use at 33.33%, while another 33.3% indicated "other" usage, which may include never having used an M-RTC. Incomplete surveys were deleted from the results. Inclusion criteria required participants to be 18 years or older with exclusions of participants with known hearing impairments.

D. METHODOLOGY

A quantitative non-experimental design through a survey was approved by the Indiana University Institutional Review Board (IRB) on March 1, 2022, protocol number 14349. The MUlti Stimuli with Hidden Reference and Anchor (MUSHRA) test, which has previously been utilized as a method of evaluation for a wide range of audio degradation [13], is utilized as the quantitative perceptual evaluation method. The MUSHRA test employs the Continuous Quality Scale (CQS) with one or more properties and assesses and detects degradations between the audio test files and the reference file. The MUSHRA test method employs original unprocessed audio files, in this case the original audio files that had been transmitted through the chosen M-RTC, and a hidden reference. ITU-R BS.1534-3 recommends audio files of a maximum length of approximately ten seconds, and no greater than 12 seconds in order to not fatigue the listener [15]. Limited audio file length also ensures retainability, as medium- and long-term aural memory is not reliable [1]. Presentation order of audio test files was randomized.

E. AUDIO TEST FILES

The audio test files (48KHz sample rate, 24-bit), including the anchor files, were recorded and created in a digital audio workstation (DAW), Logic Pro X, except the pure sine wave (440 Hz, 48KHz sample rate, 24-bit) which was downloaded from the Internet. The sine wave was included for additional subjective and objective analysis. The testing audio files were a sine wave of five seconds, a spoken word file of nine seconds, and an instrumental music sample (acoustic guitar and digital drum track) of eight seconds. The spoken word was recorded by a female voice using a Shure SM57 cardioid dynamic microphone and Focusrite interface. The spoken word data was based on International Telecommunication Union – Telecommunication (ITU-T) recommendations of audio test signals purposed to evaluate end-to-end telecommunications systems' quality in addition to assessing the compliance of applications to certain recommendation requirements [20]. Whereas in testing that is largely used for neutral assessment of speech processing systems or devices, artificial voices are used as the signal test [21] since they are more easily produced than actual speech and have a smaller variability than examples of real voices [20]. We chose to use speech recorded by a live human voice precisely for a wider frequency variability than would be offered by the artificial voice and also for the natural frequency range of the voice that would be found in M-RTC when two or more musicians transmit live voice and/or instrumentation across the Internet. The recorded text was Harvard Sentences and read as, "The ship was torn apart on the sharp reef. Sickness kept him home the third week. The box will hold seven gifts at once. Jazz and swing fans like fast music [22]."

F. AUDIO TRANSMISSION

Two MacBook Pro computers were used to send audio over the Internet through each M-RTC. Peer one computer (P1), 2019 2.6 GHz 6-Core Intel Core I7 sent the files to peer two computer (P2), 2019 3.2GHz 6-Core Intel Core I9 as shown in (fig 2). The transmission to the Internet was via a research university's Internet as the service provider to minimize extrinsic latency factors.

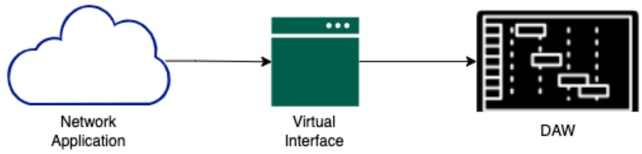

FIGURE 2. Internal routing from M-RTC (Network Application) to Virtual Interface, and DAW of P2.

Three rounds of each audio file were transmitted over Deck 10 Studio, JackTrip, JamKazam, SonoBus, and Zoom, applications which represent industry standards for academic and non-academic institutions. Each of the transmissions occurred five minutes apart on Saturdays between the dates of February 19, 2022, and March 5, 2022, between the hours of 3 p.m. – 5 p.m. EST. Additionally, all M-RTC applications utilize both User Datagram Protocol (UDP) and Transmission Control Protocol (TCP), while Differentiated Services Code Point (DSCP) is also used by Zoom (Table 1) [4], [5]. Although UDP transmission is faster than TCP, transmission of packets is dependent on the many routing protocols of the Internet Protocol, with retransmission of packets achievable only by TCP (IP) network [23]. The inherent nature of UDP's absence of the retransmission of lost data packets conceivably could result in noticeable packet loss, resulting in degradation of audio files and a poor QoE.

G. ZOOM TRANSMISSION

For the first test over Zoom, both P1 and P2 were logged in through their respective user Zoom account within the Zoom application. In order to fully optimize the tests, each machine was configured to send only audio and no video, with the camera feed disabled throughout the entire transmission. Zoom transmission was accessed through its ability to share

TABLE 1. Internet protocol information.

Network Application	Codec	Transmission Layer Protocol Audio or Audio and Video
Deck 10 Studio	Opus [24]	UDP & TCP (TCP is used only when UDP is restricted) [24] Audio and video
JackTrip	Opus [25]	UDP & TCP [26] Audio
JamKazam	No literature found.	UDP & TCP [27] Audio Paid upgrade version offers video
SonoBus	Opus [28]	UDP & TCP [29] Audio
Zoom	SILK & Opus [30]	UDP, TCP, DSCP [31] Audio and video

computer audio only under its advanced tab. P2 was configured to receive the audio directly into the DAW via Zoom as input to its audio interface preference.

H. SONOBUS AND JAMKAZAM TRANSMISSION

Similar setup was repeated for both SonoBus and JamKazam application transmission, except the audio received on P2 was routed to through the DAW using the virtual sound interface Blackhole as shown in (fig. 2). Zoom, SonoBus, and JamKazam have the ability to record media directly within their applications. However, Deck 10 and JackTrip do not have recording capability; therefore, in order to create the same recording platform for all applications, audio transmission was received directly into the Digital Audio Workstation (DAW), Logic Pro X.

I. DECK 10 AND JACKTRIP TRANSMISSION

The setup for Deck10 was different compared to the former standalone applications. Deck10 is a WebRTC tool; subsequently, Chrome browser was used for both P1 and P2. Deck 10's ability to add local media enabled the P1 user to share the files in a Deck 10 peer-to-peer (P2P) session meeting. P1 user then played each file and P2 user recorded the receiving audio in Logic Pro X using the same virtual interface, Blackhole, v0.3.0, 2 channel configuration (fig. 3). The JackTrip connection was established over JackServer and using JackRouter as the input within the DAW's audio preference, the received files were recorded.

J. DATA ANALYSIS

Quantitative analysis was carried out through MATLAB Audio Toolbox's Total Harmonic Distortion (THD) and Signal to Noise Ratio (SNR) functions to calculate degradations in comparison to the original sine wave files as shown in (Table 2). The THD was tested for the sine wave file, where 20 harmonics are accounted for in the THD calculation, based on the average of the three rounds of transmission for each network application type. The SNR is calculated in C-weighted decibels (dBC), and the average SNR in dBC is reported in (Table 2).

IV. PERCEPTUAL EVALUATION RESULTS

Audio files and survey questions were compiled through Qualtrics. Survey respondents were instructed to listen to seven versions of each audio example; a hidden reference file, an anchor file, and the recorded audio, each transmitted through the five M-RTC applications for a total of 21 audio files, and to rate them in comparison to the quality of the original reference file. Participants were instructed to listen with headphones or earbuds, with 71.43% utilizing headphones, 14.29% using earbuds, and the remainder choosing to use either computer or external speakers.

FIGURE 3. M-RTC Application (Deck 10)) sending audio from P1 to P2 over Internet and recorded through interface (Blackhole) to DAW.

A. SINE WAVE

The audio quality in comparison to the reference file was to be rated on as per the MUSHRA scale of 0-100 with the reference file at 100 (and a rating of 0-20 bad, 21-40 poor, 41-60 fair, 61-80 good, 81-100 excellent). A second rating was of annoyance, marked with levels 1-5; with 5 as imperceptible, 4 perceptible, but not annoying; 3 slightly annoying; 2 annoying; and 1 very annoying. Respondents listened to all audio files in the conditions of their own home, office, or other surroundings. Respondents had the ability to listen to the files more than once.

The mean audio quality rating of the sine wave transmitted through Deck 10 and JackTrip was similar to the reference file. The reference file was rated 86.0 and Deck 10 and JackTrip were rated 84.0 and 84.1, respectively, in the excellent category. Zoom's mean audio quality was rated good at 78.8, SonoBus was rated fair at 51.1, and JamKazam's rating was in the bad category at 18.7. The anchor file was rated at 4.3 as shown in (Table 3). The mean of Deck 10's level of annoyance was 4.5, similar to the reference file, which was rated 4.4, both ratings between perceptible, but not annoying and imperceptible. Attribution to Deck 10's

TABLE 2. Average total harmonic distortion and signal to noise ratio.

Network Application	Total Harmonic Distortion (THD)	Signal to Noise Ratio (SNR)
Deck 10	- 89.0 dB	43.3dB
JackTrip	-117.4 dB	89.4 dB
JamKazam	- 63.7 dB	19.5 dB
SonoBus	- 56.9 dB	20.0 dB
Zoom	- 90.7 dB	47.7 dB

TABLE 3. Perceptual evaluation - sinewave.

Network Application	Audio Quality Mean 81-100 Excellent 61-80 Good 41-60 Fair 21-40 Poor 20-0 Bad	Level of Annoyance 5-imperceptible 4-perceptible, but not annoying 3-slightly annoying 2-annoying 1-very annoying
Anchor File	4.3	1.3
Reference File	86.0	4.4
Deck 10	84.1	4.5
JackTrip	84.1	4.1
JamKazam	18.7	1.7
SonoBus	51.1	3.2
Zoom	78.8	4.0

score that is slightly above the reference file can relate to conditional factors of the listener such as mood, fatigue, and attention. However, the score of 4.5 is well within the range of JackTrip at 4.1 and the reference file's score. All audio files except SonoBus, JamKazam, and the anchor file were rated at levels of perceptible, but not annoying. Zoom and SonoBus' level of annoyance was rated 4.0 and 3.2 respectively. Both JamKazam at 1.7, and the anchor file at 1.3 were between the levels of very annoying and annoying as shown in (Table 3).

B. SPOKEN WORD

The mean of the spoken word reference file was rated good at 75.1, with the mean of the anchor file rated bad at 4.6. JackTrip received the same good rating category as the reference file at 75.1, and Deck 10 rated 73.9. Zoom, JamKazam, and SonoBus were respectively rated 69.9, 67.8, and 48.9, as shown in (Table 4). The mean of the level of annoyance was rated perceptible, but not annoying for Deck 10, JackTrip, and Zoom each at 4.0, similar to the reference file at 3.9, JamKazam was rated 3.8, slightly annoying, and SonoBus was rated 2.6, annoying. The anchor file's mean rating was 1.3 as shown in (Table 4).

C. MUSIC SAMPLE

The audio quality of the music sample reference file was rated 65.6, in the category of good, and the anchor file rated bad at 5.2. The mean audio quality of Zoom was rated good at 60. JamKazam and JackTrip rated in the fair category and near good at a mean of 59.9 and 59.5 respectively. Deck 10 rated at 57.5 and SonoBus at 54.1 in the fair category as shown in (Table 5). The mean level of annoyance of the music sample for the reference file was rated imperceptible at 5.0, which was the same for Deck 10. Zoom at 3.5, JamKazam at 3.4, and JackTrip 3.1 rated slightly annoying each. The anchor file was rated 1.0, very annoying (Table 5).

TABLE 4. Perceptual evaluation - spoken word.

Network Application	Audio Quality Mean	Level of Annoyance
	81-100 Excellent	5-imperceptible
	61-80 Good	4-perceptible, but not annoying
	41-60 Fair	3-slightly annoying
	21-40 Poor	2-annoying
	20-0 Bad	1-very annoying
Anchor File	4.6	1.3
Reference File	75.1	3.9
Deck 10	73.9	4.0
JackTrip	75.1	4.0
JamKazam	67.8	3.8
SonoBus	48.9	2.6

TABLE 5. Perceptual evaluation - music sample.

Network Application	Audio Quality Mean	Level of Annoyance
	81-100 Excellent	5-imperceptible
	61-80 Good	4-perceptible, but not annoying
	41-60 Fair	3-slightly annoying
	21-40 Poor	2-annoying
	20-0 Bad	1-very annoying
Anchor File	5.2	1
Reference File	65.6	5
Deck 10	57.5	5
JackTrip	59.5	3.1
JamKazam	59.9	3.4
SonoBus	54.1	3.1
Zoom	60.0	3.5

D. DATA ANALYSIS RESULTS

SonoBus scored the highest THD at −56.9 dB, followed by JamKazam at −63.7 dB. Deck 10 and Zoom had lower THD levels at −89.0 dB and −90.7 dB respectively as shown in (Table 2). JackTrip scored the lowest THD level of −117.4 dB. Research indicates that a THD of 37.8% is very high, 3% distortion is audible, and 1-2% may be audible in certain circumstances, such as during a flute solo [32]. The evaluation of the test audio files was found to be less than .5% for all applications and is noted as negligible.

The SNR of JamKazam and SonoBus scored the lowest at 19.5 dB and 20.0 dB, respectively. The Deck 10 audio measured 43.3 dB and Zoom measured 47.7 dB. JackTrip had the highest SNR at 89.4 dB. Generally, a higher SNR results indicating stronger signal strength in relationship to noise, allows for higher data rates, and when TCP is utilized results in fewer retransmissions. JamKazam's SNR of 19.5 dBC relates to approximately an amplitude factor of 90. JamKazam's SNR is lower than compact disc quality (∼96 dB). The noise in these platforms may be primarily attributed to the audio codec used, particularly since packet loss was deemed negligible in this study.

E. FREQUENCY RESPONSE

The power spectrum of the original 440 Hz sine wave, the voice audio, and the instrumental audio are depicted in (fig. 4, 5, and 6), respectively. In comparison to the original sine wave, JackTrip indicates a sharp frequency cutoff at 20 kHz on both the sine wave and the voice audio. JackTrip's frequency drops off lower at approximately 13 kHz in the instrument example. Similar to JackTrip, SonoBus also has drop offs at 20 kHz on the three audio files. JamKazam's higher THR and lower SNR analysis are visually depicted by sharp drop offs at 20 kHz for all three audio files and by periodic distortions on the sine wave. JamKazam and SonoBus indicate higher levels at 18kHz for both the voice and instrument audio files in comparison to the original files. Zoom's frequency response gradually lowers at approximately 18 kHz on all three audio files in comparison to the original files. Deck 10 also shows a gradual decline at approximately 18-19 kHz on all three of the audio files. All M-RTC applications indicate levels of drop off at approximately 18 kHz on all audio files, with the exception of JackTrip whose audio drops off at 15 kHz on the instrumental audio.

F. STATISTICAL ANALYSIS

As varying listening conditions exist in this study and there is no method that is universally appropriate in assessment of all testing conditions of audio quality, evaluation must be specific to the research scenario. An analytical and statistical

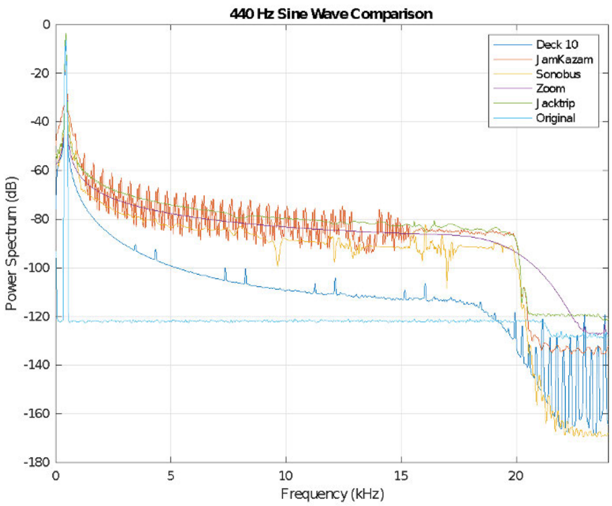

FIGURE 4. Sine wave power spectrum.

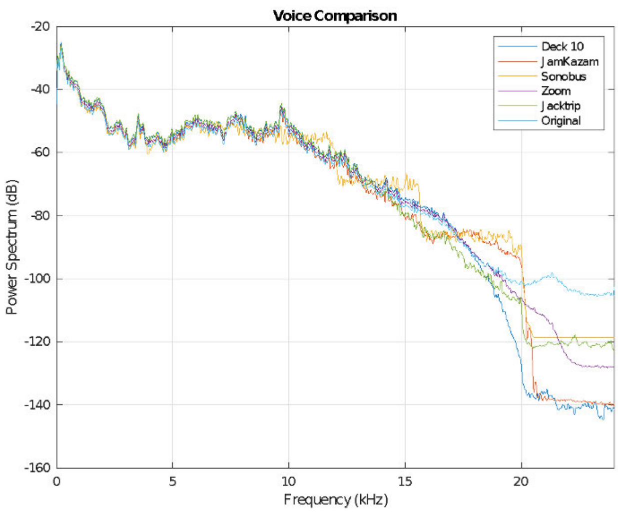

FIGURE 5. Voice power spectrum.

methodology must be implemented to alleviate experimental insensitivity that may be related to the choice of audio material or weak components of the research scenario. Otherwise, a ''null'' result cannot be considered as valid [33].

In determining appropriate statistical methods, consideration is given to the sample size and outliers. While a slightly outlying response was noted on the rating of one audio file, the respondent's other ratings were in line with the other respondents' ratings; therefore, the response is not ruled an outlier and is not excluded from the analysis. While mean values are required for some statistical methodologies, the median is an alternative and robust measure of central tendency, which is ideal for small sample sets, non-normal distributions, or datasets with noted outliers. There are testing scenarios in which these concerns are not an issue; however, there is benefit of robust analysis methods that are less affected by possible validity altering factors [33]. Therefore, several statistical methodologies are implemented.

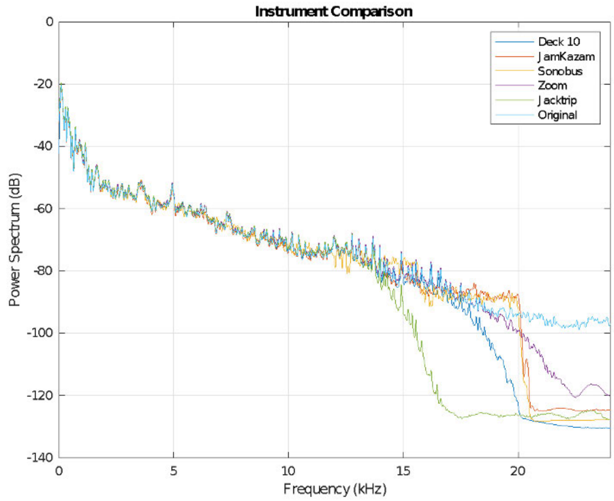

FIGURE 6. Instrument power spectrum.

First a calculation is made of the Aggregate Quality Score (AQS) for the 0-100 scale and Likert scale questions to prevent human bias in the process. Proclaiming an AQS reveals the aggregate effect of software on quality output. Scores across the three questions were calculated to generate a new metric, the AQS, which incorporates the overall quality and gives equal weight to all three aspects of Sinewave Quality, Spoken Word Quality, and Music Quality in an improved way as seen in (Table 6).

Secondly, for the 0-100 scale a regression is generated using dummy variables, where the dependent variable is the AQS score, and the independent variable is the utilized software. Regression analysis allows capture of the full variability of the dataset without violating assumptions and skewing results. The intercept of this regression represents the original audio's AQS. The interpretation of the regression of ''whether or not using a software benefits/hurts the audio quality and, if so, which software performs better than the other,'' is as shown in the summary output (Tables 7, 8, 9).

The regression yields an R-square of 34.8%. It is clear from the regression output that using SonoBus or JamKazam significantly impairs the audio quality. As seen in the provided Mean values of (Table 4), SonoBus and JamKazam have significantly lower quality than the original. Moreover, Zoom, JackTrip, and Deck 10 are kept in the analysis to demonstrate that their higher quality does not affect the AQS; therefore, they have better audio quality. Rerunning the regression while excluding the variables with high p-values slightly increases the R-square and gives similar coefficient results.

TABLE 6. Mean, median, and standard deviation for M-RTC applications.

	Zoom	SonoBus	JamKazam	JackTrip	Deck 10	Original
Mean	209	154	146	219	209	227
Median	220	163	153	234	211	239
Standard Deviation	47	47	40	49	45	44
Skewness	-0.29	-0.64	0.65	-1.15	-0.51	-0.42

TABLE 7. Regression analysis for 0-100 scale.

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	226.71	12.22	18.55	0.00	202.38	251.04
Zoom	-18.00	17.28	-1.04	0.30	-52.41	16.41
SonoBus	-72.64	17.28	-4.20	0.00	-107.05	-38.24
JamKazam	-80.29	17.28	-4.65	0.00	-114.69	-45.88
JackTrip	-8.07	17.28	-0.47	0.64	-42.48	26.34
Deck 10	-11.29	17.28	-0.65	0.52	-45.69	23.12

TABLE 8. Regression statistics.

Multiple R	0.590
R Square	0.348
Adjusted R	0.306
Standard Error	45.726
Observations	84.0

TABLE 9. ANOVA.

	df	SS	MS	F	Significance F
Regression	5	86863	17373	8.3089	2.4725E-06
Residual	78	163085	2091		
Total	83	249948			

For the Likert Scale, the AQS metric corresponds to the sum of scores across the three questions, with a range from 1 to 5 each. The Frequency Distribution (fig. 7) clearly demonstrates the outperformance of Zoom, Deck10, and JackTrip over JamKazam and SonoBus.

Zoom's output has significantly different medians from Sonobus and JamKazam, indicating, that all three (JackTrip, Deck 10, and Zoom) have quality equivalent to the original and are superior to both SonoBus and JamKazam (Table 10).

A one-way of variance (ANOVA) was made and found consistent with previous results, that SonoBus and JamKazam significantly impair the quality by bringing the AQS down from 12 to 9, leading to an average rating of 3 instead of 4. Therefore, the quality of SonoBus and JamKazam is inferior to that of the other applications (Table 11, 12, 13).

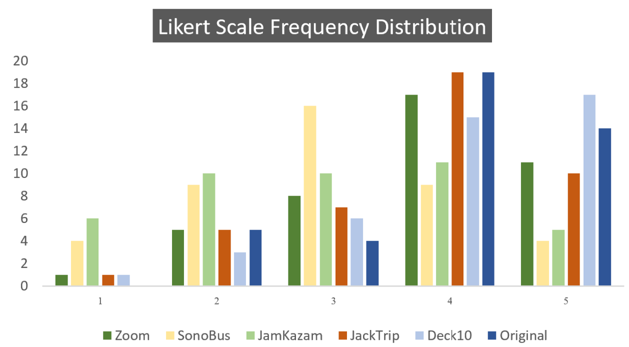

FIGURE 7. Likert scale frequency distribution.

TABLE 10. Mode and median summary likert scale for M-RTC applications.

	Zoom	SonoBus	JamKazam	JackTrip	Deck 10	Original
Mode	4	3	4	4	5	4
Median	4	3	3	4	4	4

TABLE 11. Regression statistics.

Multiple R	0.553
R Square	0.305
Adjusted R Square	0.261
Standard Error	2.091
Observations	84.0

TABLE 12. Regression analysis.

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	12.00	0.56	21.47	0.00	10.89	13.11
Zoom	-0.50	0.79	-0.63	0.53	-2.07	1.07
Sonobus	-3.00	0.79	-3.80	0.00	-4.57	-1.43
JamKazam	-3.07	0.79	-3.89	0.00	-4.64	-1.50
Jacktrip	-0.71	0.79	-0.90	0.37	-2.29	0.86
Deck10	0.14	0.79	0.18	0.86	-1.43	1.72

TABLE 13. ANOVA.

	df	SS	MS	F	Significance F
Regression	5	150	30	6.86	2.3671E-05
Residual	78	341	4		
Total	83	491			

TABLE 14. Kruskal Wallis test – by "original".

Software	p-value
Zoom	0.111
SonoBus	0.031
JamKazam	0.705
JackTrip	0.001
Deck 10	0.001

Finally, the Kruskal Wallis test was implemented to check whether the median of each software's output is different (alternative hypothesis) from the original or not as seen in (TABLE 12). The Kruskal-Wallis test's null hypothesis is sometimes stated to be that the group medians are equal. Even though the medians are the same, the Kruskal-Wallis test can reject the null hypothesis if the distributions differ. Hence, null hypothesis (H_o): the given software's median does not differ at 5% significance level (alpha value .05%) and alternative hypothesis (H_a): the given software's median does differ from the original (Table 14).

Therefore, we can reject the null hypothesis and conclude that JackTrip, Deck 10, and SonoBus have medians significantly different from the original, which means that the sample comes from the same distribution, concluding that both tests and the control perform differently.

V. DISCUSSION

Statistical analysis shows a relevant and significant difference in output of each software. The statistical analysis is commensurate with the perceptual and data analysis. Perceptual evaluation and data analysis indicate that JackTrip rates in the top of three categories (audio quality, THD, and SNR) and second in level of annoyance. Deck 10 rates best in average level of annoyance, ties for second-best in audio quality, and is close to second-best in the categories for THD and SNR. Zoom rates second-best in all categories, except audio quality in which it rates midway between the other M-RTC applications. SonoBus rates second to the lowest in audio quality and SNR categories, with the lowest rating for THD. SonoBus is equal to JamKazam for level of annoyance below the other applications, in which it ties with JamKazam for the lowest rating of the M-RTC applications. Finally, JamKazam's rating for THD scores second to the lowest, and in the other categories of audio quality and SNR, it rates the lowest. Considering both the data testing and the perceptual analysis JackTrip scored the highest overall, with Deck 10 and Zoom following close to second in best quality. SonoBus and JamKazam have lowest scores in the M-RTC applications analysis.

While an overall evaluation based on objective and perceptual assessments can be made, the respective results are not analogous between the objective and perceptual findings. More specifically, the objective metrics may indicate that SNR and THD are factors in perceptual evaluation and may play a role in perceived audio quality; however, the SNR and THD scores do not directly correspond to the MUSHRA analysis and do not adequately represent the preferences of the individual listener. Furthermore, it has not been proven that subjective assessment of audio can be exchanged by prevailing objective measurement [34] and the objective testing methods do take into consideration the human auditory system. A general quality rating can be determined by analyzing the perceptual evaluation in combination with the SNR and THD metrics, but individual analysis of data or perceptual testing does not indicate M-RTC applications' ranking similarly between these data and perceptual test categories. Since the benefits of improved M-RTC continue to be face-to-face communication, face-to-face musiking, reduction in travel costs, and depletion of travel time, further testing with statistical analysis of a larger sample size may provide more statistical power necessary to make more conclusions to determine if data analysis is consistent with perceived analysis among these platforms.

A. POTENTIAL SETBACKS AND WEAKNESSES

The perceptual evaluation of Zoom could have been analyzed at its default setting in addition to its music setting to compare the two and to compare its default settings to the other M-RTC applications. Another approach to controlling the conditions for all applications would have been to make adjustments to each M-RTC application in order to improve the sound quality. However, the adjustments that would have been made would have been based on the perceptual evaluation of the test organizers and would not necessarily represent the preferences of other listeners.

It may be feasible to assume that some survey respondents rated their level of annoyance on the level of perceived musical quality in addition to perceived audio quality for the audio instrumental example, as indicated by optional survey comments. Optional survey comments included, "I assume we are focusing on the audio recording quality, but it was hard to stop focusing on the musicians not playing together," and "The guitarist wasn't staying in time well so they were all pretty annoying. The back beat had really good presence in G though." This may account for the overall higher level of annoyance on all instrumental audio examples, including the reference file.

B. THREATS TO VALIDITY AND LIMITATIONS

Threats to validity are selection bias and sampling characteristics. Additional threats to validity include the control of the participants' listening environment and the consistency of that environment throughout the survey sessions and adjustment to the Zoom settings to music and professional audio with echo cancellation, which are not the default settings. The other applications were analyzed at their default settings. Limitations include possible degradation of the sound due to the encoding and decoding of the data packets, resulting in noise at the output level, which results in inferior QoE for M-RTC [15].

C. FUTURE WORK

Isolation of audio at points of encoding and decoding should be evaluated in future work to discern their contribution to SNR, THD, and to the MUSHRA analysis. M-RTC has the potential for increased use by musicians of all levels, in home, educational, and professional settings. M-RTC with high quality audio, absent of mediocre sound quality and degradations, are essential for a positive QoE, which enables an M-RTC experience free of unnecessary audio degradations for all musicians. A high QoE is likely to increase potential use of M-RTC to musicians of all levels. Although an overall evaluation based on objective and perceptual assessments can be made, the respective results are not analogous between the objective and perceptual findings and do not adequately represent the preferences of the individual listener. Future work would need to independently analyze each codec from factors related to Web-RTC and Internet. Studying the trade-offs between applications would provide research to offer higher audio quality, lower latency, and higher video quality when used in conjunction with processing speed and bandwidth. Additionally future work may include interviewing the application designers and engineers to understand the tradeoffs in the decision-making process.

ACKNOWLEDGMENT

The author Dana Kemack Goot offers sincere appreciation to the Donald Tavel Arts and Technology Research Center for its support of this project.

REFERENCES

[1] A. Nistico, D. Markudova, M. Trevisan, M. Meo, and G. Carofiglio, "A comparative study of RTC applications," in *Proc. IEEE Int. Symp. Multimedia (ISM)*, Dec. 2020, pp. 1–8, doi: 10.1109/ISM.2020.00007.

[2] C. Rottondi, C. Chafe, C. Allocchio, and A. Sarti, "An overview on networked music performance technologies," *IEEE Access*, vol. 4, 2016, pp. 8823–8843, doi: 10.1109/ACCESS.2016.2628440.

[3] N. Nikaein and S. Krea, "Latency for real-time machine-to-machine communication in LTE-based system architecture," in *Proc. 17th Eur. Wireless Sustain. Wireless Technol.*, 2011, pp. 1–6.

[4] Z. Lai, W. Liu, Q. Wu, H. Li, J. Xu, and J. Wu, "SpaceRTC: Unleashing the low-latency potential of mega-constellations for real-time communications," in *Proc. IEEE INFOCOM Conf. Comput. Commun.*, May 2020, pp. 1339–1348, doi: 10.1109/infocom48880.2022.9796887.

[5] J. H. Davis and J. R. Cogdell, "Calibration program for the 16-foot antenna," Elect. Eng. Res. Lab., Univ. Texas, Austin, TX, USA, Tech. Memo, Tech. Rep., NGL-006-69-3, Nov. 1987.

[6] *Transmission Systems for Communications*, 3rd ed., Western Electric Co., Winston-Salem, NC, USA, 1985, pp. 44–60.

[7] (2001). *Perceptual Evaluation of Speech Quality (PESQ): An Objective Method For End-to-End Speech Quality Assessment of Narrow-band Telephone Networks and Speech Codecs*. [Online]. Available: https://www.itu.int/rec/T-REC-P.862

[8] G. O. Young, *Synthetic Structure of Industrial Plastics*, vol. 3, 2nd ed., J. Peters, Ed. New York, NY, USA: McGraw-Hill, 1964, pp. 15–64. [Online]. Available: http://www.bookref.com

[9] *Method for Objective Measurements of Perceived Audio Quality*, document ITU-R BS.1387-1, ITU Radiocommunication Bureau (BR), 2001. [Online]. Available: https://www.itu.int/rec/R-REC-BS.1387-1-200111-I/en

[10] A. F. Khalifeh, A.-K. Al-Tamimi, and K. A. Darabkh, "Perceptual evaluation of audio quality under lossy networks," in *Proc. Int. Conf. Wireless Commun., Signal Process. Netw. (WiSPNET)*, Mar. 2017, pp. 939–943.

[11] J. S. Turner, "New directions in communications," *IEEE J. Sel. Areas Commun.*, vol. 13, no. 1, pp. 11–23, Jan. 1995.

[12] P. Pocta and J. G. Beerends, "Subjective and objective assessment of perceived audio quality of current digital audio broadcasting systems and web-casting applications," *IEEE Trans. Broadcast.*, vol. 61, no. 3, pp. 407–415, Sep. 2015, doi: 10.1109/TBC.2015.2424373.

[13] *Perceptual Objective Listening Quality Prediction*, document ITU-T P.863, ITU Radiocommunication Bureau (BR), Apr. 2020. [Online]. Available: https://www.itu.int/rec/T-REC-P.863

[14] O. Jokisch, M. Maruschke, M. Meszaros, and V. Iaroshenko, "Audio and speech quality survey of the opus codec in web real-time communication," *Elektronische Sprachsignalverarbeitung*, vol. 81, pp. 254–262, Mar. 2016.

[15] M. Maruschke, O. Jokisch, M. Meszaros, F. Trojahn, and M. Hoffmann, "Quality assessment of two full-band audio codecs supporting real-time communication," in *Proc. Int. Conf. Speech Comput.* Cham, Switzerland: Springer,Aug. 2016, pp. 571–579.

[16] J. Fang, M. Ellis, B. Li, S. Liu, Y. Hosseinkashi, M. Revow, A. Sadovnikov, Z. Liu, P. Cheng, S. Ashok, D. Zhao, R. Cutler, Y. Lu, and. J. Gehrke, "Reinforcement learning for bandwidth estimation and congestion control in real-time communications," 2019, *arXiv:1912.02222*.

[17] M. Iorwerth and D. Knox, "The application of networked music performance to access ensemble activity for socially isolated musicians," in *Proc. Web Audio Conf. Diversity Web Audio*, 2019, pp. 1–7. [Online]. Available: https://edshare.gcu.ac.uk/id/eprint/5179

[18] B, García, F. Gortázar, M. Gallego, and A. Hines, "Assessment of QoE for video and audio in web-RTC applications using full-reference models," *Electronics*, vol. 9, no. 3, p. 462, 2020, doi: 10.3390/electronics9030462.

[19] *Method for the Subjective Assessment of Intermediate Quality Level of Audio Systems*, document ITU-R BS.1534-3, Radiocommunication Sector of ITU, Oct. 2015. [Online]. Available: https://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.1534-3-201510-I!!PDF-E.pdf

[20] S. Bech and N. Zacharov, "Experimental variables," in *Perceptual Audio Evaluation—Theory, Method, and Application*. Chichester, U.K: Wiley, 2006, ch. 5, pp. 97–104.

[21] *Telephone Transmission Quality, Telephone Installations, Local Line Networks, Objective Measuring Apparatus*, document ITU-T P.50, ITU-T Telecommunication Standardization Sector, Sep. 1999. [Online]. Available: https://file:///Users/danagoot/Documents/IUPUI/IUPUI%202022%20PhD/Papers/SERIES%20P-%20TELEPHONE%20TRANSMISSION%20QUALITY%20TELEPHONE%20INSTALLATIONS%20LOCAL%20LINE%20NETWORKS%20Objective%20measuring%20apparatus%20Artificial%20voices%20.pdf

[22] Institute of Electrical and Electronics Engineers, "IEEE recommended practice for speech quality measurements," *IEEE Trans. Audio Electroacoust.*, vol. AE-17, no. 3, pp. 225–246, Jun. 1969, doi: 10.1109/IEEESTD.1969.7405210.

[23] J. F. Kurose and K. W. Ross, *Computer Networking: A Top-Down Approach*. Boston, MA, USA, Pearson Education, 2010. [Online]. Available: https://www.ucg.ac.me/skladiste/blog_44233/objava_64433/fajlovi/Computer%20Networking%20_%20A%20Top%20Down%20Approach,%207th%20converted.pdf

[24] *WebRTC*. (Apr. 26, 2022). [Online]. Available: https://webrtc.github.io/webrtc-org/faq/#what-is-the-opus-audio-codec

[25] C. Chafe and S. Oshiro. (Apr. 26, 2022). *Jacktrip on Raspberry PI*. [Online]. Available: https://www.raspberrypi.org/products/raspberry

[26] JackTrip. (Jan. 2012). *CCRMA Documentation*. Accessed: Apr. 26, 2022. [Online]. Available: https://ccrma.stanford.edu/docs/common/IETF.html

[27] (Apr. 26, 2022). *JamKazam*. [Online]. Available: https://forum.jamkazam.com/showthread.php?tid=590

[28] (Apr. 26, 2022). *SonoBus*. [Online]. Available: https://www.sonobus.net/

[29] (2022). *Sonosaurus, Sonobus*. Githubs Inc. Accessed: Apr. 26, 2022. [Online]. Available: https://github.com/sonosaurus/sonobus

[30] *Making Video Communications Frictionless*. Zoom. Accessed: Apr. 26, 2022. [Online]. Available: https://explore.zoom.us/docs/doc/Zoom-PremiumAudio.pdf

[31] (Apr. 2020). *Connection process*. Zoom video Commun. Inc. Accessed: Apr. 26, 2022. [Online]. Available: https://explore.zoom.us/docs/doc/Zoom%20Connection%20Process%20Whitepaper.pdf

[32] F. A. Everest and K. C. Pohlmann, "Signals, speech, music, and noise," in *Master handbook of acoustics*, 6th ed. New York, NY, USA: McGraw-Hill, 2015, ch. 5. [Online]. Available: https://www-accessengineeringlibrary-com.proxy.ulib.uits.iu.edu/content/book/9780071841047/chapter/chapter5

[33] *Method for Subjective Assessment of Intermediate Quality Level of Audio Systems*, document ITU-R BS.1534-3, ITU Radiocommunication Bureau (BR), 2015, pp. 14–15. [Online]. Available: https://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.1534-3-201510-I!!PDF-E.pdf

[34] N. Kaplanis, "Quality metering for high-definition audio toward objective metrics," Dept. Music Sound Recording, Univ. Surrey, Guildford, U.K., Tech. Rep., Apr. 2011. [Online]. Available: https://www.researchgate.net/profile/Neofytos-Kaplanis/publication/329943301_Quality_Metering_for_High_Definition_Audio_Towards_objective_metrics/links/5c24d282458515a4c7fbe911/Quality-Metering-for-High-Definition-Audio-Towards-objective-metrics.pdf

DANA KEMACK GOOT was born in Bitburg, West Germany, in 1964. She received the B.G.S. degree in arts and humanities and the M.S. degree in music technology from the Purdue School of Engineering and Technology, Indiana University–Purdue University Indianapolis, Indianapolis, IN, USA, in 2011 and 2021, respectively, where she is currently pursuing the Ph.D. degree in music technology.

She is a Research Fellow with the Donald Tavel Arts and Technology Center, Indiana University–Purdue University Indianapolis. She is the author of "A Spectrum of Online Rehearsal Applications: A Potential Means for Cultural Connection" (*Journal of Network Music and Arts*, 2022). With her background as a musician and in network music application and human–computer interaction, her research is focused on utilizing network music applications to cross-cultural barriers.

Ms. Goot is a member of the Association for Technology in Music Instruction and the College Music Society.

HARSHIT CHAUBEY was born in New Delhi, India, in 1985. He received the B.E. degree in applied electronics and instrumentation from Maharshi Dayanand University, India, in 2008, and the M.Sc. degree in music technology from the Dundalk Institute of Technology (DkIT), UCD, Ireland, in 2013. He is currently pursuing the Ph.D. (ABD) degree in music technology with the Purdue School of Engineering and Technology, Indiana University–Purdue University Indianapolis.

From 2013 to 2014, he carried out design and DSP work with PatchBlocks, a U.K.-based synthesizer company specializing in digital synths. Before starting as a Research Assistant with the Donald Tavel Arts and Research Center, Indiana University–Purdue University Indianapolis, in 2017, he was an Audio and Production Assistant to Los Angeles-based composer Wayne Sharpe. His research interests include the re-engineering framework for musical devices, the impact of change in tangible interaction on the end user and their sonic experiences, and both techniques and tools for live multimedia streaming over webRTC.

TIMOTHY Y. HSU was born in Birmingham, AL, USA, in 1982. He received the B.M. degree in music in piano performance, the B.M. degree in music in recording arts and sciences, and the M.A. degree in acoustics and audio design from the Peabody Institute of The Johns Hopkins University, Baltimore, MD, USA, in 2004, 2005, and 2005, respectively, and the M.S. and Ph.D. degrees in mechanical engineering (acoustics) from the Georgia Institute of Technology, Atlanta, GA, USA, in 2007 and 2012, respectively.

From 2013 to 2018, he was a Visiting Assistant Professor with the School of Music, College of Design, Georgia Institute of Technology. He is currently an Assistant Professor of music technology with Indiana University–Purdue University Indianapolis, Indianapolis, IN, USA. With his background as a musician, a conductor, and an engineer, his research is concerned with human and environmental sustainability in acoustics as it relates to music technology, investigating healthcare acoustic environments, vocal acoustics, acoustic materials, STEM education, and music signal processing.

Dr. Hsu is a member of the Acoustical Society of America, the College Music Society, and the Audio Engineering Society, where he is on the executive board of the Indiana Section of the Audio Engineering Society.

WILLIAM SCOTT DEAL was born in Ann Arbor, MI, USA, in 1957. He received the B.A. degree in music percussion performance from Cameron University, Lawton, OK, USA, in 1980, the M.M. degree in percussion performance from the University of Cincinnati College-Conservatory of Music, Cincinnati, OH, USA, in 1982, and the D.M.A. degree in percussion performance from the Frost School of Music, University of Miami, Coral Gables, FL, USA, in 1994.

From 1995 to 2007, he was an Assistant Professor, an Associate Professor, and a Professor of music with the University of Alaska Fairbanks, Fairbanks, AK, USA. In 2007, he was appointed a Professor of music technology and the Director of the Donald Tavel Arts and Technology Research Center, Purdue School of Engineering and Technology, Indiana University–Purdue University Indianapolis (IUPUI), Indianapolis, IN, USA. He is currently with IUPUI. With a background in music performance and technology, his research is concerned with musical-computer interactivity as it relates to internet-based telematics, machine learning, electro-acoustics, STEM musical education, and diversity, equity, and inclusion of all musicians.

Dr. Deal is a member of the Percussive Arts Society, the International Computer Music Association, the Society of Electro-Acoustic Musicians in the United States, the American Society for Acoustic Ecology, the College Music Society, and the Association for Instruction in Music Technology.