Do You Have the Energy for This Meeting?

An Empirical Study on the Energy Consumption of the Google Meet and Zoom Android apps

Leonhard Wattenbach
Vrije Universiteit Amsterdam
l.j.wattenbach@student.vu.nl

Basel Aslan
Vrije Universiteit Amsterdam
b.aslan@student.vu.nl

Matteo Maria Fiore
Vrije Universiteit Amsterdam
m.m.fiore@student.vu.nl

Henley Ding
Vrije Universiteit Amsterdam
h2.ding@student.vu.nl

Roberto Verdecchia
Vrije Universiteit Amsterdam
r.verdecchia@vu.nl

Ivano Malavolta
Vrije Universiteit Amsterdam
i.malavolta@vu.nl

Abstract

Context. With "work from home" policies becoming the norm during the COVID-19 pandemic, videoconferencing apps have soared in popularity, especially on mobile devices. However, mobile devices only have limited energy capacities, and their batteries degrade slightly with each charge/discharge cycle.
Goal. With this research we aim at comparing the energy consumption of two Android videoconferencing apps, and studying the impact that different features and settings of these apps have on energy consumption.
Method. We conduct an empirical experiment by utilizing as subjects Google Meet and Zoom. We test the impact of multiple factors on the energy consumption: number of call participants, microphone and camera use, and virtual backgrounds.
Results. Zoom results to be more energy efficient than Google Meet, albeit only to a small extent. Camera use is the most energy greedy feature, while the use of virtual background only marginally impacts energy consumption. Number of participants affect differently the energy consumption of the apps. As exception, microphone use does not significantly affect energy consumption.
Conclusions. Most features of Android videoconferencing apps significantly impact their energy consumption. As implication for users, selecting which features to use can significantly prolong their mobile battery charge. For developers, our results provide empirical evidence on which features are more energy-greedy, and how features can impact differently energy consumption across apps.

ACM Reference Format:
Leonhard Wattenbach, Basel Aslan, Matteo Maria Fiore, Henley Ding, Roberto Verdecchia, and Ivano Malavolta. 2022. Do You Have the Energy for This Meeting?: An Empirical Study on the Energy Consumption of the Google Meet and Zoom Android apps. In *IEEE/ACM 9th International Conference on Mobile Software Engineering and Systems (MOBILESoft '22), May 17–24, 2022, Pittsburgh, PA, USA.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3524613.3527812

MOBILESoft '22, May 17–24, 2022, Pittsburgh, P A, USA

ACM ISBN 978-1-4503-9301-0/22/05…$15.00
https://doi.org/10.1145/3524613.3527812

1 Introduction

A great impact on communication was brought into our lives by the internet. The way people were able to communicate drastically changed with all the services that it provided to everybody's homes. Given the widespread use of IT technologies, the sustainability impact of this sector can no longer be neglected [1]. An even bigger change was pushed by mobile devices and the sprout of videoconferencing apps. Software like Skype, Google Talk, and Windows Messenger allowed family members and friends to feel closer with their video call functionalities [2]. During the COVID-19 pandemic years, these types of applications have entered our daily lives and in some cases have become vital to performing in our respective fields, especially for students and "work from home" employees [3]. Zoom, Google Meet, Microsoft Teams, Webex, and GoToMeeting are just some of the names of applications that have been used by educational institutions and companies to organize their everyday activities. Based on market statistics mobile video conferencing apps experienced a staggering usage increase of 250% between March 2020 and June 2021 [4].

By considering the drastic increase of mobile videoconferencing apps, and the limitations in battery capacities of mobile devices, developers need to take another look at their software stack, and carefully evaluate what they want to spend their energy budget on. While historically continuous improvements in efficiency of chips implied that we could rely on new generations of devices to satisfy our efficiency demands, we have seen a slowdown on these fronts [5]. In addition, as lithium-ion batteries degrade in capacity over time as they get charged and drained repeatedly [6], it becomes paramount for users to be able to choose the apps and related settings to optimize energy efficiency.

Despite mobile videoconferencing apps entered our daily lives and users concerns, little research is available on the energy consumption of these apps. To study the energy usage of videoconferencing apps on mobile, our experiments measure the energy consumption of popular videoconferencing apps on Android. We study multiple scenarios to improve the generality of our results, allowing us to represent the real world usage patterns of these apps.

Our study provide in-depth empirical insights into the energy consumption of mobile videoconferencing apps. The ultimate goal is to empower users to make their devices last longer, both during the usage of the apps and in preventing the battery from degrading in capacity. To provide insightful and actionable results, in addition to comparing the energy consumption of two popular videoconferencing apps, our results document empirical evidence on which

features are more energy-greedy, and how features can impact differently energy consumption across apps.

The study focuses exclusively on the energy consumed by videoconferencing apps on Android devices, i.e., we do not aim to assess also the energy consumed by other hardware components involved in the videoconferences (e.g., servers and network infrastructures).

The main contributions of this research are the following:

- An in-depth comparison of the energy consumption of two among the most popular Android videoconferencing apps, namely Zoom and Google Meet;mobile
- An investigation of the impact on energy consumption of different videoconferencing app features and settings, namely *number of participants*, and *camera*, *microphone*, and *virtual background use*;
- A comprehensive replication package, where the entirety of the data and scripts used for this study are made available.[1]

2 Related Work

Two literature reviews conducted by Ong et al. [7, 8] in 2012 and 2014 demonstrated that videoconferencing is more environmentally sustainable (in terms of energy and CO_2 emissions) than face-to-face meetings. Ong et al. noted that videoconferencing uses at most 6.7%-7% of the energy/CO_2 required by face-to-face meetings. The higher sustainability of videoconferences is mostly due to commuting fuel savings, operating transportation infrastructure, and time saved by meeting participants. Differently from the papers of Ong et al. [7, 8], rather than surveying the literature, our study presents an in-depth empirical experiment, executed to measure energy consumption characteristics of mobile videoconferencing apps.

In the work of Maiti and Challen [9] a method to measure the perceived "value" of apps by users is presented. The value is then utilized to manage energy on battery-bound mobile devices more efficiently, by prioritizing apps execution. For example, to compare two videoconferencing applications to improve their energy efficiency, we first need to have a numerator to divide their energy usage by their perceived usefulness, as the apps cannot be compared directly. Maiti and Challen collected energy data from volunteers, and used the data to rank applications in terms of energy efficiency using their model for value as the numerator. The authors conclude that, while the effort to measure an application's value resulted to be a failure, the authors hope that the mobile systems community will continue work in this direction. Our research focuses exclusively on energy consumption of apps, and does not attempt to measure the intrinsic "value" of apps. While utilizing exclusively energy efficiency does not allow us to compare completely different apps to each other, this perspective is out of scope for our research, and enables us regardless of documenting an "apples to apples" comparison between videoconferencing apps.

Faber [10] introduced an adaptable framework for the systematic measurement of CO_2 emissions of video conferences, by considering among others hardware waste, internet energy intensity, network data transmission rates, and server power. The author reports strategies to reduce emissions of video conferences, and use their framework to measure the CO_2 emission of a case-study conference. Faber concludes that, while video conferences emit far less CO_2 than their in person counterparts, there is still a significant amount of CO_2 which we should aim to reduce. Similarly to [7] and [8], our focus is narrower than the one of Faber [10], allowing us to conduct an in-depth investigation on the energy consumption of videoconferencing apps based on concrete empirical values, rather than relying on approximations and estimates.

Trestian et al. [11] investigated the power consumption of Android video streams. The authors tested various playback quality levels, video codecs, and different scenarios, and measured the related energy consumption. The four scenarios considered were the user being located near the access point (with and without any background traffic), and the user being located far away from the access point (with and without any background traffic). Trestian et al. discovered that significant energy efficiency factors are signal quality and network load, followed by codec and playback quality. Compared to the work of Trestian et al., we do not test network factors, and instead focus on videoconferencing app functionalities, by ensuring that we control network related independent variables to avoid potential confounding experimental factors.

3 Experiment Definition

3.1 Research Goal

By utilizing the formulation proposed by Basili et al. [12], the goal of this study is to *analyze* videoconferencing applications *for the purpose of* evaluation *with respect to* their energy consumption *from the point of view of* users *in the context of* Android applications.

3.2 Research Questions

Our study considers the following main research question (RQ1) and sub-research questions (RQ1.1-RQ1.4):

[RQ1]: *To what extent does the energy consumption differ between videoconferencing apps?* To answer this question, we compare the two most popular and free videoconferencing apps from the Google Play Store, namely Zoom[2] and Google Meet[3]. We measure the energy consumed during a fixed length video conference, executed under different settings, as further detailed below.

In videoconferencing apps, settings that might affect the energy consumption are present, e.g., if the camera is activated during the call. We identified the following sub-questions to investigate the impact on energy consumption of meeting settings and parameters.

[RQ1.1]: *What is the impact of different numbers of participants on the energy consumption of videoconferencing apps?* With an increasing number participants, the incoming data stream is also increased, as the video and/or audio of more participants has to be transmitted. In addition, the rendering of the additional received data is expected to increase the energy consumed by the apps.

[RQ1.2]: *What is the impact of using the camera on the energy consumption of videoconferencing apps?* Activating the camera is expected to lead to a higher energy consumption, as the camera has to be powered on, the video stream has to be processed, and transmitted to the server.

[1] https://github.com/S2-group/mobilesoft-2022-replication-package

[2] https://play.google.com/store/apps/details?id=us.zoom.videomeetings

[3] https://play.google.com/store/apps/details?id=com.google.android.apps.meetings

[RQ1.3]: *What is the impact of using virtual backgrounds on the energy consumption of videoconferencing apps?* Virtual backgrounds require computing power to identify which portions of the video belong to the foreground and background. However, the resulting video stream is expected to be smaller, due to the static background, and thus less changing pixels that need to be transmitted.
[RQ1.4]: *What is the impact of using the microphone on the energy consumption of videoconferencing apps?* Similar to activating the camera, an activated microphone is also expected to increase the transmitted data, and hence, the energy consumed.

4 Experiment Planning

4.1 Subjects Selection

To achieve our goal, this research focuses on testing two among the most popular conferencing apps: Google Meet and Zoom. The app selection was made based on the idea to test scenarios as close as possible to real world scenarios and the purpose of having a positive impact to as many users as possible. The two apps were picked among the most popular ones, in terms of total number of downloads from the Google Play Store. The popularity of apps was further verified by considering web articles in which the usage and number of adoptions of these apps over the last two years is reported [3, 4].

4.2 Experimental Variables

The dependent variable for all RQs is the energy consumption, measured in Joules, collected as total energy consumed over the duration of a video conference.

For RQ1.1, the independent variable is the number of participants in the video conference. In the experiment, video conferences with either 2 or 5 participants are considered (as 1 participant is required to start the call, and with 6 participants Google Meet splits camera feeds in a different tab). The number of participants does not vary within a single experimental run.

For RQ1.2, the independent variable is camera use. In the experiment, video conferences with the local camera on and off are conducted. The setting of the camera does not change during the run. The camera of the other participant(s) is turned by default on.

For RQ1.3, the independent variable is virtual background use. In the experiment, video conferences with the virtual background on or off are conducted. Since it is only possible to use virtual backgrounds when the camera is on, the scenarios where the camera is off and the virtual background is on are left out of the experiment.

Lastly, for RQ1.4, the independent variable is microphone use. In the experiment, video conferences with the microphone on or off are conducted. The camera setting does not vary during the run.

4.3 Experimental Hypotheses

To be able to answer our RQs, we formulate the following null and corresponding alternative hypotheses. μ stands for the mean value of the measured energy consumption for the treatment considered.

- $H1_0$: There is no difference in energy consumption between videoconferencing apps.
  $H1_a$: There is a difference in energy consumption between videoconferencing apps.

$$H1_0 : \mu_{\text{zoom}} = \mu_{\text{meet}}$$
$$H1_a : \mu_{\text{zoom}} \neq \mu_{\text{meet}}$$

- $H1.1_0$: The energy consumption is the same for all numbers of participants.
  $H1.1_a$: the energy consumption is higher when there is a higher number of participants.

$$H1.1_0 : \mu_{\text{pair}} = \mu_{\text{multi}}$$
$$H1.1_a : \mu_{\text{pair}} \neq \mu_{\text{multi}}$$

- $H1.2_0$: There energy consumption is the same when using the camera and when not using the camera.
  $H1.2_a$: The energy consumption is higher when using the camera than when not using the camera.

$$H1.2_0 : \mu_{\text{camon}} = \mu_{\text{camoff}}$$
$$H1.2_a : \mu_{\text{camon}} \neq \mu_{\text{camoff}}$$

- $H1.3_0$: The energy consumption is the same when using a virtual background and when not using a virtual background.
  $H1.3_a$: The energy consumption is higher when using a virtual background than when not using a virtual background.

$$H1.3_0 : \mu_{\text{virtualon}} = \mu_{\text{virtualoff}}$$
$$H1.3_a : \mu_{\text{virtualon}} \neq \mu_{\text{virtualoff}}$$

- $H1.4_0$ The energy consumption is the same when using the microphone and when not using the microphone.
  $H1.4_a$ The energy consumption is higher when using the microphone than when not using a microphone.

$$H1.4_0 : \mu_{\text{micon}} = \mu_{\text{micoff}}$$
$$H1.4_a : \mu_{\text{micon}} \neq \mu_{\text{micoff}}$$

4.4 Experiment Design

In Table 1 we document how the treatments are assigned to each of our subjects in order to answer our RQs.

Table 1: Trials of the experiment

Webcam	Background	Mic	Participants
on	on	on	2
on	on	off	2
on	off	on	2
on	off	off	2
off	off	on	2
off	off	off	2
on	on	on	5
on	on	off	5
on	off	on	5
on	off	off	5
off	off	on	5
off	off	off	5

To ensure the correctness and feasibility of the experiment, a pre-testing phase is designed and executed. In this pre-testing phase, the energy consumption of each application for different videoconferencing lengths is studied, as further detailed in Section 5.1.

The experiment phase instead involves executing a combination of independent variables, selected by following our RQs. The

investigation uses a complete design [13], since it allows testing all combination of independent variables, hence reflecting the potential usage of the apps in the real world. Since RQ1 is our main question, the sub-questions are investigated first. The results are then evaluated to reach a conclusion for RQ1.

4.5 Data Analysis

We analyze the experimental data collected *via* 4 main phases, namely: data exploration, normality checking, hypothesis testing, and effect size estimation.

Data exploration: In this initial step we visualize the measured energy consumption values *via* a combination of descriptive statistics, namely histograms, violin plots, and boxplots, to gain preliminary insights into data trends [14]. Following the visualization, we adopt a fitting statistical techniques to test our hypotheses.

Normality testing: We analyze the distribution of the measured energy consumption, to detect whether a sample comes from a normal or non-normal distribution *via* (i) Q-Q Plots, (ii) the Shapiro-Wilk test, and (iii) frequency distribution plots. This analysis allows us to check whether parametric or nonparametric statistical tests can be utilized for hypothesis testing [15][13].

Hypothesis testing: In this phase, we answer the RQs of our study by applying statistical tests. If the data we collected is normally distributed, we first use a t-test as parametric test [16] to evaluate our hypothesis. If the null hypothesis is rejected ($p < 0.05$), we then use Dunn's Test to perform pairwise comparisons between each independent group, and identify which groups are statistically significantly different. If our data is instead not normally distributed, we use Mann-Whitney nonparametric test to evaluate our hypothesis, and determine if statistically significant differences between groups exist.

Effect size estimation: We use Cohen's d (in case of normality) or Cliff's delta statistic (in case of non-normality) to measure effect size and quantify differences between observation groups beyond p-values interpretation [17]. In other words, we use effect size estimation to assess the magnitude of the experimental effect of our independent variables on the measured energy consumption.

5 Experiment Execution

5.1 Preliminary Experiment

For the purpose of correctness and feasibility of our experiment, a pre-testing phase is performed. In this preliminary phase, the energy consumption of each application is compared for different video conference lengths, in order to verify if the energy consumption grows linearly with time. A linear correlation allows us to reduce the time of each run, therefore making it possible to increase the total number of runs. Table 2 shows the treatments for each trial used in the preliminary experiment, executed for both considered apps. Each treatment is applied 20 times, in order to mitigate potential threats to conclusion validity (see also Section 8). In case our hypothesis of linearity is confirmed, we perform 20 runs of 3 minutes, otherwise the number of runs is reduced to 10, with a length of 10 minutes per run. By considering the total number of runs, treatments, apps, and cooldown time, the total experiment duration is estimated to be between 32 and 44 hours.

Table 2: Preliminary experiment

Webcam	Background	Mic	Participants	Time
on	on	on	2	10 min
on	on	on	2	9 min
on	on	on	2	7 min
on	on	on	2	5 min
on	on	on	2	3 min

Table 3: Pixel 3 – technical specifications

CPU	2.8 GHz Qualcomm Kryo 385 Octa Core
GPU	Adreno 630
Memory	4 GB
Disk space	64 GB
Battery capacity	2915 mAh Lithium-Ion
Screen	5.5 inch, 2160x1080, Amoled
OS	Android 9.0

5.2 Experiment Setup

In our experiment we are testing the mobile videoconferencing apps Google Meet (version 2021.10.17.404394895.Release) and Zoom (version 5.6.6). The two apps are both installed from the Google Play Store, and are executed on a Google Pixel 3 smartphone, whose technical specifications are listed in Table 3.

To automatically execute our experiment, we adopt the experimental setup depicted in Figure 1, and further described below. The

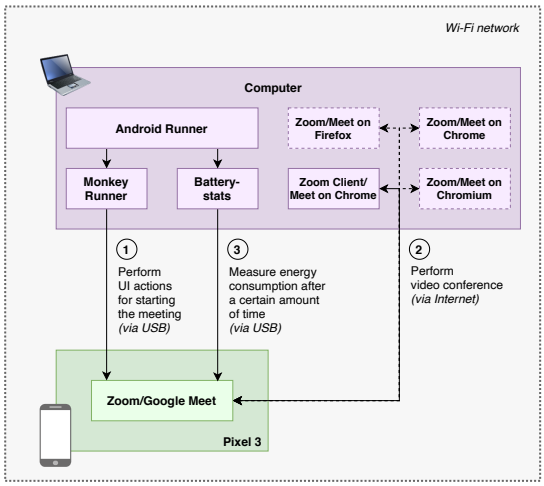


Figure 1: Structure of the experiment

experiment is run on a single computer. Care is taken to ensure the computer is powerful enough to run 5 instances of the videoconferencing application, as all the participants are joining the conference from this device. The computer is also used to control the smartphone. Specifications of the computer used are documented in (Table 4). The computer and smartphone are connected via USB and communicate via the Android Debug Bridge (ADB).[4]

The Android Runner framework [18], is used to configure and run the experiment. Specifically, such framework is used, among

[4]https://developer.android.com/studio/command-line/adb

Table 4: Asus A15 – technical specifications

CPU	AMD Ryzen™ 7 4800H (8C/16T, 12MB Cache, 4.2 GHz)
GPU	NVIDIA® GeForce RTX™ 2060, 6GB GDDR6
Memory	16GB DDR4-3200 SO-DIMM
Disk space	256GB M.2 NVMe™ PCIe® 3.0 SSD
Camera	720P HD camera
Screen	15.6-inch, FHD (1920 x 1080) 16:9 Anti-Glare
OS	Ubuntu 18.04.6 Desktop (64-bit)

others, to specify *via* a configuration file the apps to be tested, the execution/waiting times, and the plugins to be used. Batterystats [18] is one of such plugins, which is utilized to measure the amount of energy consumed in our experiment. One should be aware that Batterystats measurements are software-based, and therefore are not as precise as hardware-based measurements. Further consideration on this threat are discussed in Section 8.

To implement a fully-automatic and replicable execution of our experiment, we use MonkeyRunner[5], a record-replay tool allowing to first record sequences of user interface actions, and then replay the actions in every run of the experiment.

The video conferences the mobile device joins are set up manually with the camera feed being spoofed by a script which runs a video in a loop[6]. Thus the experiment is divided up into 4 parts, as we need to set up a conference with 2 and 5 participants for both Zoom and Google Meet apps. The tests for a conference are run by executing a Python script that starts one instance of Android Runner for each of our experimental configuration files sequentially. One Android Runner instance is configured to do 20 runs with a treatment for 3 minutes, with 1 minute of cool down time in between each run. One run starts by launching the app, and subsequently using an after-launch script hook to join the conference and applying the settings corresponding to the current treatment. Then the Batterystats profiler starts measuring the energy consumption for 3 minutes, after which Android Runner shuts down the app, and finally enters the 1 minutes cool down period.

6 Results

6.1 Preliminary Experiment

As described in Section 5.1, we conduct a preliminary experiment to assess if the energy consumption increases linearly over time, allowing us to potentially consider an execution time shorter than usual video conferences. Figure 2 depicts the energy consumption of video conferences with different duration. The slight duration fluctuations visible for each duration level are attributable to the time required to trigger each individual run. As clearly visible, the consumed energy grows linearly over time. This becomes even more evident when dividing the energy consumption by duration, which gives us the mean power consumption. From the data collected, we observe that the mean power consumption is always around 300 to 350 Joules per minute, independently of the meeting duration considered. These observations hold both for the Zoom and Meet apps. Therefore, we set the duration of a video conference to 3 minutes, enabling us to execute 20 runs per treatment, instead of only 10 runs.

[5]https://developer.android.com/studio/test/monkeyrunner
[6]https://github.com/whokilleddb/Fake-Stream

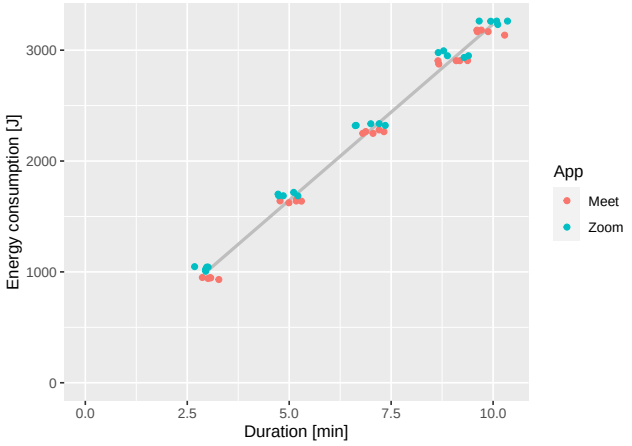


Figure 2: Preliminary experiment: Energy consumption over time

6.2 Data Exploration

Table 5 shows the summary statistic of the measured energy consumption, grouped by the two apps, for the whole dataset. Zoom results to have a slightly lower energy consumption compared to Meet in all shown statistical characteristics. The mean energy consumption of Meet (770.5 J) is 4% higher than the one of Zoom (738.2 J), while the median energy consumption is 3% higher (995.7 J for Meet and 970.5 J for Zoom).

Table 5: Energy consumption per app [J]

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Meet	245.2	282.5	995.7	770.5	1014.8	1109.5
Zoom	155.9	259.9	970.5	738.2	987.2	1041.4
Combined	155.9	267.6	977.8	754.3	1003.5	1109.5

Figure 3 shows the distribution of the energy consumption values for Meet and Zoom as histograms. In both figures two distinct clusters of particularly frequent values can be identified, with a large area without any values between approximately 300 J and 900 J. Such trend is further discussed when considering the results for each sub-RQ reported from Section 6.2.1 to Section 6.2.4.

Table 6 and 7 show the statistical characteristics for Meet and Zoom respectively, grouped by the different treatments for the factors participants, camera, virtual background, and microphone. Table 8 instead shows the relative change of the median when changing one of these factors. In the following subsections the impact of the different treatments are discussed in more detail.

6.2.1 Number of Participants. We conducted the experiment with either 2 or 5 conference participants. Figure 4 shows the distribution of the energy consumption in a combined violin- and boxplot grouped by the number of participants and the used app. All groups have in common that the distribution has two distinct peaks, as further discussed in Section 6.2.2. This matches the observation of the whole dataset, whose values could also be grouped into two clusters (cf. Figure 3).

For Meet, the energy consumption is higher when having five participants as compared to two. The mean increases from 750.7 Joules

Table 6: Energy consumption using Meet [J]

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
2 participants	245.2	265.6	991.8	750.7	997.8	1022.6
5 participants	252.3	289.2	1014.8	790.2	1068.5	1109.5
Camera off	245.2	264.1	267.2	272.6	281.7	292.3
Camera on	968.3	995.7	1003.0	1019.4	1028.9	1109.5
Virt. backg. off*	970.7	992.5	1002.5	1002.5	1014.8	1021.0
Virt. backg. on*	968.3	997.9	1035.1	1036.2	1074.8	1109.5
Microphone off	245.2	289.2	997.7	771.0	1013.4	1082.1
Microphone on	247.4	279.8	992.9	769.9	1015.2	1109.5
Combined	245.2	282.5	995.7	770.5	1014.8	1109.5

Table 7: Energy consumption using Zoom [J]

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
2 participants	242.7	254.2	987.2	752.8	1010.7	1041.4
5 participants	155.9	265.0	953.0	723.6	973.5	985.0
Camera off	242.7	248.9	252.5	254.5	261.1	269.6
Camera on	155.9	970.7	979.4	980.1	1004.9	1041.4
Virt. backg. off*	155.9	956.8	975.5	972.8	1008.6	1041.4
Virt. backg. on*	943.5	971.8	984.2	987.4	999.9	1028.9
Microphone off	242.7	251.0	955.3	731.6	975.5	1011.9
Microphone on	155.9	265.0	978.6	744.9	1010.1	1041.4
Combined	155.9	259.9	970.5	738.2	987.2	1041.4

*Since a virtual background can only be used when the camera is on, only runs where the camera is turned on are considered here

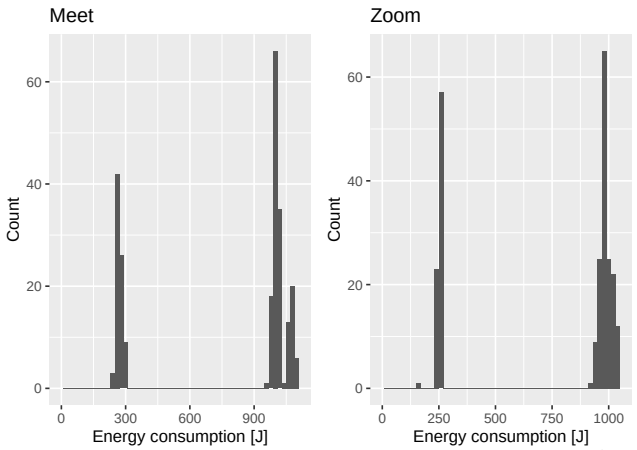


Figure 3: Histogram: Energy consumption per app (bin width: 20 J)

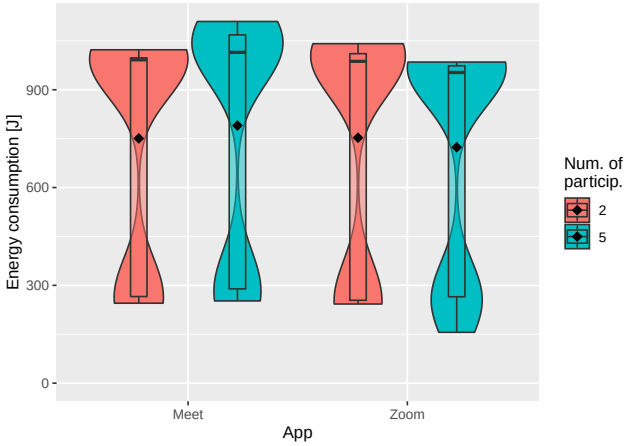


Figure 4: Violin- & boxplot: Energy consumption depending on the number of participants

Table 8: Relative increase of the mean energy consumption when changing one factor

	Meet	Zoom
2 ⇝ 5 participants	+5%	−4%
Camera off ⇝ on	+274%	+285%
Virt. backg. off ⇝ on*	+3%	+2%
Microphone off ⇝ on	−0.1%	+2%

*Since a virtual background can only be used when the camera is on, only runs where the camera is turned on are considered here

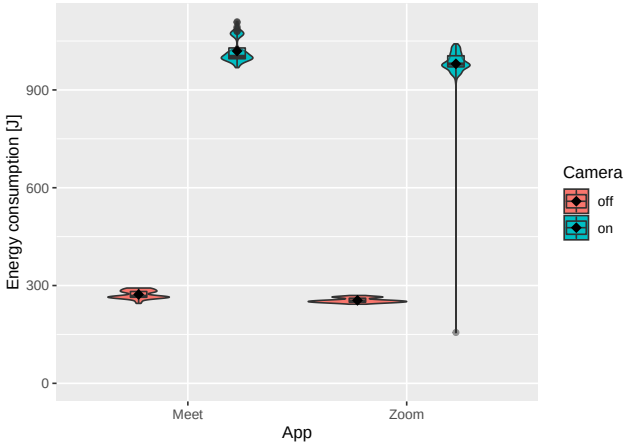


Figure 5: Violin- & boxplot: Energy consumption depending on whether the camera is turned on or off

to 790.2 Joules (+5%). For Zoom the mean energy consumption is surprisingly higher for two participants (752.8 J) when compared to five (723.6 J; −4%).

6.2.2 Camera. The violin- and boxplot for the camera, shown in Figure 5, clearly differs from the one for the number of participants. Here we have a large difference between the runs where the camera was active and the runs where it was not. For Meet the mean increases from 272.6 J to 1019.4 J (+274%), for Zoom the mean increases from 254.5 J to 980.1 J (+285%).

Although the mean energy consumption is clearly higher when the camera is on, the run with the lowest energy consumption in Zoom was a run with activated camera (155.9 J). But since the second lowest run with activated camera has a much higher energy consumption of 926 J, it can be assumed that the outlier data point is caused by some external factor influencing the experimental run; for the sake of completeness we decided to keep this data point in our dataset.

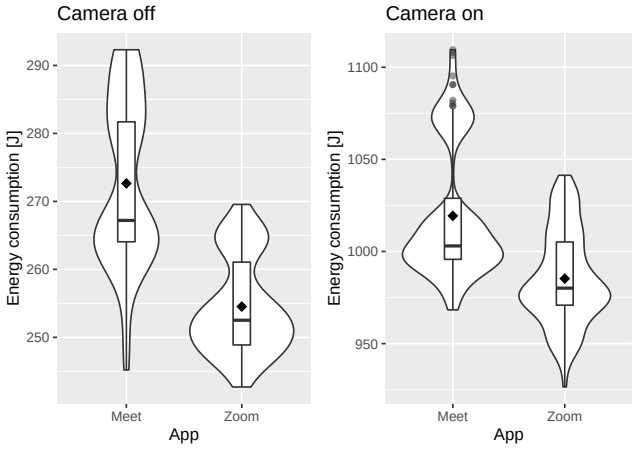


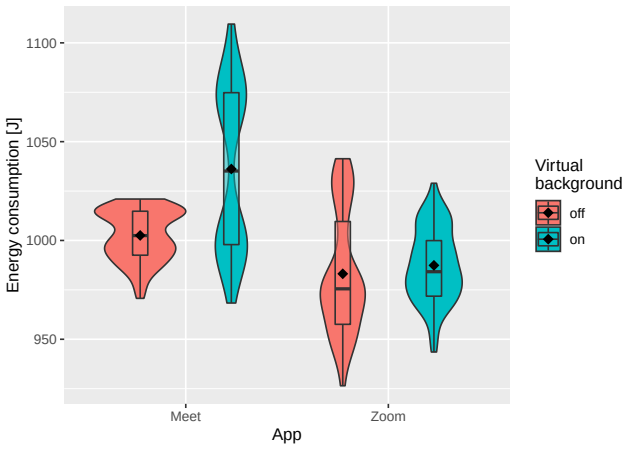
Figure 6: Violin- & boxplot: Energy consumption depending on whether the camera is turned on or off (zoomed in, outliers cut off)

Despite this outlier, the range of values for each boxplot is much smaller than in the previous boxplots, showcasing a much higher concentration towards median values. To further understand the distribution trends within each violin- and boxplot, Figure 6 depicts the same data, but with separated scales for camera off and on. As can be clearly noticed, except for Zoom with activated camera, all the plots display two distinct peaks due to the camera factor, which can be noticed also in the distributions of the data analysis conducted for the other factors.

6.2.3 Virtual Background. When the camera is active, it is also possible for users to replace their video stream background with a virtual background. It is of course possible to use a virtual background only if the camera is active. This is why we only consider the runs where the camera is active when assessing the impact of the virtual background feature.

Figure 7 shows the impact of the virtual background as a combined violin- and boxplot. Both apps have a small increase in energy consumption when activating the virtual background. For Meet, the mean energy consumption increases slightly from 1002.5 J when the virtual background is off to 1035.1 J when it is on (+3%). For Zoom, it increases even less noticeably, from 972.8 J to 987.4 J (+2%). As discussed in the previous subsection about the impact of the camera, we again observe an outlier data point for the Zoom app. As for the data on camera use, depicted in Figure 5, a single outlier measurement is present for Zoom without virtual background. For presentation purposes, such outlier is excluded from Figure 7 as, given the distribution of remaining data points, such measurement is with high probability due to an external factor.

*6.2.4 Microphone.*When considering the impact of the microphone, we get four nearly identical violin- and boxplots, as shown in Figure 8. For Meet the mean energy consumption slightly decreases when turning on the microphone from 771.0 J to 769.9 J (−0.1%), for Zoom it slightly increases from 731.6 J to 744.9 J (+2%).



*Since a virtual background can only be used when the camera is on, only runs where the camera is turned on are considered here

Figure 7: Violin- & boxplot: Energy consumption depending on whether the virtual background is turned on or off* (zoomed in, outliers cut off)

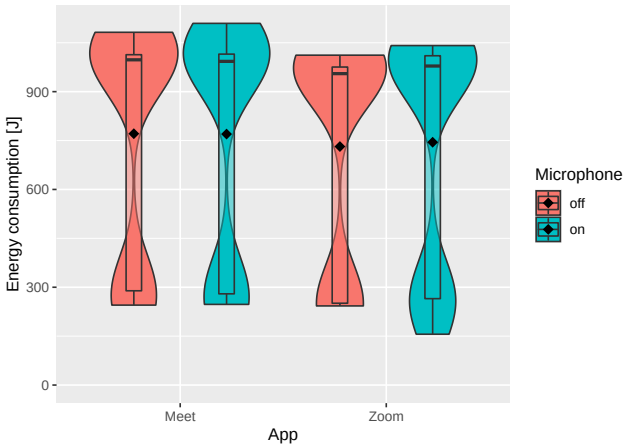


Figure 8: Violin- & boxplot: Energy consumption depending on whether the microphone is turned on or off

6.3 Normality Testing

From the data exploration section, we have a strong suspicion that the data is not normally distributed. Figure 9 shows the density and Q-Q plots of the Google Meet and Zoom data. Both density plots do not show a classical bell curve, which is generally present in normally distributed data. The Q-Q plots do not show a diagonal straight line, which would have been visible in normally distributed data. Lastly, the Shapiro-Wilk test gives a p-value < 2.2e-16 for both datasets, indicating that both for Zoom and Google Meet the energy measurements are not normally distributed. With the results of these three statistical methods, we can confidently conclude that the datasets are not normally distributed.

Since the datasets are further split for hypothesis testing, we also test each sub-dataset for normality. Figure 10 shows the resulting

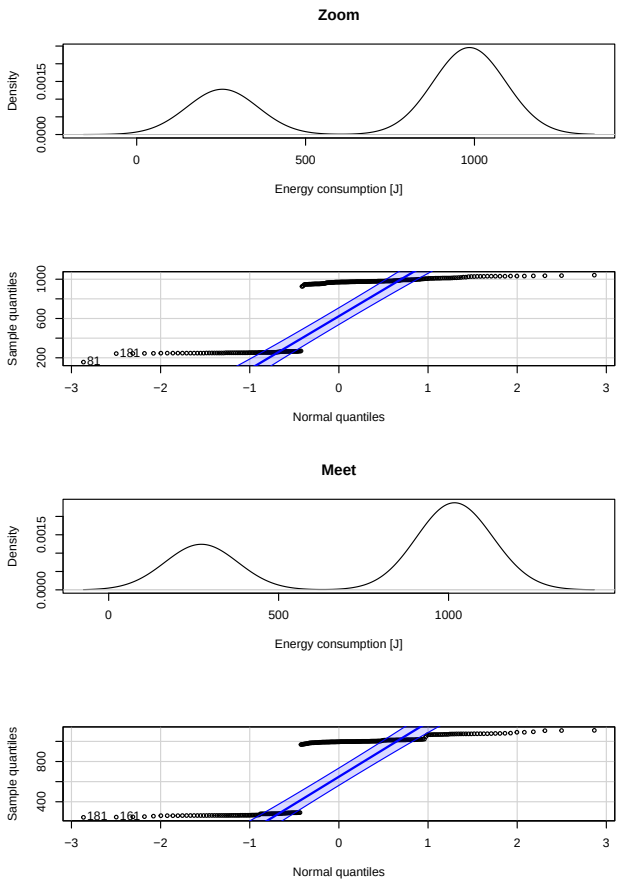


Figure 9: Density and Q-Q plots of the Zoom and Google Meet datasets

density and Q-Q plots and Table 9 documents the sub-dataset categories with the resulting Shapiro-Wilk test p-values. It is evident for most sub-datasets that they are not normally distributed. As exception, the sub-dataset of Zoom with the virtual background on is the only sub-dataset that, from a visual inspection followed by a Shapiro-Wilk test, results to be normally distributed. To ease the hypothesis testing and effect size analysis, we utilize exclusively non-parametric tests, namely the Mann-Whitney test and Cliff's delta (cf. Section 4.5), as we can consistently use such tests throughout all gathered sub-datasets.

Table 9: Shapiro-Wilk test p-values for sub datasets

	Meet	Zoom
2 participants	2.78e-16	8.01e-16
5 participants	2.75e-15	7.18e-16
Camera off	1.75e-05	1.22e-05
Camera on	2.912e-12	< 2.2e-16
Virt. backg. off	< 2.2e-16	< 2.2e-16
Virt. backg. on	3.9e-08	0.18
Microphone off	3.01e-15	7.20e-16
Microphone on	4.74e-15	1.90e-15

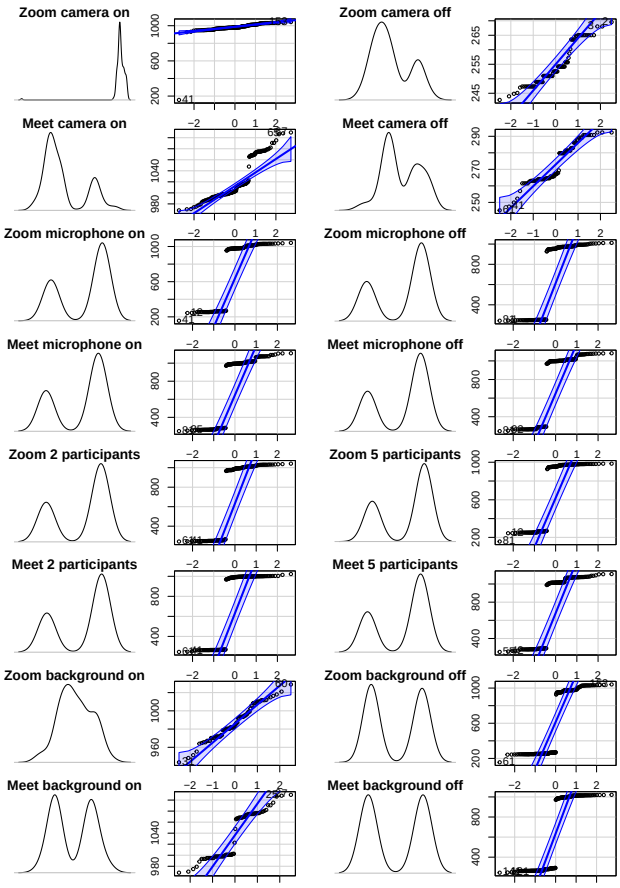


Figure 10: Density and Q-Q plots of the sub datasets

6.4 Hypothesis Testing.

After exploring the data and testing for normality, we can answer our RQs by statistically testing the null hypotheses defined in Section 4.3. In all null hypotheses we are considering one factor with exactly two treatments. Since nearly all of the distributions are not following a normal distribution, as shown in Section 6.3, we use the Mann-Whitney test to statistically test our hypotheses.

In Table 10, we report the p-values of the Mann-Whitney tests, which we apply to evaluate our null hypotheses. In addition, we also document the p-values corrected *via* the Benjamini-Hochberg procedure, in order to further mitigate potential type I errors.

Except for the microphone hypothesis for Meet ($H1.4_0$), all null hypotheses are confidently rejected. This means that a statistical significant difference is present between the energy consumption of Zoom and Meet, and also between the different configurations within both apps. As only exception, turning on the microphone when using Meet does not result to have a statistical significant impact on the energy consumption.

6.5 Effect Size Estimation

As documented in the previous section, most of the factors have a statistically significant impact on the energy consumption. To

Table 10: Statistical significance using the Mann-Whitney test and the Benjamini-Hochberg correction method (p-values < 0.05 marked in bold)

Hypothesis	Description	p-value	Corrected p-value
$H1_0$	Meet vs. Zoom	**6.905e-11**	**1.553625e-10**
Meet			
$H1.1_0$	Number of participants	**1.402e-11**	**4.206000e-11**
$H1.2_0$	Camera	**2.2e-16**	**9.900000e-16**
$H1.3_0$	Virtual background	**4.97e-05**	**6.390000e-05**
$H1.4_0$	Microphone	0.3577	0.3577
Zoom			
$H1.1_0$	Number of participants	**3.745e-06**	**6.741000e-06**
$H1.2_0$	Camera	**2.2e-16**	**9.900000e-16**
$H1.3_0$	Virtual background	**0.04005**	**0.04505625**
$H1.4_0$	Microphone	**4.518e-06**	**6.777000e-06**

determine the different impact on energy consumption of the factors we consider, we utilize Cliff's delta. This test returns a value between −1 and 1, whose absolute value is used to classify the effect size as either negligible (< 0.147), small (< 0.33), medium (< 0.47) or large (otherwise).

Table 11 shows the results of applying Cliff's delta on our different RQs. For the overall comparison between Meet and Zoom (RQ1) we get a value of −0.344 which can be classified according to the thresholds as a medium effect size. For Meet we see a large effect size for the number of participants (RQ1.1) and the camera (RQ1.2), and a medium effect size for the virtual background (RQ1.3). The effect size of the microphone (RQ1.4) would be negligible, but this is regardless not meaningful, as the difference is not statistical significant. For Zoom there is a large effect size for the camera (RQ1.2), a medium effect size for both the number of participants (RQ1.1) and the microphone (R1.4), and a small effect size for the virtual background (RQ1.3).

Table 11: Effect size estimation using Cliff's delta

Research Question	Description	δ-value	Classification
$RQ1$	Zoom vs. Meet	−0.344	Medium
Meet			
$RQ1.1$	Number of participants	0.505	Large
$RQ1.2$	Camera	1.000	Large
$RQ1.3$	Virtual background	0.372	Medium
$RQ1.4$	Microphone	−0.069	Negligible
Zoom			
$RQ1.1$	Number of participants	−0.345	Medium
$RQ1.2$	Camera	0.988	Large
$RQ1.3$	Virtual background	0.188	Small
$RQ1.4$	Microphone	0.343	Medium

7 Discussion

In this section, we revisit our research questions, and further discuss implications for the users of videoconferencing apps.

[RQ1]: *To what extent does the energy consumption differ between different videoconferencing apps?* We observe a statistically significant difference between the two apps considered, with a medium effect size. For a user who wants to save energy on their phone, Zoom would be the better choice. However, as the mean energy consumption of Zoom is only 4% lower compared to Meet, other factors such as usability or restrictions of the free versions should be taken into account as well.

[RQ1.1]: *What is the impact of different numbers of participants on the energy consumption of videoconferencing apps?* For both apps the number of participants has a statistically significant impact on the energy consumption, leading up to a 5% energy consumption increase. The effect size is large for Meet and medium for Zoom. An interesting observation is that, while the energy consumption in Meet increases with the number of participants, it decreases in Zoom. This could possibly be due to the fact that, by default, the Zoom app always displays exclusively the speaking person, while Meet displays the video streams of the participants as a grid.

[RQ1.2]: *What is the impact of using the camera on the energy consumption of videoconferencing apps?* Using the camera has by far the largest impact on energy consumption. With an active camera, the mean energy consumption is up to 285% higher than without camera. In addition, the statistical test demonstrate that for both apps camera use leads to statistically significant differences in energy consumption, with large effect size. Thus, when not required during a conference call, turning off the camera will lead to considerable savings in energy consumption for the call.

[RQ1.3]: *What is the impact of using virtual backgrounds on the energy consumption of videoconferencing apps?* Both apps exhibit a statistically significant increase of consumed energy when using a virtual background compared to using the original camera stream. Nevertheless, the mean energy consumption only increases by 3% for Google Meet and or 2% for Zoom. We conjecture that, albeit isolating the foreground image implies an additional computation task, the impact on the additional energy consumed by this tasks is mitigated by the need to stream less data. In addition, it is possible that virtual backgrounds might be implemented to benefit from hardware acceleration, potentially further reducing the overall impact on energy consumption of this feature.

[RQ1.4]: *What is the impact of using the microphone on the energy consumption of videoconferencing apps?* For Meet, we could not observe statistically significant evidence that microphone use influences energy consumption. As explanation of this finding, we conjecture that the microphone does not get physically deactivated, and instead the data is simply not transmitted to the other participants. However further investigation needs to be conducted in order to verify this claim. For Zoom the observations were statistically significant, and the effect size was classified as medium. With the microphone active the mean energy consumption increased by 2%, which is nevertheless a much lower increase with respect to the use of the camera.

8 Threats To Validity

The validity of the experiment has been analyzed based on four classification types, namely internal, external, construct, and conclusion validity, as proposed by Campbell and Cook [19].

8.1 Internal Validity

8.1.1 History. This threat is closely related to the experiment design and operation. A potential threat in this category is constituted by the temperature of the mobile device utilized for experimentation, which may have influenced the data of sequential runs. To mitigate this threat, we ensured that an appropriate cool down period, equal to 1 minute, was present between each subsequent run.

8.1.2 Maturation. A potential maturation threat is constituted by possible caching functionalities of the apps used as experimental subjects. In fact, data caching may have influenced how the apps reacted to our factors, as these were sequentially applied throughout our experiment. To mitigate this threat, we ensured the cache of the mobile device utilized for experimentation was cleared before each run.

8.1.3 Reliability of measures. Several factors could influence the reliability of measures, such as brightness of the mobile screen, notifications, distance to the router, and interference with other processes which consume more energy. To mitigate this threat, we ensured that throughout the entirety of the experiment notifications were turned off, brightness was set to the same intensity, the device was positioned at the same distance from the router, and only the functionalities required for the experiment were running on the mobile device.

8.2 External Validity

8.2.1 Interaction of selection and treatment. In order to mitigate potential threats to external validity, we adopted as experimental subjects two among the most popular Android videoconferencing apps, namely Zoom and Google Meet, which are utilized by millions of users every day. An additional threat to external validity is constituted by the mobile device utilized in our experiment. To mitigate this threat, we selected and used a popular mobile device, namely the Pixel 3, which is distributed by a prominent multinational technology company (Google), and reflects in terms of specification the hardware of commonly utilized mobile devices (cf. Table 3).

8.2.2 Interaction of setting and treatment. A threat in this category is constituted by the video conference duration, which was set to 3 minutes each, and hence might not be representative for real video conferences duration. To mitigate this threat, we performed a preliminary experiment to assess that the duration of the experiment runs do not influence significantly our findings (cf. Section 6.1). In addition, changes in internet bandwidth could have influenced our results. We deem this threat as minor, as the experimental environment relied on a dependable internet connection, and no devices other than the ones used for the experiment were running.

8.3 Construct Validity

8.3.1 Definition of constructs. In order to mitigate potential threats of this category, in Section 4 we reported an in-depth description of our experimental setting, including a documentation of the design phase and the related rationale, and description of all metrics, factors, and measurement tools utilized. In order to mitigate threats related to insufficient interpretation, we formulated systematic hypotheses and scenarios to answer our RQs, and tested our hypothesis by utilizing exclusively fitting statistical tests.

8.3.2 Mono-method bias. Our experimental results are based on one dependent variable, namely energy consumption. Thus, our experiment might be affected by a mono-operation bias. Per se, we do not deem utilizing only energy consumption measurements a notable threat, and utilizing such single metric is a common practice in software energy efficiency research [20–22]. However, our measurements relied on energy estimations, rather than direct energy measurements, as provided by the tool Batterystats we used. To mitigate potential threats related to the adoption of this measurement method, we ensured that the tool was peer-reviewed [18], based on a sound theoretical foundation [23], provided as open source (hence allowing for independent scrutiny), and previously used in other academic peer-reviewed studies (e.g., in the work of Malavolta et al. [24]).

8.4 Conclusion Validity

8.4.1 Low statistical power. In order to mitigate potential threats of low statistical power, we ensured to systematically collect the required volume of data, and conducted statistical analysis to draw conclusions. More specifically, we collected the experimental data by using a fixed number of 12 treatment combinations for our experiments, conducted by considering 2 experimental subjects. Each treatment was applied 20 different times for both applications, resulting in a total of 480 data samples which were collected for this study.

8.4.2 Violated assumptions of statistical tests. To mitigate potential threats of violated test assumptions, before performing our statistical analysis, we tested the normal distribution of our collected sample. This process was conducted to select the most appropriate statistical test from a list of tests systematically defined *a priori*.

8.4.3 Treatment implementation. All treatments were implemented as fitful as possible w.r.t. a real-world videoconference scenario. The only slight difference is constituted by the use of a script to generate the incoming camera videos. We do not deem this as a prominent threat, as this variation may have influenced only the mocked users devices, and not the device receiving the videos (i.e., where the energy consumption was measured).

9 Conclusions

In this research, we document an empirical experiment conducted to study the energy consumption of Android videoconferincing apps. In addition to investigating the differences between two popular videoconferencing apps, we study if distinct features, such as use of web camera, microphone, and virtual background, significantly influence the energy consumption of the apps. From our findings emerge that, both when considering different apps and features, the energy consumed by mobile devices drastically changes. Camera use has the larges impact on energy, with an increase up to 285% energy consumed. In contrast, microphone use does not influence considerably the energy consumption, with only a negligible 2% energy consumption increase for one of the two apps considered.

As future work, we plan to replicate our experiment in order to further mitigate some of the threats to validity of our study, e.g., by considering different mobile devices and more videoconferencing apps. In addition, we plan to broaden the scope of our research, by considering additional factors related to videoconferencing apps that could influence energy consumption. For instance, we envision to test the impact on energy consumption of turning off the screen when only audio is needed, using different connection types, e.g., 4G/5G, and utilizing the chat available in videoconferencing apps during the meetings. Finally, we envision to investigate potential tradeoffs between energy consumption and other non-functional properties of the apps (*e.g.,* security, usability, performance).

References

[1] R. Verdecchia, P. Lago, C. Ebert, and C. De Vries, "Green IT and Green Software," *IEEE Software*, vol. 38, no. 6, pp. 7–15, 2021.

[2] C. N. Tejinder K. Judge, "Sharing Conversation and Sharing Life:Video Conferencing in the Home," 2010.

[3] N. P. Ella Koeze, "The Virus Changed the Way We Internet," 2020.

[4] S. Chan, "Usage of Mobile Video Conferencing Apps Including Zoom Grew 150 in the First Half of 2021," 2020.

[5] H. Esmaeilzadeh, E. Blem, R. S. Amant, K. Sankaralingam, and D. Burger, "Dark silicon and the end of multicore scaling," in *2011 38th Annual international symposium on computer architecture (ISCA)*. IEEE, 2011, pp. 365–376.

[6] X. Han, L. Lu, Y. Zheng, X. Feng, Z. Li, J. Li, and M. Ouyang, "A review on the key issues of the lithium ion battery degradation among the whole life cycle," *ETransportation*, vol. 1, p. 100005, 2019.

[7] D. Ong, T. Moors, and V. Sivaraman, "Complete life-cycle assessment of the energy/co2 costs of videoconferencing vs face-to-face meetings," in *2012 IEEE Online Conference on Green Communications (GreenCom)*. IEEE, 2012, pp. 50–55.

[8] ——, "Comparison of the energy, carbon and time costs of videoconferencing and in-person meetings," *Computer Communications*, vol. 50, pp. 86–94, 2014.

[9] A. Maiti and G. Challen, "The missing numerator: Toward a value measure for smartphone apps," in *Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications*, 2015, pp. 99–104.

[10] G. Faber, "A framework to estimate emissions from virtual conferences," *International Journal of Environmental Studies*, pp. 1–16, 2021.

[11] R. Trestian, A.-N. Moldovan, O. Ormond, and G.-M. Muntean, "Energy consumption analysis of video streaming to android mobile devices," in *2012 IEEE Network Operations and Management Symposium*. IEEE, 2012, pp. 444–452.

[12] V. R. Basili, G. Caldiera, and D. Rombach, "The Goal Question Metric Approach," in *Encyclopedia of Software Engineering*. Wiley, 1994, pp. 528–532.

[13] C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell, and A. Wesslén, *Experimentation in software engineering*. Springer Science & Business Media, 2012.

[14] M. Hoogendoorn and B. Funk, "Machine learning for the quantified self," *On the art of learning from sensory data*, 2018.

[15] J. J. Faraway, *Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models*. CRC press, 2016.

[16] T. V. Hecke, "Power study of anova versus kruskal-wallis test," *Journal of Statistics and Management Systems*, vol. 15, no. 2-3, pp. 241–247, 2012.

[17] G. Macbeth, E. Razumiejczyk, and R. D. Ledesma, "Cliff's delta calculator: A non-parametric effect size program for two groups of observations," *Universitas Psychologica*, vol. 10, pp. 545–555, 2010.

[18] I. Malavolta, E. M. Grua, C.-Y. Lam, R. de Vries, F. Tan, E. Zielinski, M. Peters, and L. Kaandorp, "A Framework for the Automatic Execution of Measurement-based Experiments on Android Devices," in *35th IEEE/ACM International Conference on Automated Software Engineering Workshops (ASEW '20)*. ACM, 2020. [Online]. Available: https://github.com/S2-group/android-runner/blob/master/documentation/A_Mobile_2020.pdf

[19] D. T. Campbell and T. D. Cook, "Quasi-experimentation," *Chicago, IL: Rand McNally*, 1979.

[20] A. Carette, M. A. A. Younes, G. Hecht, N. Moha, and R. Rouvoy, "Investigating the energy impact of android smells," in *2017 IEEE 24th International Conference on Software Analysis, Evolution and Reengineering (SANER)*. IEEE, 2017, pp. 115–126.

[21] R. Verdecchia, G. Procaccianti, I. Malavolta, P. Lago, and J. Koedijk, "Estimating energy impact of software releases and deployment strategies: The kpmg case study," in *2017 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*. IEEE, 2017, pp. 257–266.

[22] A. Hindle, A. Wilson, K. Rasmussen, E. J. Barlow, J. C. Campbell, and S. Romansky, "Greenminer: A hardware based mining software repositories software energy consumption framework," in *Proceedings of the 11th working conference on mining software repositories*, 2014, pp. 12–21.

[23] D. Di Nucci, F. Palomba, A. Prota, A. Panichella, A. Zaidman, and A. De Lucia, "Petra: a software-based tool for estimating the energy profile of android applications," in *2017 IEEE/ACM 39th International Conference on Software Engineering Companion (ICSE-C)*. IEEE, 2017, pp. 3–6.

[24] I. Malavolta, K. Chinnappan, L. Jasmontas, S. Gupta, and K. A. K. Soltany, "Evaluating the impact of caching on the energy consumption and performance of progressive web apps," in *Proceedings of the IEEE/ACM 7th International Conference on Mobile Software Engineering and Systems*, 2020, pp. 109–119.