Analyzing the Intelligibility of Real-Time Mobile Sign Language Video Transmitted Below Recommended Standards

Jessica J. Tran[1], Ben Flowers[1], Eve A. Riskin[1], Richard E. Ladner[2], Jacob O. Wobbrock[3]

[1]Electrical Engineering
DUB Group
University of Washington
Seattle, WA 98195 USA
{jjtran, blow, riskin}@uw.edu

[2]Computer Science & Engineering
DUB Group
University of Washington
Seattle, WA 98195 USA
ladner@cs.washington.edu

[3]The Information School
DUB Group
University of Washington
Seattle, WA 98195 USA
wobbrock@uw.edu

ABSTRACT

Mobile sign language video communication has the potential to be more accessible and affordable if the current recommended video transmission standard of 25 frames per second at 100 kilobits per second (kbps) as prescribed in the International Telecommunication Standardization Sector (ITU-T) Q.26/16 were relaxed. To investigate sign language video intelligibility at lower settings, we conducted a laboratory study, where fluent ASL signers in pairs held real-time free-form conversations over an experimental smartphone app transmitting real-time video at 5 fps/25 kbps, 10 fps/50 kbps, 15 fps/75 kbps, and 30 fps/150 kbps, settings well below the ITU-T standard that save both bandwidth and battery life. The aim of the laboratory study was to investigate how fluent ASL signers adapt to the lower video transmission rates, and to identify a lower threshold at which intelligible real-time conversations could be held. We gathered both subjective and objective measures from participants and calculated battery life drain. As expected, reducing the frame rate/bit rate monotonically extended the battery life. We discovered all participants were successful in holding intelligible conversations across all frame rates/bit rates. Participants did perceive the lower quality of video transmitted at 5 fps/25 kbps and felt that they were signing more slowly to compensate; however, participants' rate of fingerspelling did not actually decrease. This and other findings support our recommendation that intelligible mobile sign language conversations can occur at frame rates as low as 10 fps/50 kbps while optimizing resource consumption, video intelligibility, and user preferences.

Categories and Subject Descriptors

K.4.2. [**Social Issues**]: Assistive technologies for persons with disabilities; H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems – Video.

General Terms

Performance; Experimentation; Human Factors.

Keywords

Intelligibility; comprehension; American Sign Language; bit rate; frame rate; video compression; laboratory study; Deaf community.

*ASSETS'*14, October 20–22, 2014, Rochester, NY, USA.
Copyright © 2014 ACM 978-1-4503-2720-6/14/10…$15.00
http://dx.doi.org/10.1145/2661334.2661358

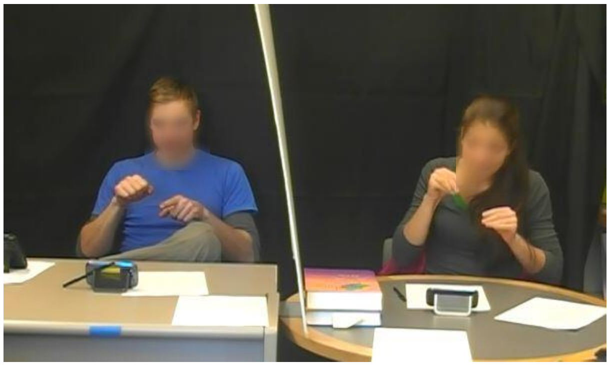

Figure 1: Two study participants holding an intelligible sign language conversation over an experimental smartphone application transmitting video at frame rates and bit rates well below industry recommended rates.

1. INTRODUCTION

Smartphones are rapidly changing the way people communicate and receive information, with over 1.9 billion smartphone users worldwide at the end of 2013 [34]. The growth of smartphone users has led to video being the fastest growing contributor to mobile data traffic [34]. Streaming video providers like YouTube, Hulu, and Netflix contribute to mobile video traffic by consuming 51% of all network traffic. Mobile video telephony is also contributing to the acceleration of video data consumption with the numerous available mobile video chat applications like Skype, Facetime, and Google Hangouts. In 2010, Skype received 7 million downloads onto Apple's iPhone alone [25]. Figure 1 is an example of two people signing over a mobile device.

Often, high fidelity video quality with little-to-no delay is a priority for mobile video telephony; however, this performance usually comes at the cost of high bandwidth consumption. Apple's Facetime app provides high quality video over Wi-Fi or cellular networks with an average bandwidth consumption of 5 MB of data per minute [11]. The high data rate cost of using FaceTime over limited data plans can quickly become expensive [6]. Other mobile video chat apps, like Skype, transmit video at lower dynamic transmission rates ranging from 40-450 kilobits per second (kbps) depending on network traffic [10]. Video intelligibility is sacrificed when relying on the available network bandwidth to regulate video transmission rates.

Deaf and hard-of-hearing people benefit from advancements in mobile video communication because it facilitates sign language communication. American Sign Language (ASL) is a visual language with its own grammar and syntax unique from any spoken languages. Intelligible video content is required for successful sign language conversations; therefore, the International Telecommunication Standardization Sector (ITU-T) Q.26/16 recommends at least 25 frames per second (fps) and 100 kbps for sign language video transmission [24]. However, total network bandwidth is limited and network congestion can lead to unintelligible content due to delayed and dropped video. U.S. cellular networks do not provide unlimited data plans and may throttle back network speeds for high data rate consumers [33].

We conducted a laboratory study in which pairs of fluent ASL signers held free-form conversations over an experimental smartphone app transmitting real-time video at 5 fps/25 kbps, 10 fps/50 kbps, 15 fps/75 kbps, and 30 fps/150 kbps, well below the ITU-T standard, for the purpose of saving bandwidth and battery life. The objectives of this study were: (1) to identify the minimum video quality settings allowable for intelligible sign language communication; (2) to learn what adaptation techniques participants use to compensate for the lowered transmission rates; (3) to objectively measure user perceived intelligibility of video content used in mobile sign language conversations; and (4) to quantify how much battery life is extended. We gathered both subjective and objective measures from participants and measured battery life drain. As expected, reducing the frame rate/bit rate monotonically extended the battery life. Video transmitted at 5 fps/25 kbps averaged 264 minutes of battery life, while video at 30 fps/150 kbps averaged 209 minutes of battery life. Subjective results revealed video transmitting at 5 fps/25 kbps had the most negative impact on perceived video quality ($\chi^2_{3,N=20}$=11.01, *p*<.05), fingerspelling ($\chi^2_{3,N=19}$=8.11, *p*<.05), and how often a participant needed to guess what the other signer was signing ($\chi^2_{3,N=20}$=29.75, *p*<.0001). However, frame rate/bit rate was not found to significantly impact perceived video intelligibility ($\chi^2_{3,N=20}$=5.08, *n.s.*).

Participants were successful in holding intelligible conversations across all frame rates/bit rates. All participants did perceive the lower quality of video transmitted at 5 fps/25 kbps and perceived they were signing more slowly to compensate; however, participants' rate of fingerspelling did not significantly decrease. Exit interviews revealed four recurring themes when it came to signing on mobile devices: (1) there was noticeable lower quality of video transmitted at 5 fps/25 kbps; (2) desire for larger screens; (3) different adaptation techniques were used to compensate for lower video quality; and (4) comparison of video quality used in the experimental app to commercially available apps. These and other findings compel our recommendation that mobile video software used by deaf people should support frame rates as low as 10 fps /50 kbps.

2. RELATED WORK

2.1 Bandwidth Requirements

The bandwidth requirements for transmission of sign language have been under consideration since the early 1990s. Sperling [28] investigated the ability for deaf people to transcribe ASL and fingerspelling from reduced television displays at bandwidths of 86 kHz, 21 kHz, 4.4 kHz, and 1.1 kHz. Intelligibility was found to drop to 90% at 21 kHz and to 10% at 4.4 kHz. Fingerspelling intelligibility was found to be more sensitive to bandwidth reduction, with intelligibility dropping to only 70% at 21 kHz.

Sosnowski and Hsing [27] evaluated moving images, finding that reducing the frame rate from 30 to 15 fps only produced slightly less intelligible video; however, video displayed below 15 fps resulted in intelligibility dropping dramatically. Harkins *et al*. [14] compared the outline of signers to a videotaped control, which consisted of the video transmitted at the original recording rate and found that video shown below 10 fps resulted in poor intelligibility. Ultimately, these prior works suggest that frame rates between 15-30 fps are the recommended rates at which video should be transmitted to maintain intelligibility. Our work will demonstrate that intelligible sign language conversations can occur below 15 fps.

Manoranjan and Robinson [19] investigated a method to reduce bandwidth consumption by transmitting binary sketches of cartoon signers. They implemented their video processing technique on a computer that simulated the bandwidth used over telephone lines. In a laboratory study with two total participants, participant 1 signed a sentence and participant 2 wrote down what he viewed. Participants evaluated four picture sizes of video displayed at 80×60, 160×120, 120×160, and 320×240 pixels/frame with video transmitted at 8 fps. The computer simulated transmission rates at 33.5 kbps for phone lines and 100 Mbps for the LAN data rate. Participants were unable to complete the task at 320×240 pixels/frame because of the low number of bits allocated per pixel. At such a low frame rate, participants preferred to view the binary sketches of the signer at the 80×60 pixels/frame resolution. A major limitation of this prior work was the small sample size of 2 total participants, which made results hard to generalize to mobile video communication. Our laboratory study uses up-to-date technology with more participants to produce more generalizable recommendations for mobile video communication.

2.2 Prior Laboratory Studies

2.2.1 MobileASL Project

MobileASL, an experimental smartphone application running on the Windows Mobile 6 platform, was created in 2008 and provides two-way, real-time sign language video at very low bandwidth: 30 kilobits per second at 8-12 frames per second. Prior research evaluated intelligibility of pre-recorded ASL video and reducing the power consumption of MobileASL through various techniques.

Cavender *et al.* [5] conducted a laboratory study evaluating perceived video intelligibility of pre-recorded ASL videos transmitted at two frame rates (10 and 15 fps), three bit rates (15, 20, and 25 kbps), and three region-of-interest (ROI) encoding levels (0, -6, and -12 ROI). They discovered a frame rate preference of 10 fps for viewing ASL video at a fixed bit rate.

Cherniavsky *et al*. [9] conducted a laboratory study where participants in pairs were observed signing over MobileASL with an algorithm that lowered the frame rate to 1 fps during not-signing sections of a conversation. They found that applying that algorithm led to degradation in video quality, which resulted in respondents having to guess more frequently during conversations. Overall, participants expressed that having the power saving algorithm applied during their conversations did not deter their potential adoption of MobileASL for mainstream mobile video communication.

These previous studies demonstrate the potential lower limits in which intelligible mobile sign language video communication can occur. Our new laboratory study is different from prior work because we investigate intelligibility of *real-time* conversations held over smartphones with video transmitted at higher frame rates and bit rates than were explored in prior work.

2.2.2 Sign Language Learning and Comprehension

Sign language learning is more nuanced than holding sign language conversations because linguistic accuracy is most important. Therefore, the effect of frame rate reduction on sign language learning has been extensively researched [7, 16, 18, 29]. Johnson and Caird [18] investigated whether perceptual ASL learning was affected by video transmitted at 1, 5, 15, and 30 fps. In a discrimination task, participants made a *yes-no* decision about whether the displayed sign and the English word shown matched. They found that frame rates as low as 1 fps and 5 fps were sufficient for novice ASL learners to recognize learned ASL gestures. Although this work demonstrates the potential lower limits at which video can be transmitted, this work did not evaluate conversational sign language, which we evaluate in our laboratory study.

Sperling *et al.* [29] investigated sign recognition when ASL video was transmitted at 10, 15, and 30 fps displayed at 96×64, 48×32, and 24×16 spatial resolutions. They found that common isolated ASL signs shown at 96×64 pixels at 15 fps and 30 fps did not have a noticeable effect on intelligibility, but signs at 10 fps did. While prior work showed that lower frame rates can impact isolated sign recognition, these results may not hold true for mobile sign language video conversations. Our work goes beyond sign recognition and investigates video intelligibility to support two-way conversations.

3. Laboratory Study

Up until now, we have conducted web-based studies [30, 31] evaluating perceived video intelligibility of pre-recorded conversational sign language videos transmitted at frame rates, bit rates, and spatial resolutions lower than the recommended ITU-T standard. Findings from our prior work have suggested an “intelligibility ceiling effect” [32], where increasing the frame rate above 10 fps and bit rate above 60 kbps does not significantly improve perceived video intelligibility.

In a continued effort to reduce total bandwidth consumption and extend battery life for mobile sign language video telephony, we conducted a laboratory study, where fluent ASL signers in pairs held free-form conversations over an experimental smartphone app transmitting real-time video at 5 fps/25 kbps, 10 fps/50 kbps, 15 fps/75 kbps, and 30 fps/150 kbps. The objectives of this study were: (1) to identify the minimum video quality settings allowable for intelligible sign language communication; (2) to learn what adaptation techniques participants use to compensate for the lowered transmission rates; (3) to objectively measure user perceived intelligibility of video content used in mobile sign language conversations; and (4) to quantify how much battery life is extended. Results from the laboratory study also demonstrate that intelligible conversations can occur at transmission rates lower than the ITU-T standard.

3.1 Technology Used

3.1.1 Mobile Phone

The Samsung Galaxy S3 smartphone was used to run an open source video chat software app called IMSDroid[1], whose encoder was modified to transmit video at 5, 10, 15, and 30 fps. The bit rate averaged 5 kb/frame, resulting in the bit rate increasing as the frame rate increased, namely 25, 50, 75, and 150 kbps, respectively. The spatial resolution of the video transmitted was held constant at 320×240 pixels and displayed horizontally on the phone to maximize the screen size. Prior to the selection of the Samsung Galaxy S3 phone, the Sprint EVO, Samsung Galaxy S2, Samsung Galaxy S4, HTC One, and Google Nexus Phone 4 were investigated as alternatives, but each of these phones’ encoders failed to allow for the lowered frame rates. Only the Samsung Galaxy S3 encoder was compatible with the IMSDroid frame rate modifications and thus, the Galaxy S3 was selected for the laboratory study.

3.1.2 IMSDroid

IMSDroid is an open source video conferencing application running on Doubango [17], a 3GPP IMS/LTE (IP Multimedia Subsystem) framework for embedded systems. IMSDroid is a Java-based front-end to Doubango, which is open source VoIP client that references implementation to the Doubango framework. IMSDroid has a GUI interface allowing for both audio and video calls with the robustness of selecting different video encoder. Doubango is the backend framework running 3GPP IMS/LTE which can run many different types of protocols like SIP/SDP, HTTP/HTTPS, and DNS. In this study, the Session Initiation Protocol (SIP) was selected for the VoIP.

3.1.3 Asterisk Server

An Asterisk [2] server was set up as the communication server for the laboratory study. Asterisk is an open source framework that supports the server side of facilitating Voice over Internet Protocol (VoIP) video communication, where we used the Session Initiation Protocol. A specific configuration file was modified to regulate the bit rate at which video was transmitted, specifically averaging 5 kb/frame. Asterisk uses User Datagram Protocol, which is suitable for fast efficient transmission of data for video conversations.

3.1.4 Unobtrusive Logging

Network traces were conducted on the Asterisk server monitoring the frame rate and bit rate at which video was transmitted for each video call. The battery drain of each phone was also unobtrusively logged on the mobile device using an open source mobile application called AndroSensor [1]. AndroSensor logged the battery life percentage every 30 seconds.

3.2 Participants

Social media and email listservs were used to recruit fluent ASL signers to participate in the study. Participant inclusion criteria included: (1) deaf and/or hard-of-hearing people for whom ASL is the primary language; (2) hearing people who fluently sign ASL (over 5 years of signing experience); and (3) people 18 years old or older. Participants received a $25 gift card upon completing the 75-minute laboratory study. Those who responded to the e-mail were either paired with a random person to sign with or brought a friend fluent in ASL. Demographic questions asked in the laboratory study (described below) were used to further ensure language fluency.

The laboratory study had 20 participants (11 women), all of whom fluently signed ASL. Their age ranged from 26-74 years old (median=48.5 years, SD=13.5 years). Of the 20 participants, 18 were deaf (2 of 18 wore hearing aids) and 2 were Children of Deaf Adults with full hearing. Eight participants were randomly assigned to their signing partner (4 sessions) and the other participants were paired with a friend (6 sessions). Thirteen participants indicated that ASL was their daily language, and the number of years they had spoken ASL ranged from 26-74 years (mean=47 years, SD=13 years). All but one participant owned a smartphone and everyone had sent text messages; 19 participants indicated they use video chat; and 17 use video relay services.

[1] http://doubango.org/. Accessed on May 9, 2012.

3.3 Study Design

3.3.1 *Apparatus*

Participants sat on the same side of a table with a black drape behind them. They were separated by a board. Two phones were propped up with a business card holder and placed, one each in front of the participants. Participants were told to adjust the location of the phone for comfortable conversation. Figure 2 is a photo of the experimental setup.

3.3.2 *Conversation Task*

Participants were instructed to hold five, 5-minute free-form conversations over the provided smartphones. The first conversation was a practice round for participants to familiarize themselves with the phone and available signing space. Participants were instructed to talk about whatever they liked, but for each subsequent conversation, they were asked to discuss a different topic than the conversation before. After each session, participants filled out a paper questionnaire, described below. All participants were video recorded during the study. The smartphone did not record conversations. A randomized Latin Square was used to assign the order in which video frame rate was used on IMSDroid. Participants were not told how the video quality was altered, only that they were using different versions of the smartphone app. A certified ASL interpreter was present during all study sessions and facilitated communication between the study participants and the first author, who conducted the studies.

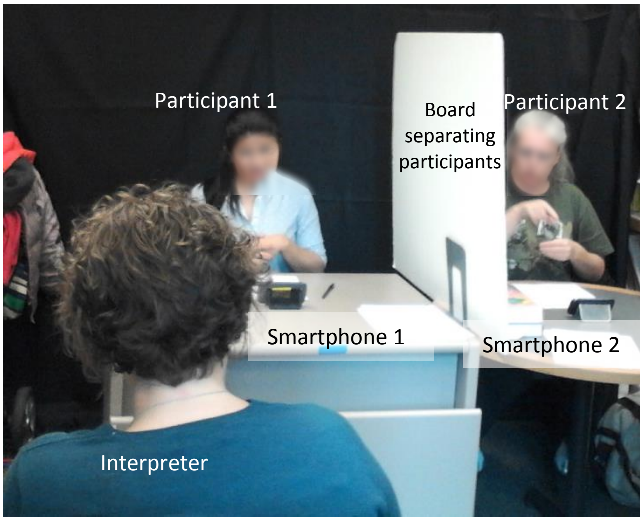

Figure 2: Experimental setup with two participants separated by a board. A certified ASL interpreter was always present.

3.3.3 *Subjective Measures*

Participants were asked to fill out a subjective questionnaire after each 5-minute conversation. The questions are listed below and respondents circled the response that best answered the question.

- Question 1: How easy was it to understand the video?
  (7-point Likert scale ranging from very easy to very difficult)
- Question 2: Rate the video quality for sign language.
  (7-point Likert scale ranging from excellent to poor)
- Question 3: Rate the video quality for fingerspelling.
  (7-point Likert scale ranging from excellent to poor)
- Question 4: Rate the video quality for lip reading.
  (7-point Likert scale ranging from excellent to poor);
- Question 5: During the conversation, indicate how often you had to guess what the other signer was signing.
  (0% never, 25% sometimes, but not often, 50% half the time; 75% most of the time, and 100% all of the time).

After all trials were completed, participants filled out a demographic questionnaire which included questions such as, "how long have you been signing ASL?'; "what language do you prefer to sign with family?"; and, "do you own a smartphone?" Lastly, participants were asked exit interview questions regarding their overall experience while signing over the different frame rates and bit rates. Examples of questions asked included, "did you notice changes in video quality?"; "at any time were you frustrated with the video quality provided?"; and, "would you use the lower video quality if you knew you could save battery life?"

3.3.4 *Objective Measures*

A conversation with low intelligibility may contain a lot of requests for repetitions, called "repair requests" [35], which may include signing "what?" or "again" and "conversational breakdowns," where a signer may sign the equivalent of, "I didn't understand what you said." Also, the rate of signing may decrease with the lowered frame rate/bit rate. Therefore, we analyzed the rate of fingerspelling. Fingerspelling occurs when a signer spells out the name of something, which is usually for titles, proper names, and technical words. Signs that are lexicalized "loan signs," which are common words that have become the stylized fingerspelling, are not counted in our fingerspelling measure.

The objective measures were the number of repair requests, average number of turns associated with repair requests, number of conversational breakdowns, and speed of fingerspelling. These measures were calculated from the videotaped sessions with the assistance of a certified ASL interpreter. For each repair request, the number of turns was counted until the concept was understood. Conversational breakdowns were counted as the number of times the participant signed the equivalent of "I can't see you" due to the video being blurry, choppy, or frozen. An unresolved repair request was also counted as a conversational breakdown. Finally, the speed of fingerspelling was measured as the time it took to sign each letter of the word, divided by the number of characters in that word, producing the characters per second.

4. RESULTS

4.1 Perceived Intelligibility

Nonparametric analyses were used to analyze each question, which captured responses on 7-point Likert scales. Since data gathered were ordinal and dichotomous responses, a Friedman test [13] was used to analyze the main effect of frame rate/bit rate for each question. Separate pairwise Wilcoxon tests [36] with Holm's Sequential Bonferroni procedure [15] were performed to investigate the effect of frame rate/bit rate. Results will be reported for each question.

Question 1 asked participants to rate how easy it was to understand the video from 7-very easy to 1-very difficult. The Friedman test did not indicate a significant main effect of frame rate on perceived video intelligibility ($\chi^2_{3,N=20}$=5.08, *n.s.*).

Question 2 asked participants to rate the video quality for sign language communication from 7-excellent to 1-poor. The Friedman test indicated a significant main effect of frame rate on perceived video quality ($\chi^2_{3,N=20}$=11.01, *p*<.05). Wilcoxon tests with Holm's Sequential Bonferroni procedure were performed to identify the effect of frame rate on perceived video quality.

Increasing the frame rate from 5 fps/25 kbps vs. 10 fps/50 kbps, 15 fps/75 kbps, and 30 fps/150 kbps, respectively, was found to increase perceived video quality ($\chi^2_{3,N=20}$=46.5, *p*<.05). However, comparing perceived video quality between 10 fps/50 kbps, 15 fps/75 kbps, and 30 fps/150 kbps was not found to significantly increase perceived video quality ($\chi^2_{3,N=20}$=9.0, *n.s.*).

Question 3 asked participants to rate the video quality for fingerspelling from 7-excellent to 1-poor. The Friedman test indicated a significant main effect of frame rate on perceived video quality for fingerspelling ($\chi^2_{3,N=19}$=8.11, *p*<.05). Wilcoxon tests with Bonferroni procedure were performed to identify the effect of frame rate on perceived video quality for fingerspelling. Increasing the frame rate from 5 fps/25 kbps vs. 10 fps/50 kbps, 15 fps/75 kbps, and 30 fps/150 kbps, respectively, was found to increase perceived video quality ($\chi^2_{3,N=20}$=35.5, *p*<.05). However, comparing perceived video quality between 10 fps/50 kbps vs. 15 fps/75 kbps vs. 30 fps/150 kbps was not found to significantly increase perceived video quality for fingerspelling ($\chi^2_{3,N=20}$=10.0, *n.s.*).

Only half of the participants indicated that they lip read during signing. Therefore, analysis for question 4, which asked participants to rate the perceived video quality for lip reading from 7-excellent to 1-poor, was performed for 10 participants. The Friedman test did not indicate a significant main effect of frame rate on perceived video intelligibility for lip reading ($\chi^2_{3,N=10}$=2.92, *n.s.*).

Question 5 asked participants to rate how often they had to guess what the signer was signing during their conversation (0% never, 25% sometimes, but not often, 50% half the time; 75% most of the time, and 100% all of the time). The Friedman test indicated a significant main effect of frame rate on the rate at which participants had to guess what their signing partner was signing ($\chi^2_{3,N=20}$=29.75, *p*<.0001). Wilcoxon tests with Bonferroni procedure were performed to identify the effect of frame rate on participants guessing what the other signer was signing. Increasing the frame rate from 5 fps/25 kbps vs. 10 fps/50 kbps, 15 fps/75 kbps, and 30 fps/150 kbps, respectively, was found to decrease how often a participant had to guess what the other signer was signing ($\chi^2_{3,N=20}$=52.5, *p*<.001). However, comparing how often a signer had to guess what their partner was signing for video transmitted between 10 fps/50 kbps vs. 15 fps/75 kbps vs. 30 fps/150 kbps was not found to significantly reduce how often they guessed what the other person was signing ($\chi^2_{3,N=20}$=6.0, *n.s.*).

4.2 Objective Measures

All sessions were video recorded to be objectively analyzed in post-analysis with a certified ASL interpreter. Each conversation was analyzed to identify and count instances of (1) repair requests during a conversation; (2) conversational breakdowns; and (3) speed of fingerspelling (reported as characters per second - 1). Examples of repair requests include instances when a signer signs the equivalent of "what?" or "again."

A Friedman test was performed for each objective measure to determine how varying the frame rate affected it. Frame rate was found to significantly impact the number of repair requests ($\chi^2_{3,N=10}$=11.0, *p*<.05) and the number of conversation breakdowns made during a conversation ($\chi^2_{3,N=10}$19.8, *p*<.001); however, varying the frame rate was not found to statistically significantly impact the speed of fingerspelling ($\chi^2_{3,N=10}$=2.48, *n.s.*). Table 1 lists the number of instances of fingerspelling and the average characters signed per second at each frame rate. As Table 1 demonstrates, the average number of characters per second did not change as the frame rate increased, even though participants perceived changes in video quality. Perhaps, participants adapted quickly to the temporal video quality or used alternative methods, which are discussed further below.

Table 1: Count of the number of fingerspelled words and the average, max, min, and standard deviation of the number of characters signed per second.

frame rate/bit rate (fps/kbps)	5/25	10/50	15/75	30/150
Total count of finger spelled words (over all sessions)	153	191	166	180
average characters/sec	4.08	4.16	4.03	4.29
SD of characters/sec	1.99	2.03	1.45	1.97

Sign language conversations held over video transmitted at 5 fps/25 kbps received the most counts for both repair requests and conversational breakdowns, as expected. Video transmitted at 10 fps/50 kbps, 15 fps/75 kbps, and 30 fps/150 kbps did not have any instances of repair requests or conversational breakdowns across all sessions. Figure 3 lists the number of repair requests and conversational breakdowns that occurred for each session.

Figure 3 shows that sessions 6 and 7 received the highest counts for conversational breakdowns with 11 total breakdowns occurring in a 5 minute conversation. Participants in sessions 4, 5, 6, 7, 8, and 9 were friends while the other sessions had participants paired with strangers.

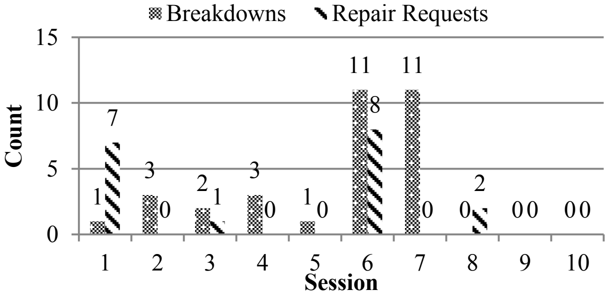

Figure 3: Count of conversational breakdowns and repair requests that occurred for each session when video was transmitted at 5 fps.

4.3 Exit Interviews

During the exit interviews, participants were asked to indicate which version of the video app they preferred use. There were four recurring themes that arose during the exit interviews, which were: (1) there was noticeable lower quality of video transmitted at 5 fps; (2) desire for larger screens; (3) different adaptation techniques were used to compensate for lower video quality; and (4) comparison of video quality used in the experimental app to commercially available apps. (Note that consent was obtained from study participants to include excerpts in publication.)

4.3.1 5 FPS Video Quality

All participants voiced their observations that video transmitted at 5 fps was noticeably more "choppy" or "frozen" than other versions of the app that they used. When asked what they liked or disliked about signing over video shown at 5 fps, many participants said they "would not want to use the video at all." P3 signed that she really could not express herself like she normally would when signing to someone in-person because of the lower video quality. P13 and P14 said they chose to have a "lighter conversation," *i.e.*, not talk about anything that required a lot of background information to be signed first. They were unsure how often they would need to repeat themselves so they wanted to keep the conversation short.

Many participants signed that they would not use mobile video communication at 5 fps, even though the video quality provided intelligible content. When asked if they would "give up" signing to each other at video transmitted at 5 fps, participants expressed that they probably would turn to texting to clarify what they wanted to say since texting is more reliable than mobile video at 5 fps. P17 and P18 said they would rather text message instead of sign over video transmitted at 5 fps. When asked why, they said because more energy was needed to repeat themselves over video, while texting required only one message. P17 did acknowledge that texting was asynchronous, but believed texting was more reliable than current mobile video apps. P18 followed up by saying she didn't use mobile video chat on her phone, so texting was her solution for mobile communication.

4.3.2 Desire for Larger Screens

During the exit interviews, many participants spoke about the form factor of the device, specifically desire for larger screen sizes. P13 and P14 made comments that they preferred to sign over a larger device with a bigger screen similar to the screens available on the iPad or Samsung Galaxy Note. P14 expressed she did not feel like she could express everything she wanted to say because of the confined signing space. Also, the angle at which video was shown made it more difficult to understand her signing partner. Mainly, the hands were closer to the screen, but the signer's head appeared to look like a "pin head" because of the camera angle. P14 also said that lip reading was hard to do because of the "pin head" appearance of her signing partner.

4.3.3 Adaptation Techniques

When participants were asked what adaptation techniques they used to compensate for the lower video quality, a majority of the participants said they deliberately fingerspelled more slowly than their regular signing speed. They also had to ask their signing partner to repeat what was signed and slow down whatever they were signing. Some participants also said doing this often disrupted what they were trying to say, which caused some frustration for both the signer and receiver. Interestingly, participants did not actually fingerspell more slowly when the frame rate varied (mean characters per second: 4.97 at 5 fps vs. 5.22 at 30 fps), as listed in Table 1, even though they were perceived to sign more slowly.

When participants were asked which version of the video app they preferred to use, many participants indicated they preferred signing over video transmitted at 15 and 30 fps; however, many participants indicated that they could not tell the difference between video transmitted at 15 fps and 30 fps. When asked if they noticed changes in video quality when video transmitted at 10 fps, participants did say it was better than video transmitted at 5 fps, but not as good as video transmitted at 15 or 30 fps.

4.3.4 Comparisons to Commercial Video Apps

In many of the laboratory sessions, participants compared the video quality they were using to commercially available apps like Skype and FaceTime. Those participants who referred to FaceTime said that FaceTime's video quality was clearer and smoother. This particular comment was expected since FaceTime transmits video at 30 fps at 1-3 Mbps at 960×640 screen resolution [20]. In one of the sessions, P7 and P8 were signing over video transmitted at 15 fps and began to discuss how IMSDroid's video quality compared to FaceTime:

P7: How does this compare to FaceTime?

P8: FaceTime is more clear, but this is fine… your hands are a little more blurry. I understand you fine though.

P7: Am I signing too fast?

P8: No, you're signing fine.

P7: Well...I'm signing normal, just trying to test the limitations. Is the fingerspelling clear?

P8: Yeah, I can see you fine.

P7: So when I spelled 'amoeba'

P8: Yes, amoeba

P7: Did you see all the signs or did you just catch the 'b' 'a'?

P8: …I saw the full spelling, but deaf [people] understand what you're saying anyhow. We're used to doing that.

This snippet of the conversation is an example of how people who are deaf naturally interpolate what they view to understand the overall message of a conversation. For instance, when words are fingerspelled, all the letters of the word may not have been viewed by the receiver, but the word can be discerned from the context of the conversation.

4.4 Battery Drain

The battery drain was unobtrusively logged using an open source app called AndroSensor, which ran in the background and logged the percentage battery drain every 30 seconds for each 5 minute conversation. Data were collected from the phones after each session for later analysis.

The rate at which the battery percentage depleted was calculated for each 5 minute video call. We verified that the battery drain was linear, which allowed us to use linear regression to model the data. The estimated average battery duration for each frame rate was calculated for every conversation and shown in Figure 4. As anticipated, the higher the frame rate at which video was transmitted, the higher the rate at which the battery drained. We found that the Samsung Galaxy S3 has an average battery life of 1000 minutes in standby mode and an average battery life of 750 minutes if IMSDroid was "active" but not transmitting video.

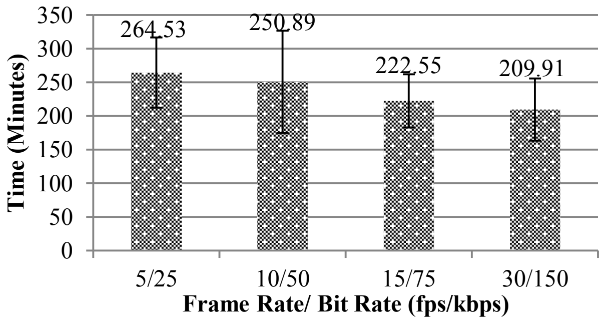

Figure 4: Estimated average battery life (in minutes) for sign language video transmitted on IMSDroid at each frame rate.

4.5 Bandwidth Consumption

Network traces were performed on the Asterisk server to monitor the average rate at which data was transmitted. Bit rate control is an active area of research [8, 12, 21, 26] and was not the focus of this study. Table 2 lists the average bit rate at which video was transmitted for each frame rate. The bit rate was controlled by the Asterisk server and the network traces confirmed that the frame rate dictated the bit rate at which video was transmitted.

Table 2: Average, min, max, and SD of the bit rate when varying the frame rate as captured by the network traces.

frame rate (fps)	target bit rate (kbps)	average bit rate (kbps)	min bit rate (kbps)	max bit rate (kbps)	SD (kbps)
5	25	23.89	20.87	32.19	3.38
10	50	50.00	39.78	67.76	8.67
15	75	73.04	64.43	91.25	8.67
30	150	129.89	114.78	147.38	9.91

5. DISCUSSION

Participants were successful at holding intelligible conversations across all frame rates. All participants did notice and complain about the lower quality of video transmitted at 5 fps; however, participants' rate of fingerspelling did not decrease, even though they perceived their signing speed to be slower. Video transmitted at 5 fps had more instances of conversational breakdowns and repair requests. Sessions 6 and 7 received the most counts for conversational breakdowns (11 instances); the frequencies at which breakdowns occurred were low across other sessions. Closer inspection of the conversations held in sessions 6 and 7, where the most breakdowns and repair requests occurred, revealed that the topic of conversation was very detailed and required more explanation. For example, P11 and P12 from session 6 were talking about a trip to Iceland. P12 asked if P11 was going to see the Aurora Borealis. It took multiple attempts by P11 asking the question to clarify what P12 was asking. The frame rate at which the video was signing was 10 fps/50 kbps. The conversational breakdown could have resulted from the conversation topic and not because of the video transmission rate.

5.1 Signing Adaptation Techniques

Signers are versatile when it comes to adapting their signing to the technology they use to communicate. The context of a conversation, signs used, loan signs (signs that represent an English word that has developed a unique movement), and fingerspelling words all assist in filling in missing information [4]. Signers may be naturally taking advantage of the "word superiority effect" where people are more successful recognizing letters presented within words that just isolated letters [3]. This may explain why the rate of fingerspelling did not vary across the frame rates.

During objective analysis of the video conversations, there were instances in which a participant would begin to finger-spell a word; however, she did not spell every letter within that word. For example, a participant was talking about the different seasons, but when she fingerspelled "season," she only signed "s" and "n" of the word. The receiver of the message was still able to infer the word. The receiver may also have been able to infer the word from the context of the message. Often the context of a conversation can aid in understanding a word that was not seen during the conversation [22].

5.2 Willingness to Use Lower Video Quality

When asked if they were willing to use a low video quality to hold conversations, all participants said they would be willing to use the mobile technology if there were a guarantee that video would be transmitted at 15 fps/75 kbps or 30 fps/150 kbps. However, video transmitted at lower frame rates would only be used for very short conversations, such as asking a quick question. When given the option between texting and mobile video chatting, participants said they always would prefer to sign over video; however, if the person they are communicating with does not sign, texting is considered necessary.

5.3 Technology Position Adjustments

Participants were allowed to adjust the mobile device to a position that felt comfortable. Some of the participants adjusted the phone to increase the angle at which it was displayed or raised the phone to increase their signing space. Figure 5(a) shows the original position of the phone placed in front of the participants. Figure 5(b) shows how a participant placed a pen behind the phone to increase the angle at which he viewed the phone. Figure 5(c) and 5(d) are two different examples of how participants requested to use stacks of books located in the room to raise the smartphone's position.

5.4 Recommendations

As anticipated, reducing the frame rate at which sign language video is transmitted increases the average battery life of IMSDroid. From the laboratory results, it is recommended that conversational video transmitted at 10 fps/50 kbps best balances resource consumption, video intelligibility, and user preferences. Transmitting video at 10 fps/50 kbps, 15 fps/75 kbps, and 30 fps/150 kbps received, on average, the same subjective responses from participants when asked to rate how easy it was to understand the video; rate the video for picture quality, fingerspelling, and lip-reading; and how often the signer had to guess what the other person was signing. While the battery life lasted the longest when video was transmitted at 5 fps/25 kbps, video transmitted at 5 fps/25 kbps also received the most counts for repair requests and conversational breakdowns. Finally, in the exit interviews, participants voiced their dissatisfaction of communicating at video transmitted at 5 fps/25 kbps because of the choppy video quality. Although some participants were able to tell that there was a difference between video transmitted at 10 fps/50 kbps vs. 15 fps/75 kbps vs. 30 fps/150 kbps in the exit interviews, both the subjective and objective results support that video transmitted at 10 fps/50 kbps is the lowest threshold at which intelligible sign language conversations can be comfortably held.

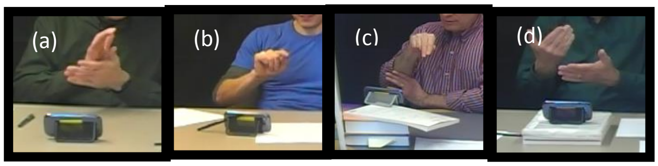

Figure 5: Four examples of how participants adjusted the phone position. (a) Original phone setup using a business card holder. (b) Phone propped up with a pen. (c) Increased height and viewing angle. (d) Increased height from table.

6. CONCLUSION AND FUTURE WORK

The ITU-T standard recommends that video should be transmitted at least at 25 fps and 100 kbps for intelligible conversations. Our laboratory study clearly demonstrates that there is a lower limit at which intelligible mobile sign language video can be transmitted. Our findings suggest that video transmitted at 10 fps with a bit rate averaging 50 kbps can facilitate intelligible sign language conversations, and can extend battery life by almost 20% compared to transmitting at 30 fps and 150 kbps.

The findings from this study provide the motivation for the creation of video technology specifically designed for use during emergencies and natural disasters, where the full cellular network infrastructure may become unavailable. In 2005, it was estimated that 50% of the total phone lines and wireless subscribers lost access to phone service for multiple days after Hurricane Katrina hit land [23]. In the laboratory study, people were still successful at holding intelligible conversations at 5 fps (averaging 23.89 kbps) even though participants did not prefer communicating at those video transmission rates. Having the capability to transmit emergency videos, even at these low transmission rates, would be useful to relay important information.

7. ACKNOWLEDGMENTS

The authors wish to thank the participants; Tobias Cullins, Coordinator of Deaf & Hard of Hearing Services; Jennifer Austin for interpreting video content; Rafael Rodriguez for the implementation of IMSDroid. Lydia Runnels from ZVRS. Samsung Research for the smartphones. This work was funded in part by Google.

8. REFERENCES

[1] AndroSensor: 2013. *http://www.fivasim.com/androsensor.html*.

[2] Asterisk: 2014. *http://www.asterisk.org/*. Accessed: 2014-01-04.

[3] Baron, J. and Thurston, I. 1973. An analysis of the world-superiority effect. *Cognitive Psychology*. 4, 2, 207–228.

[4] Battison, R. 1978. *Lexical borrowing in American Sign Language*.

[5] Cavender, A., Ladner, R. and Riskin, E. 2006. MobileASL: Intelligibility of sign language video as constrained by mobile phone technology. *Proc. ASSETS*, 71–78.

[6] Chen, B. 2013. AT&T allows FaceTime for limited data users. What about unlimited? *The New York Times*.

[7] Chen, J.Y.C. and Thropp, J.E. 2007. Review of low frame rate effects on human performance. *IEEE Trans. on Systems, Man, and Cybernetics - Part A: Systems and Humans*. 37, 6, 1063–1076.

[8] Chen, Z. and Ngan, K. 2007. Recent advances in rate control for video coding. *Signal Processing: Image Communication*. 22, 1, 19–38.

[9] Cherniavsky, N., Chon, J., Wobbrock, J.O., Ladner, R. and Riskin, E. 2007. Variable Frame Rate for Low Power Mobile Sign Language Communication. *Proc. ASSETS*, 163–170.

[10] Cicco, L., Mascolo, S. and Palmisano, V. 2008. Skype video responsiveness to bandwidth variations. *NOSSDAV*.

[11] Costs associated with using FaceTime: 2013. *http://www.ilounge.com/index.php/articles/comments/costs-associated-with-using-facetime/*.

[12] Ding, W. and Liu, B. 1996. Rate control of MPEG video coding and recording by rate-quantization modeling. *IEEE Trans. Circuits Syst. Video Technol*. 6, 1, 12–20.

[13] Friedman, M. 1937. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association*. 32, 200, 675–701.

[14] Harkins, J., Wolff, A., Korres, E., Foulds, R. and Galuska, S. 1990. Intelligibility experiments with a feature extration system designed to simulate a low-bandwidth video telephone for deaf people. *RESNA*, 92–95.

[15] Holm, S. 1979. A simple sequentially rejective multiple test procedure. *Scand J Stat*. 6, 2, 65–70.

[16] Hooper, S., Miller, C., Rose, S. and Veletsianos, G. 2007. The effects of digital video quality on learner comprehension in an American sign language assessment environment. *Sign Language Studies*. 8, 1, 42–58.

[17] IMSDroid-High Quality Video SIP/IMS client for Google Android: *http://code.google.com/p/imsdroid/*. Accessed: 2012-05-23.

[18] Johnson, B.F. and Caird, J.K. 1996. The effect of frame rate and video information redundancy on the perceptual learning of American sign language gestures.

[19] Manoranjan, M.D. and Robinson, J. a 2000. Practical low-cost visual communication using binary images for deaf sign language. *IEEE transactions on rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society*. 8, 1, 81–8.

[20] Ou, G. 2010. Estimate of network bandwidth for iPhone 4 FaceTime. *Digital Society*.

[21] Reed, E. and Lim, J. 2002. Optimal multidimensional bit-rate control for video communication. *IEEE Trans. Image Process*. 11, 8, 873–885.

[22] Reicher, G. 1969. Perceptual recognition as a function of meaningfulness of stimulus material. *Experimental Psychology*. 81, 2, 275–280.

[23] Reilly, G., Jrad, A., Nagarajan, R., Brown, T. and Conrad, S. 2006. Critical infrastructure analysis of telecom for natural disaters. *Telecom. Network Strategy and Planning*, 1–6.

[24] Saks, A. and Hellström, G. 2006. Quality of conversation experience in sign language , lip - reading and text. *ITU-T Workshop on End-to-end QoE/QoS*.

[25] Skype Statistics: 2012. *http://www.statisticbrain.com/skype-statistics*.

[26] Song, H. and Kuo, C. 2001. Rate control for low-bit-rate video via variable-encoding frame rates. *IEEE Trans. Circuits Syst. Video Technol.* 11, 4, 512–521.

[27] Sosnowski, T. and Hsing, T. 1983. Toward the conveyance of deaf sign language over public telephone networks. *RESNA*.

[28] Sperling, G. 1981. Video transmission of American Sign Language and finger spelling: present and projected bandwidth requirements. *IEEE Transactions on Communications*. 29, 12, 1993–2002.

[29] Sperling, G., Landy, M., Cohen, Y. and Pavel, M. 1985. Intelligible encoding of ASL image sequences at extremely low information rates. *Computer Vision Graphics, and Image Processing*. 31, 335–391.

[30] Tran, J.J., Kim, J., Chon, J., Riskin, E., Ladner, R. and Wobbrock, J.O. 2011. Evaluating quality and comprehension of real-time sign language video on mobile phones. *Proc. ASSETS*, 115–122.

[31] Tran, J.J., Riskin, E., Ladner, R. and Wobbrock, J.O. 2013. Increasing mobile sign language video accessibility by relaxing video transmission standards. *Third Mobile Accessibility Workshop at Proc. CHI*.

[32] Tran, J.J., Rodriguez, R., Riskin, E. and Wobbrock, J.O. 2013. A web-based intelligibility evaluation of sign language videotransmitted at low frame rates and bitrates. *Proc. ASSETS*.

[33] Verizon begins throttling iPhone unlimited 3G customers who use 2GB/month | 9to5Mac | Apple Intelligence: *http://9to5mac.com/2011/09/17/verizon-begins-throttling-iphone-2gbmonth-unlimited-3g-customers/*. Accessed: 2012-01-04.

[34] Video is fastest growing mobile data traffic source: 2013. *http://www.humanipo.com/news/36341/video-is-fastest-growing-mobile-data-traffic-source/*.

[35] Watson, A. and Sasse, M.A. 1998. Measuring perceived quality of speech and video in multimedia conferencing applications. *Multimedia*, 55–60.

[36] Wilcoxon, F. 1945. Individual comparisons by ranking methods. *Biometrics Bulletin*. 1, 6, 80–83.