Increasing Mobile Sign Language Video Accessibility by Relaxing Video Transmission Standards

Jessica J. Tran
Electrical Engineering
DUB Group
University of Washington
Seattle, WA 98195 USA
jjtran@uw.edu

Jacob O. Wobbrock
The Information School
DUB Group
University of Washington
Seattle, WA 98195 USA
wobbrock@uw.edu

Eve A. Riskin
Electrical Engineering
University of Washington
Seattle, WA 98195 USA
riskin@u.washington.edu

Richard E. Ladner
Computer Science & Engineering
University of Washington
Seattle, WA 98195 USA
ladner@cs.washington.edu

CHI'13, April 27 – May 2, 2013, Paris, France.

Abstract
The current recommended video transmission standards, Telecommunication Standardization Sector (ITU-T) Q.26/16, of 25 frames per second at 100 kilobits per second or higher make mobile sign language video communication less accessible than it could be with a more relaxed standard. The current bandwidth requirements are high enough that network congestion may occur, causing delays or lost information. In addition, limited data plans may cause higher cost to video communication users. To increase the accessibility and affordability of video communication, we explore a relaxed standard for video transmission using lower frame rates and bitrates. We introduce a novel measure, the Human Signal Intelligibility Model, to accomplish this. We propose web and laboratory studies to validate lower bounds on frame rates and bitrates for sign language communication on small mobile devices.

Author Keywords
mobile video communication; video compression; American Sign Language (ASL); intelligibility; Deaf community; ITU-T standards

ACM Classification Keywords
H.5.1. [Information Interfaces and Presentation]: Multimedia Information Systems- Video.

Introduction

Commercial mobile video applications, intended for sign language communication, transmit video at the current recommended Telecommunication Standardization Sector (ITU-T) Q.26/16 standards of 25 frames per second (fps) at 100 kilobits per second or higher [7]. However, the high bitrate makes mobile sign language video communication less accessible. First, congestion is more likely to cause delays and lost information. Second, the cost of calling can be high because of limited data plans. We argue that the frame rate and bitrate recommended for the ITU-T standards are higher than needed to support sign language communication on mobile devices.

We suggest either reducing the frame rate or the number of bits allocated per frame to make mobile sign language video communication more accessible. Prior research indicates that frame rates lower than 25 fps yield intelligible sign language [3,4,9]. It is always the case that more bits per frame gives better quality, but there is a tradeoff between cost and video quality. Major U.S. cellular networks are throttling down network speeds in response to high data consumption rates [10]. People who communicate using mobile video or mobile video-relay services consume network bandwidth faster than average data users. Currently, cellular phone companies do not subsidize the extra cost of mobile sign language video communication.

Our research investigates the lower limits of frame rates and bitrates to yield intelligible sign language communication over small mobile devices. We introduce a novel measure, the *Human Signal Intelligibility* Model, to evaluate sign language video intelligibility. Our research demonstrates that intelligible sign language video content can be transmitted at lower frame rates and bitrates than the current recommended ITU-T standards.

Related Work

MobileASL(American Sign Language)

MobileASL is an experimental smart phone application providing two-way, real-time sign language video at very low bandwidth (30 kilobits per second at 8-12 frames per second) [11]. Cavender *et al.* [3] conducted preliminary research investigating intelligibility of video quality transmitted at various frame rates, bitrates, and region-of-interest (ROI) (10, 15 fps; 15, 20, 25 kbps; and 0, -6, -12 ROI). They found respondents preferred video transmitted at the lower frame rate of 10 fps, given a fixed bitrate. In a laboratory study, Cherniavsky *et al*., [4] evaluated the impact of lowering the frame rate when a person was not signing during a mobile video conversation. They found more conversation breakdowns occurred such as the need for participants to repeat themselves. However, participants expressed that reduced video quality did not prevent potential use of mobile sign language video. Tran *et al.* [9] evaluated video quality perception when different power saving algorithms were applied, specifically reducing the spatial resolution and/or frame rate of not signing content. They found that reducing both the frame rate *and* spatial resolution during not signing content extended the battery life the most and was perceived with the fewest changes in video quality.

We aim to demonstrate more rigorously that intelligible mobile sign language video content can be transmitted at frame rates and bitrates lower than the current recommended ITU-T standards. We introduce a novel model, the Human Signal Intelligibility Model (HSIM), to accomplish this.

Evaluating Sign Language Video Intelligibility

We are creating (1) an intelligibility model, the Human Signal Intelligibility Model, to distinguish the components comprising video intelligibility from video comprehension, and (2) using this model to evaluate sign language video intelligibility in a web study and laboratory study.

Human Signal Intelligibility Model

The Human Signal Intelligibility Model (HSIM) was created to address the lack of uniformity in the way that signal intelligibility and signal comprehension are operationalized. Often, intelligibility and comprehension are loosely defined and used interchangeably to validate video quality. Some researchers define higher objective video quality to mean greater intelligibility of content, assuming the users has sufficient knowledge for content comprehension [5,6]. Peak signal-to-noise (PSNR) ratio is one commonly used objective measure, measuring quality of image reconstruction after lossy compression. Existing communication models [1,2] that attempt to define intelligibility from comprehension are poorly defined.

The HSIM (1) extends Shannon's theory of communication [8] to include the human and environmental influences on signal intelligibility and signal comprehension, and (2) identifies the components that make up *intelligibility* of a communication signal and separate those from the *comprehension* of a communication signal. Signal intelligibility and signal comprehension need to be distinguished because intelligibility does not entail comprehension. Intelligibility depends on signal quality, specifically how the signal was captured, transmitted, received, and perceived by the receiver, including the environmental conditions affecting these steps. Comprehension relies on signal quality *and* the human receiver having the prerequisite knowledge to process the information.

Study Design

Using the HSIM, we are creating a web study and laboratory study to evaluate how much video quality can be reduced in terms of frame rate and bitrate before intelligibility is compromised. In a web study, four low frame rates (1, 5, 10, 15 frames per second) and four low bitrates (15, 30, 60, 120 kilobits per second) were selected for evaluation in a full factorial mixed-methods design. Note that the frame rates and bitrates selected are much lower than the recommended ITU-T standards.

In the web study, participants watch 16 short ASL videos of a native male ASL signer signing short sentences shown at different frame rate and bitrate pairs. After each video, participants rate how easy the video was to understand. To date there are 99 respondents. We anticipate finding two specific frame rate and bitrate pairs: one where video quality begins to affect intelligibility too negatively and one where increasing resource allocation no longer provides significant gains.

The results of the web study will inform the parameter settings used in the laboratory study to evaluate two-way, real-time sign language video conversations. We aim to discover the minimum video quality settings needed for intelligible sign language communications while objectively measuring signal intelligibility.

Conclusion

People who choose to communicate via mobile sign language video should not have to pay more for video phone service. ITU-T standards recommend that mobile sign language video content needs to be transmitted at 25 fps at 100 kbps or higher to have intelligible conversations. However, delay and lost video content often occur because total network bandwidth is limited.

Our research makes three significant contributions: (1) making real-time mobile video communication more accessible to deaf and hard-of-hearing people by reducing bandwidth consumption; (2) proposing the Human Signal Intelligibility Model (HSIM), which addresses the lack of a universal definition for signal intelligibility, and (3) demonstrating intelligible mobile sign language video can be transmitted at relaxed frame rates and bitrates recommended by the ITU-T standards. We anticipate that the knowledge gained will improve how sign language video is transmitted on mobile video applications and empower users to communicate in their native language on mobile devices.

References

1. Barnlund, D.C. A transactional model of communication. In *Communication Theory*. New Brunswick, New Jersey, 2008, 47–57.
2. Berlo, D.K. *The Process of Communication*. Holt, Rinehart, & Winston, New York, New York, USA, 1960.
3. Cavender, A., Ladner, R., and Riskin, E. MobileASL: Intelligibility of Sign Language Video as Constrained by Mobile Phone Technology. *Proceedings of ASSETS 2006: ACM SIGACCESS Conference on Computers and Accessibility*, (2006).
4. Cherniavsky, N., Chon, J., Wobbrock, J., Ladner, R., and Riskin, E. Variable Frame Rate for Low Power Mobile Sign Language Communication. *Proceedings of ASSETS 2007: The Ninth International ACM SIGACCESS Conference on Computers and Accessibility*, (2007), 163–170.
5. Ciaramello, F. and Hemami, S. Quality versus Intelligibility: Studying Human Preferences for American Sign Language Video. *SPIE Human Vision and Electronic Imaging 7865*, (2011).
6. Feghali, R., Speranza, F., Wang, D., and Vincent, A. Video Quality Metric for Bit Rate Control via Joint Adjustment of Quantization and Frame Rate. *IEEE Transactions on Broadcasting 53*, 1 (2007), 441–446.
7. Saks, A. and Hellström, G. Quality of conversation experience in sign language , lip - reading and text. 2006.
8. Shannon, C.E. A mathematical theory of communication. *The Bell System Technical Journal 27*, 379-426 (1948), 623–656.
9. Tran, J.J., Johnshon, T., Kim, J., et al. A Web-Based User Survey for Evaluating Power Saving Strategies for Deaf Users of MobileASL. *The 12th International ACM SIGACCESS Conference on Computers and Accessibility (Assets 10')*, (2010), 115–122.
10. 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/.
11. MobileASL. University of Washington. 2012. http://mobileasl.cs.washington.edu/.