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 2013 Mobile Accessibility Workshop, April 28, 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 cause lost information. In addition, capped 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 sign language video transmission using lower frame rates and bitrates. We propose web and laboratory studies to validate lower bounds on frame rates and bitrates for sign language communication on small mobile devices. We introduce a new model, the *Human Signal Intelligibility Model*, for informing video intelligibility evaluations.

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 [6]. However, the high bitrate makes mobile sign language video communication less accessible. Network congestion is more likely to cause delays and lost information, and the cost of calling can be high. We argue that the frame rate and bitrate recommended for the ITU-T standards are higher than needed to support intelligible 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 video [3,4,8]. While more bits per frame always gives better quality, 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 [9]. People who communicate using mobile video or mobile video-relay services consume network bandwidth faster than other 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 new model, the *Human Signal Intelligibility Model*, to inform evaluations of 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) [10]. 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 conversational breakdowns occurred when lowering the frame rate during not-signing portions of a conversation. It was observed participants needed to repeat themselves; however, participants expressed that reduced video quality did not prevent potential use of mobile sign language video. In prior work [8], we evaluated video quality perception when different power-saving algorithms were applied, specifically reducing the spatial resolution and/or frame rate of not-signing content. We found that reducing both the frame rate *and* spatial resolution during not-signing extended the battery life most and was perceived with the fewest negative 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

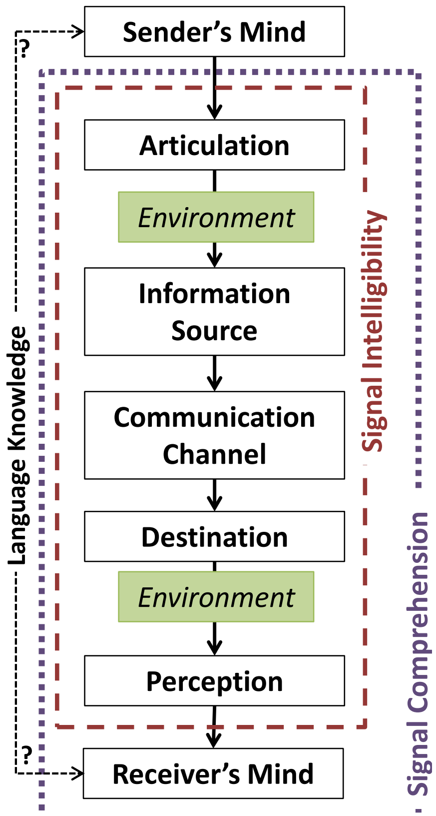

Figure 1: Block diagram of the Human Signal Intelligibility Model. Here the information source and destination are mobile phones. The communication channel represents Shannon's theory of communication. Note that the components comprising signal intelligibility are a subset of signal comprehension.

recommended ITU-T standards. We introduce a novel model, the Human Signal Intelligibility Model (HSIM), for informing our and others' human-centered evaluations of video content.

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 quality and video comprehension, which we argue are three distinct and separable things; and (2) using this model to evaluate sign language video intelligibility in web and laboratory studies.

Human Signal Intelligibility Model

We are developing the Human Signal Intelligibility Model (HSIM), as shown in Figure 1, to address the lack of uniformity in the way that signal intelligibility and signal comprehension are operationalized, especially in contrast to objective video quality measures like peak signal-to-noise ratio (PSNR), which measures the quality of image reconstruction after lossy compression. Too often, intelligibility and comprehension are loosely defined and used interchangeably to validate video quality, and some researchers treat higher PSNR to mean greater intelligibility, assuming users have sufficient knowledge for content comprehension [5]. Furthermore, existing communication models [1,2] that attempt to distinguish intelligibility from comprehension are poorly defined.

The HSIM (1) extends Shannon's theory of communication [7] 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, which lies outside the purview of those seeking to improve sign language video.

Study Design

The HSIM is informing our web and laboratory studies evaluating how much frame rate and bitrate can be reduced before intelligibility is compromised. The web study has participants watch 16 short ASL videos of a native male ASL signer signing short sentences shown at different frame rates (1, 5, 10, 15 frames per second) and bitrates (15, 30, 60, 120 kbps). Note that the frame rates and bitrates selected are much lower than the recommended ITU-T standards. After each video, fluent ASL participants rate how easy the video was to understand. Although we are measuring comprehension, we screen participants to ensure they are fluent in ASL and therefore comprehension is controlled for, allowing us to isolate intelligibility. 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 web study findings will inform the parameter settings implemented in an experimental smartphone application used in our laboratory study. We aim to discover the minimum video quality settings needed for real-time mobile sign language video conversations 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 real-time mobile video communication more accessible to deaf and hard-of-hearing people by (1) providing an alternative method of communication to text or audio and (2) increasing affordability of video communication by reducing bandwidth consumption. We propose the Human Signal Intelligibility Model (HSIM) for informing evaluations of sign language video and isolating intelligibility as distinct from objective video quality and video comprehension. Preliminary findings from our web study suggest, but do not confirm, that intelligible ASL video can be transmitted at frame rates and bitrates lower than the ITU-T standard. In future work, our laboratory study will have participants making real-time mobile video calls transmitted at low frame rates and bitrates, in a controlled environment, to help demonstrate intelligible sign language conversations can occur at lower frame rates and bitrates recommended than those by the ITU-T standards.

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. 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.
6. Saks, A. and Hellström, G. Quality of conversation experience in sign language , lip - reading and text. 2006.
7. Shannon, C.E. A mathematical theory of communication. *The Bell System Technical Journal 27*, 379-426 (1948), 623–656.
8. 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.
9. 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/.
10. MobileASL. University of Washington. 2012. http://mobileasl.cs.washington.edu/.