Towards Human-Centered Optimization of Mobile Sign Language Video Communication

Jessica J. Tran
Electrical Engineering | DUB Group

University of Washington
Seattle, WA 98195 USA
jjtran@uw.edu

Abstract

The mainstream adoption of mobile video communication, especially among deaf and hard-of-hearing people, is heavily reliant on cellular network capacity. Video compression lowers the rate at which video content is transmitted; however, intelligibility may be sacrificed. Currently, there is not a standard method to evaluate video intelligibility, or a good communication model on which to base evaluation. I am developing a better theoretical model, the *Human Signal Intelligibility (HSI)*, to evaluate intelligibility of lowered video quality for the purpose of reducing bandwidth consumption and extending cell phone battery duration. The goal of my dissertation is to advance mobile sign language video communication so it does not rely on higher cellular network bandwidth capacities. I will conduct this work by (1) identifying the components in the HSI model that make up *intelligibility* of a communication signal and separating those from the *comprehensibility* of a communication signal; and (2) using this model to identify how low video quality can get before the intelligibility of video content is sacrificed. Thus far, I have evaluated the use of mobile video communication among deaf and hard-of-hearing teenagers; developed two new power-saving algorithms; and quantified battery savings and evaluated user perception when those algorithms are applied.

Introduction

Mobile video is becoming a technology of choice for communication among deaf and hard-of-hearing people in the United States because their native language is American Sign Language (ASL), which is a visual language. Most major U.S. cellular networks no longer provide unlimited data plans, or are throttling down network speeds in response to high data consumption rates. This limits mainstream adoption of mobile video communication because video users consume network bandwidth faster than average data users. Currently, cellular phone companies do not subsidize the extra cost of mobile video communication used by deaf and hard-of-hearing people.

The goal of my dissertation research is to use video compression algorithms to reduce bandwidth consumption and increase battery duration for mobile sign language video communication. My research will answer how low video quality can get in terms of bitrate and frame rate before intelligibility is compromised, and will quantify how much battery life can be extended. These findings will make mobile video communication more accessible while providing intelligible content, reducing bandwidth consumption, and extending cell phone battery life.

Related Work

Measuring human intelligibility and comprehension of a signal (audio or video) is challenging because it relies on multiple factors such as common ground [4]; signal quality; and the environment in which signals are transmitted. Researchers have attempted to link higher objective signal quality to greater intelligibility of content, assuming that the user has sufficient knowledge and perception abilities for comprehension [3,5]. One such measure of signal quality is peak signal-to-noise ratio (PSNR), which measures quality of image reconstruction after lossy compression. Thu and Ghanbari [10] have demonstrated that PSNR is a reasonable measurement when used across the same content; however, PSNR may not necessarily reflect comprehension, which is most important for sign language communication. Other researchers have focused on measuring signal intelligibility with the intent that if one finds the signal intelligible, then comprehension of content may follow [6,7]. Often human intelligibility and comprehensibility are used interchangeably to validate signal quality. Furthermore, existing communication models only focus on the communication channel itself [9] without considering the environment or the human sender and receiver. Models that have attempted to do so have been poorly defined and do not clearly identify the components of video intelligibility and comprehensibility [1,2].

Proposed Solution

A clear theoretical model designed for evaluating signal intelligibility and signal comprehensibility is lacking. I am developing a better theoretical model, the *Human Signal Intelligibility (HSI)*, which distinguishes the components of *signal intelligibility* from *signal comprehensibility* for evaluating communication signals. The HSI model will (1) extend Shannon's theory of communication [9] to include the human and environmental influences on signal intelligibility and signal comprehensibility, and (2) identify the components that make up *intelligibility* of a communication signal and separate those from the *comprehensibility* of a communication signal as demonstrated in Figure 1.

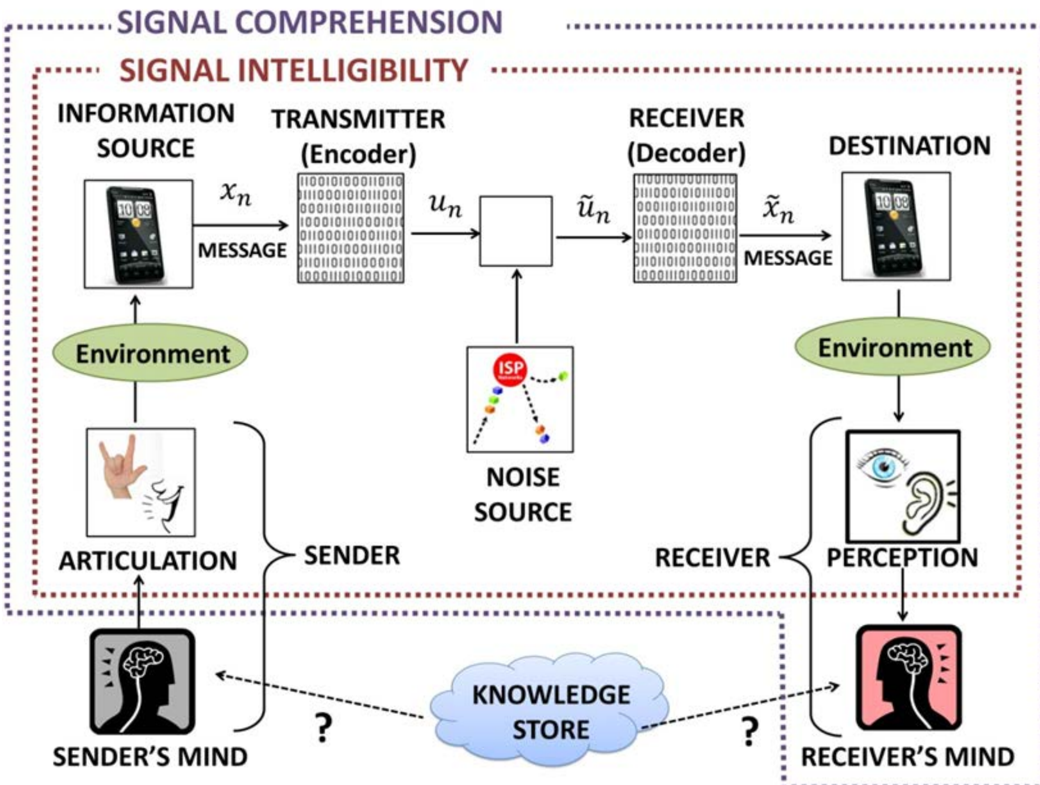

Figure 1: Block diagram of the Human Signal Intelligibility Model. Note that the components comprising intelligibility is a subset of signal comprehension.

The distinction between signal intelligibility and signal comprehension is important because signal intelligibility does not imply signal comprehension. Intelligibility depends on signal quality, specifically how the signal was captured, transmitted, received, and perceived by the receiver. Comprehension relies on signal quality *and* the human receiver having prerequisite knowledge to process the information. In the HSI model, the components that comprise intelligibility are a subset of the components that comprise comprehension. Identifying where intelligibility breaks down is important to determine how much video quality can be reduced. For the purpose of this work, I will focus on mobile sign language video.

Motivated by the HSI model, I will investigate how low sign language video quality can get before intelligibility is sacrificed. Specifically, I will reduce the frame rate and bitrate at which video is transmitted because these directly impact bandwidth consumption and the battery power needed to capture, transmit, and receive video.

Progress and Future Work

My Master's work created two new power saving algorithms; quantified battery savings; and evaluated user perception when those algorithms were applied [11,13]. The major finding was that the algorithm that extended the battery life the most (reducing both the frame rate and the frame size of video content) resulted in the least amount of perceived changes in video quality. I have also led a pilot field study with deaf and hard-of-hearing teenagers investigating use of an experimental smart phone application, *MobileASL*, that transmits video at extremely low bandwidths [8]. My colleagues and I found that mobile video communication was preferred over text messaging; however, participants expressed that the short battery life limited their use of the technology. This and other findings further motivate the need for longer battery duration for successful mobile video communication.

Future work will conduct a full factorial mixed-methods study to investigate how lowering video quality by reducing the rate at which video is transmitted (specifically, bitrates of 15, 30, 60, 120 kbps and frame rates of 1, 5, 10, 15 fps) will impact intelligibility of sign language video and resource consumption. In a web study, I 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 (*i.e.*, a point of diminishing returns). The laboratory study will use a subset of parameter settings between these frame rate and bit rate pairs to evaluate intelligibility. In pairs of two, participants will be video recorded signing to each other over MobileASL at lowered video qualities. In post analysis, intelligibility will be measured by counting the number of repair requests; average number of turns associated with repair requests; and the number of conversational breakdowns. Finally, battery life will be quantified once transmission rates are known.

Contributions

I foresee my research contributing to effective mobile sign language video communication and to the theory of video signal intelligibility and signal comprehension. This work builds on my prior work presented at ASSETS 2010 [11] and ASSETS 2011 [12], which demonstrates the continued importance of making mobile communication more accessible to everyone.

Goals for the Consortium

My proposed dissertation research was approved by my thesis committee in June 2012. Although the main goals are set, there is much to be determined in the specifics of the study design such as how to control variables that make up the HSI model, and how to measure the effects of varied video quality on intelligibility. Validation of the HSI model is reliant on my proposed web and laboratory studies. Feedback on the HSI model and learning of best practices for the study design will increase the validity and model robustness. From the DC, I would like to explore future directions and applications of the Human Signal Intelligibility model.

Conclusion

While I have made progress in my dissertation research, I believe the ASSETS 2012 Doctoral Consortium is the ideal community to present the Human Signal Intelligibility model and empirical study designs. I am eager to receive feedback to make the model more robust; meet fellow researchers in the accessibility field; and contribute to thoughtful discussions.

References

[1] Barnlund, D. (2008). A transactional model of communication. *Communication Theory*. 47-57.

[2] Berlo, D. (1960). *The Process of Communication*. Holt, Rinehart, & Winston.

[3] Ciaramello, F. and Hemami, S. (2011). Quality vs. Intelligibility: Studying Human Preferences for American Sign Language Video. *SPIE Human Vision and Electronic Imaging*. 7865.

[4] Clark, H. and Brennan, S. (1991). Perspectives in Socially Shared Cognition, Grounding in Communication. *American Psychological Association*. 127-149.
[5] Feghali, R., Speranza, F., Wang, D., and Vincent, A. (2007). Video Quality Metric for Bit Rate Control via Joint Adjustment of Quantization and Frame Rate. *IEEE Transactions on Broadcasting*. 53, 1, 441-446.
[6] Harrigan, K. (1995). The SPECIAL System: Self-Paced Education with Compressed Interactive Audio Learning. *Journal of Research on Computing in Education*. 3, 27, 361-370.
[7] 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.
[8] Kim, J., Tran, J., Johnson, T., Ladner, R., Riskin, E. and Wobbrock, J. (2011). Effect of MobileASL on Communication Among Deaf Users. *Extended Abstracts of the ACM Conference on Human Factors in Computing Systems (CHI '11)*, 2185-2190.
[9] Shannon, C.E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*. 27, 379-426, 623-656.
[10]Thu, H. and Ghanbari, M. (2008). Scope of Validity of PSNR in image/video quality assessment. *Electronic Letters*. 44, 13, 800-801.
[11]Tran, J., Johnson, T., Kim, J., Rodriguez, R., Yin, S., Riskin, E., Ladner, R. and Wobbrock, J. (2010). 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')* (Orlando, FL), 115-122.
[12]Tran, J., Kim, J., Chon, J., Riskin, E., Ladner, R. and Wobbrock, J. (2011). Evaluating Quality and Comprehension of Real-Time Sign Language Video on Mobile Phones. *The 13th International ACM SIGACCESS Conference on Computers and Accessibility (Assets 11')* (Dundee, Scotland, UK, 2011), 115-122.
[13]Tran, J. (2010). *Power Saving Strategies for Two-Way, Real-Time Video-Enabled Cellular Phones*. M.S. Thesis, Electrical Engineering Department, University of Washington, Seattle.

About the Author:

Jessica J. Tran is a Ph. D. Candidate in Electrical Engineering (EE) at the University of Washington (UW), Seattle. She is tri-advised by Professor Eve A. Riskin (EE), Professor Richard E. Ladner (CSE), and Associate Professor Jacob O. Wobbrock (iSchool). She earned her MSEE in 2010 and BSEE in 2008 from UW. Her research interests are in digital signal processing-video compression- and Human Computer Interaction. Her research contributes to the MobileASL project. She focuses on making mobile video communication more accessible to deaf and hard-of-hearing people while considering video quality intelligibility tradeoffs. Website: http://bit.ly/jjtran