3-2016

Eyetracking Metrics Related to Subjective Assessments of ASL Animations

Matt Huenerfauth

Hernisa Kacorri
Carnegie Mellon University

2016
THE JOURNAL ON
TECHNOLOGY AND
PERSONS WITH
DISABILITIES

Eyetracking Metrics Related to Subjective Assessments of ASL Animations

Matt Huenerfauth
Rochester Institute of Technology
Golisano College of Computing and Information Sciences
matt.huenerfauth@rit.edu
Hernisa Kacorri
Carnegie Mellon University
Human Computer Interaction Institute
hkacorri@gmail.com

Abstract

Analysis of eyetracking data can serve as an alternative method of evaluation when assessing the quality of computer-synthesized animations of American Sign Language (ASL), technology which can make information accessible to people who are deaf or hard-of-hearing, who may have lower levels of written language literacy. In this work, we build and evaluate the efficacy of descriptive models of subjective scores that native signers assign to ASL animations, based on eye-tracking metrics.

Keywords

Eye-Tracking, Sign Language, Animation.

Introduction

Automatic synthesis of sign language animations can increase information accessibility for people who are deaf and use signing as a primary means of communication. In the US, this population is estimated to be over half a million (Mitchell et al. 328-329). Standardized testing has revealed that many US deaf adults have lower levels of English reading literacy (Traxler), and thus complexity in the reading level of the text on websites or media can be too high. Linguistically accurate and natural-looking animations of American Sign Language (ASL) that are automatically synthesized from an easy-to-update script would make it easier to add ASL content to websites and media.

Researchers must regularly evaluate whether animations are grammatically correct and understandable, often through participation of signers, e.g. (Gibet et al. 18-23; Kipp et al. 107-114; Schnepp et al. 250). We have previously proposed the use of eyetracking to evaluate participants' reactions to animations *without obtrusively directing their attention to any particular aspect of the animation* (Kacorri, Harper, and Huenerfauth, Comparing; Kacorri, Harper, and Huenerfauth, Measuring 549-559). In this work, through multiple regression analysis on data from a user study, we identify relationships between (a) eyetracking metrics defined on recorded eye movements of participants watching ASL animations and (b) the subjective scores on grammaticality, understandability, and naturalness that participants assigned to those animations.

Discussion

Eyetracking and Sign Language Animations

As discussed in (Kacorri, Lu, and Huenerfauth, 514-516; Huenerfauth and Kacorri), in the context of research on incorporating new capabilities into ASL animation technology, it is difficult to design experimental stimuli and questions to measure participants' comprehension of information content specifically conveyed by some new feature of an animation.

To address this concern, we examined research using eyetracking to unobtrusively probe where participants are looking during an experiment, which can allow researchers to infer the cognitive strategies of those users, e.g. (Jacob and Karn). In fact, researchers have used eyetracking with participants who are deaf to investigate comprehension of *videos* of humans performing sign language (Cavendar et al.; Muir and Richardson; Emmorey et al.), but not of sign language *animations*. In our prior work (Kacorri, Harper, and Huenerfauth, Comparing; Kacorri, Harper, and Huenerfauth, Measuring), we examined whether these eyetracking methods could be adapted to the evaluation of sign language animations. However, in this earlier work, we examined one-to-one correlation relationships between the evaluation scores that participants assigned to the stimuli (video and animations) and specific eyetracking metrics. In this paper we focus on sign language animations only, and we systematically investigate the contribution of *multiple metrics* in indicating the subjective responses that native signers assign to ASL animations via multiple regression modeling.

User Study and Collected Data

Participants. Eleven ASL signers were recruited using ads posted on New York City Deaf community websites: 4 men and 7 women of ages 24-44 (average age 33.4). Seven learned ASL since birth, three prior to age 4, and one learned ASL at age 8 (attending schools for the deaf with instruction in ASL until age 18 and continuing to use ASL at home and work).

Experiment. Participants viewed 21 short stories in ASL performed by an animated character, created by a native ASL signer using the VCom3D (2015) SignSmith animation tool; we previously shared these stimuli with the research community (Huenerfauth and Kacorri). The video size, resolution, and frame-rate for all stimuli were identical. During the study, after viewing a story, participants responded to 1-to-10 scalar-response questions about their subjective impression of the animation. All questions were presented onscreen (embedded in the stimuli interface) as HTML forms to minimize possible loss of tracking accuracy due to head movements of participants between the screen and a paper questionnaire on a tabletop. The following English question text was shown onscreen:

(a) Good ASL grammar? (10=Perfect, 1=Bad)

(b) Easy to understand? (10=Clear, 1=Confusing)

(c) Natural? (10=Moves like person, 1=Like robot)

An initial sample animation familiarized the participants with the experiment and the eye tracking system. All of the instructions and interactions were conducted in ASL; subjective questions were explained in ASL. Some introductory information about the study was conveyed via a video recording of a native ASL signer. As discussed in (Kacorri, Harper, and Huenerfauth, Comparing), participants were seated in front of an Applied Science Labs D6 desktop-mounted eye-tracker, which sat below a 19-inch computer screen at a typical viewing distance.

Eyetracking Metrics. We recorded eye-tracking data while the participant viewed each animation, and then participants answered the questionnaire. Since eyetrackers occasionally lose the tracking of the participant's eye (e.g., if the participant rubbed their face with their hand), we needed to filter out any eye-tracking data in which there was a loss of tracking accuracy, as discussed in (Kacorri, Harper, and Huenerfauth, Comparing). For analysis, we defined 4 areas of interest in our stimuli: the virtual signer's head/face, body (including hands), upper face, and lower face; eye fixations elsewhere were coded as "off." Based on these areas of interest, we describe a participant's eye movements during each animation with 28 eyetracking metrics.

Table 1. Eyetracking Metrics

Category	Eyetracking Metrics
Total Fixation Time: *duration when the eyes are on this area of interest*	BodyTotalFixTime, UpperFaceTotalFixTime, LowerFaceTotalFixTime, FaceTotalFixTime
Proportional Fixation Time: *percentage of time with the eyes on this area of interest*	PercentFaceFix, PercentUpperFaceFix, PercentLowerFaceFix
Proportional Fixation Time (discounting "Off" time): *same as above, but the fixation time spent "off" is not included in denominator*	PercentFaceFixNoOff, PercentUpperFaceFixNoOff, PercentLowerFaceFixNoOff
Transitions: *count of the movements of the eyes from one area of interest to another*	NumFaceToBody, NumBodyToFace, NumBodyToOff, NumOffToBody, UpperFaceToBody, LowerFaceToBody, UpperFaceToLowerFace, LowerFaceToUpperFace, BodyToUpperFace, BodyToLowerFace, OffToUpperFace,
Proportional Transitions: *same as above, but normalized by the total time duration of the stimulus*	NormFaceToFromHands, NormUpperFaceToFromHands, NormLowerFaceToFromHands, NormUpperFaceToFromLowerFace
Overall: *counts of transitions or length of the eye movement trail*	NumTotalTran, TotalDetailedTrans, NormTrailDistance

Results and Analysis

The goal of our analysis is to examine how eye movements of participants relate to their responses to subjective questions about ASL animations. In addition, we wanted to know which eyetracking metrics best capture variance in score for each of the subjective questions evaluating the grammar, understandability, naturalness of the animations. We therefore used multiple regression to analyze the data. Our independent variables included all of the eyetracking metrics, listed in the above table. We trained a separate model for each of our dependent variables (Grammar, Understand, and Natural).

Since we have calculated many eyetracking variables, it was important to explore combinations of variables in a systematic manner. We used the 'leaps' package (Lumley) to build models of all possible subsets of features to identify the model with the highest adjusted R-squared value, i.e. the percentage of total variability accounted for by the model.

For a meaningful interpretation of the relative contribution of each of the eyetracking metrics, we calculated the relative importance of each independent variable in the Grammar, Understand, and Natural models, using the Linderman-Merenda-Gold (LMG) metric (Lindeman, Merenda, Gold), using the 'relaimpo' package (Grömping). This analysis assigns an R-squared percent contribution to each correlated variable obtained from all possible orderings of the variables in the regression model. Higher bars in Figures 1-3 indicate that the metric had greater importance in the model. We employed bootstrap to estimate the variability of the obtained relative importance value, to determine 95% confidence intervals (whiskers in the graphs). Importance values may be considered significant when whiskers do not cross the zero line in the graph. As illustrated by Figures 1-3, we see that the eyemetrics relating to the 'Head/Face' area of interest features prominently in many of the best models.

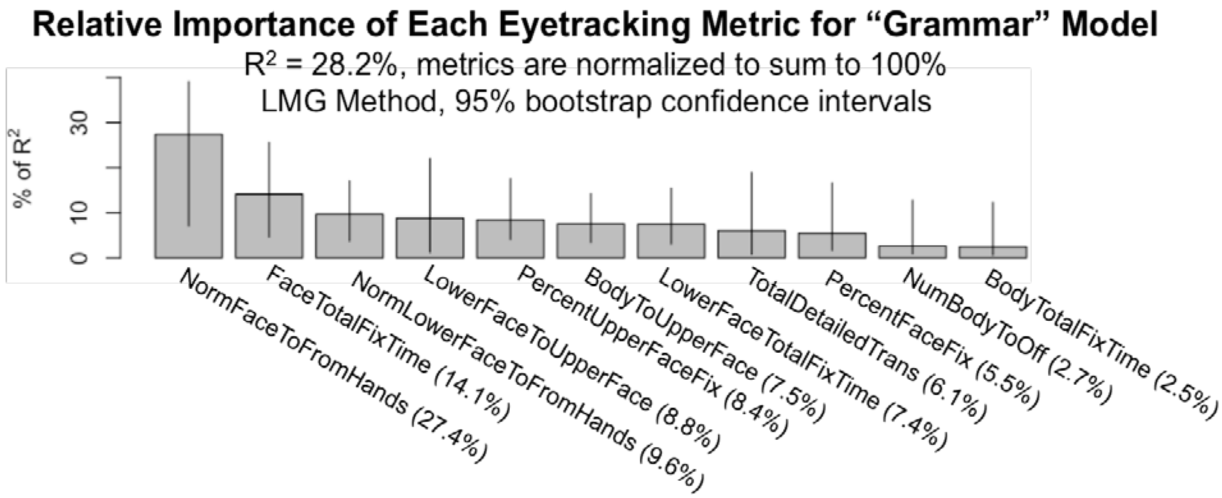


Fig. 1. Relative importance of each eyetracking metric in the model with the highest R-squared value (28.2%) for the "Grammar" subjective response score; the most important metrics include: NormFaceToFromHands and FaceTotalFixTime.

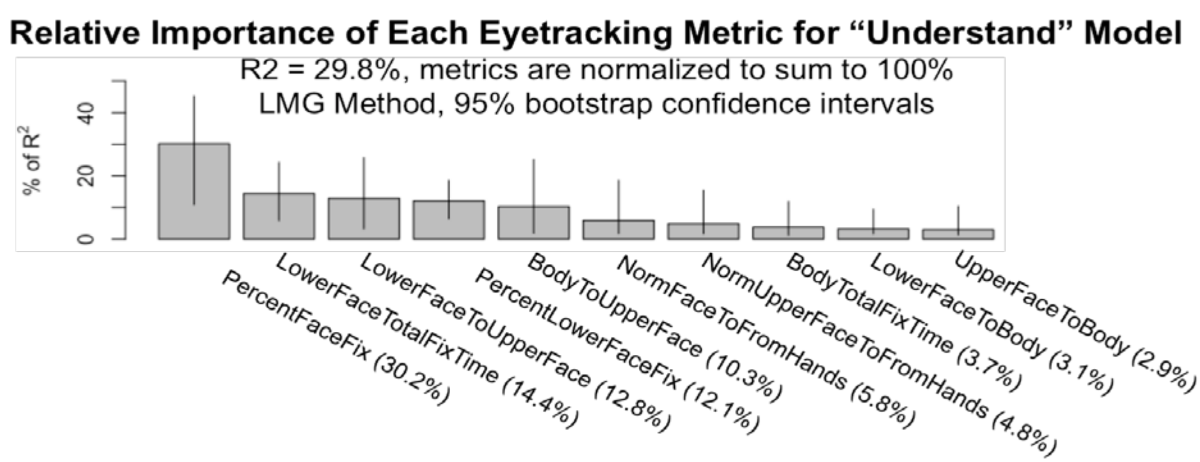


Fig. 2. Relative importance of metrics in the model with highest R-squared value (29.83%) of the "Understand" subjective response score; the most important metrics include: PercentFaceFix and LowerFaceTotalFixTime.

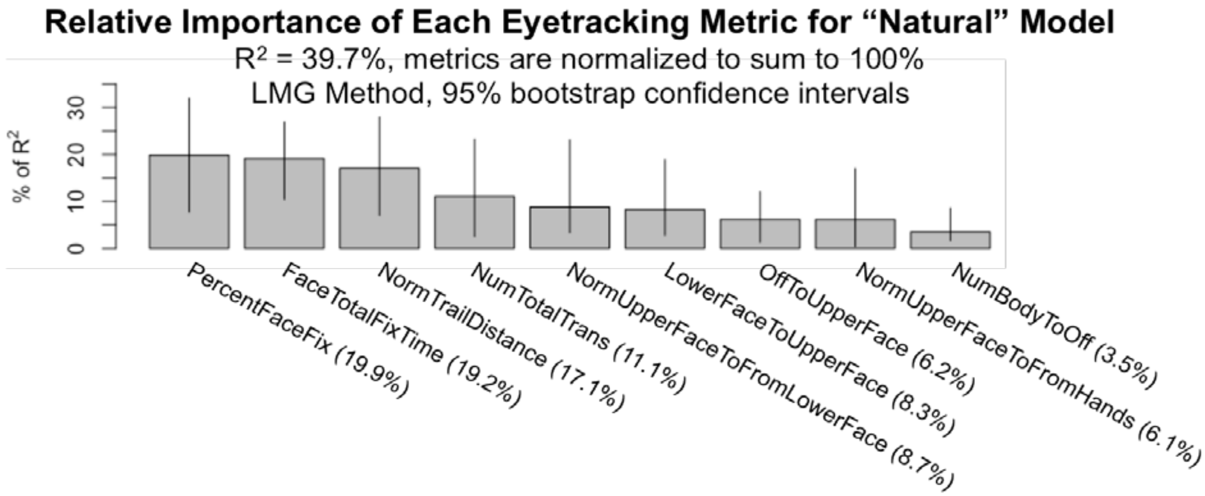
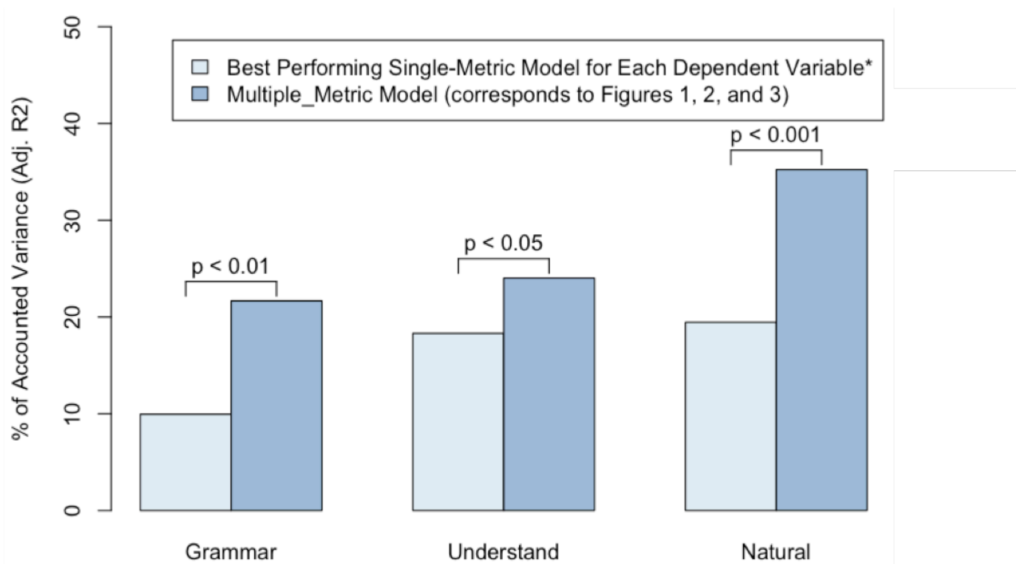


Fig. 3. Relative importance of each eyetracking metric in the model with the highest R-squared value (39.7%) for the "Natural" subjective response score; the most important metrics include: PercentFaceFix and FaceTotalFixTime.



* The best single-metric model for Grammar uses the **NormFaceToFromHands** eyetracking metric. The best single-metric model for Understand uses **FaceTotalFixTime**, and the best for Natural uses **PercentFaceFix**.

Fig. 4. Comparison of the best multiple-metric regression model and best single-metric regression model for each of the subjective response scores.

In order to determine whether these multiple-metric models outperformed single-metric models (as we had explored in earlier work), for each of the subjective scores we build a model using a single eyetracking metric (chosen by ‘leaps’ as the one yielding the highest adjusted R-squared value). As shown in Figure 4, we found that in each case, the single-metric model accounts for significantly less variance than the multiple-metrics model (ANOVA, $p<0.05$).

Conclusions

We have offered guidance on which eyetracking metrics can be used to predict ASL signers’ subjective judgments about the grammaticality, understandability, and naturalness of sign language animations. Future researchers who need to unobtrusively collect subjective judgments about sign language animations can use regression models based on these metrics.

Works Cited

Cavender, Anna, E. A. Rice, K.M. Wilamowska. SignWave: Human Perception of Sign Language Video Quality as Constrained by Mobile Phone Technology. Retrieved from http://courses.cs.washington.edu/courses/cse510/05sp/project-reports/cse510-signwave.pdf

Emmorey, K. and Thompson, R. Colvin. Eye gaze during comprehension of American Sign Language by native and beginning signers. J Deaf Stud Deaf Educ. 14, 2, 237-43, 2009.

Gibet, Sylvie, Nicolas Courty, Kyle Duarte, and Thibaut Le Naour. "The SignCom System for Data-driven Animation of Interactive Virtual Signers: Methodology and Evaluation." ACM Transactions on Interactive Intelligent Systems (TiiS) 1.1 (2011): 6.

Grömping, U. Relative importance for linear regression in R: the package relaimpo. Journal of statistical software 17(1): 1-27. 2006.

Huenerfauth, Matt, and Hernisa Kacorri. "Release of Experimental Stimuli and Questions for Evaluating Facial Expressions in Animations of American Sign Language." Proceedings of the 6th Workshop on the Representation and Processing of Sign Languages: Beyond the Manual Channel, The 9th International Conference on Language Resources and Evaluation (LREC 2014), Reykjavik, Iceland. 2014.

Jacob, R. J. K. and K. S. Karn. Eye Tracking in Human-Computer Interaction and Usability Research: Ready to Deliver the Promises. The Mind's Eye (First Edition). In: Hyönä, J., Radach, R., and Deubel, H. (eds.). Amsterdam: 573-605, 2003.

Kacorri, Hernisa, Allen Harper, and Matt Huenerfauth. "Measuring the Perception of Facial Expressions in American Sign Language Animations with Eye Tracking." Universal Access in Human-Computer Interaction. Lecture Notes in Computer Science 8516 (2014): 549-59. Switzerland: Springer International Publishing, 2014.

Kacorri, Hernisa, Allen Harper, and Matt Huenerfauth. "Comparing Native Signers Perception of American Sign Language Animations and Videos via Eye Tracking." Proceedings of the

15th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS 2013). New York: ACM Press, 2013.

Kacorri, Hernisa, Pengfei Lu, and Matt Huenerfauth. “Evaluating Facial Expressions in American Sign Language Animations for Accessible Online Information.” Universal Access in Human-Computer Interaction. Design Methods, Tools, and Interaction Techniques for eInclusion, Lecture Notes in Computer Science 8009 (2013): 510-19. Switzerland: Springer International Publishing, 2013.

Kipp, Michael, Quan Nguyen, Alexis Heloir, and Silke Matthes. "Assessing the Deaf User Perspective on Sign Language Avatars." Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS 2011). New York: ACM Press, 2011.

Lindeman, R.H., P.F. Merenda, R.Z. Gold. Introduction to bivariate & multivariate analysis. Scott Foresman, Glenview, IL. 1980.

Lumley, T. Leaps: Regression subset selection. R package version 2.9. 2009.

Mitchell, Ross E., Travas A. Young, Bellamie Bachleda, and Michael A. Karchmer. “How Many People Use ASL in the United States? Why Estimates Need Updating.” Sign Language Studies 6.3 (2006): 306-35. Gallaudet University Press. 2006.

Muir, L.J., and I.E. Richardson. Perception of sign language and its application to visual communications for deaf people. J Deaf Stud Deaf Educ 10, 4, 390-401, 2005.

Schnepp, Jerry C., Rosalee J. Wolfe, John C. McDonald, and Jorge A. Toro. "Combining emotion and facial nonmanual signals in synthesized american sign language." Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS 2012). New York: ACM Press, 2013.

Traxler, Carol Bloomquist. "The Stanford Achievement Test: National Norming and Performance Standards for Deaf and Hard-of-hearing Students." Journal of Deaf Studies and Deaf Education 5.4 (2000): 337-348. Oxford University Press. 2000.

Journal on Technology and Persons with Disabilities

ISSN 2330-4216

LIBRARY OF CONGRESS * U.S. ISSN CENTER
ISSN Publisher Liaison Section
Library of Congress
101 Independence Avenue SE
Washington, DC 20540-4284
(202) 707-6452 (voice); (202) 707-6333 (fax)
issn@loc.gov (email); www.loc.gov/issn (web page)