Effect of Ranking and Precision of Results on Users' Satisfaction with Search-by-Video Sign-Language Dictionaries

Saad Hassan, Oliver Alonzo, Abraham Glasser, and Matt Huenerfauth

Rochester Institute of Technology, Rochester NY 14623, USA
{sh2513,oa7652,atg2036,matt.huenerfauth}@rit.edu

Abstract. Advances in sign-language recognition technology can enable users of American Sign Language (ASL) dictionaries to search for a sign, whose meaning is unknown, by submitting a video of themselves performing the sign they had encountered, based on their memory of how it appeared. However, the relationship between the performance of sign-recognition technology and user satisfaction of such search interaction is unknown. In two Wizard-of-Oz experimental studies, we found that in addition to the position of the desired word in a list of results, the similarity of the other words in the results list also affected user satisfaction.

Keywords: American Sign Language (ASL); Dictionary; Search.

1 Background and Introduction

In addition to people who are Deaf and Hard of Hearing (DHH), other groups are interested in learning ASL and may benefit from related technologies, including students in ASL classes [5], parents of DHH children [7], teachers in ASL/English bilingual programs [6]. While it is straightforward to look up an English word in a bilingual ASL-English dictionary, it is more difficult for students to look up an unfamiliar ASL sign whose meaning is unknown. Without a common writing system for ASL, students cannot type a text string to search for a sign. Instead, the user may need to use a query interface to select linguistic properties of the sign, e.g. handshape or movement, which may be challenging for students [4].

Advancement in sign-recognition technology can enable users to submit a video (of themselves or a clip of someone else performing the desired ASL sign) to search for a matching sign in an ASL dictionary collection, e.g. [3]. However, given the challenges inherent in recognizing the 3D human motion and complex linguistic features of ASL, e.g. [10], sign-language recognition technologies are still improving. Moreover, poor lighting while recording a sign, camera motion, diverse camera viewpoints, occlusions, poor video quality, and cluttered background can also reduce the accuracy of these technologies [9]. Thus, even if a user of an ASL dictionary has remembered the appearance of an ASL sign and has attempted to produce it as faithfully as possible when creating their video

query, the system may not return the desired sign as the top result. Instead, the user may need to browse a list of possible "matches" to find the word they seek.

Given the challenge of sign-recognition technology and the affect of mis-recognition on the experience of users of dictionary systems, there is a need identify properties of the output of ASL dictionary match algorithms that affect users' opinion of the system's quality. Such findings would inform designers of match algorithms which characteristics to optimize – in order to determine whether a dictionary-search experience would have sufficient usability. To investigate these issues for a video-based-search ASL dictionary, we have created a prototype dictionary system to support experimental studies with ASL students (or other potential users). Some of the findings presented in this paper were originally presented at the 21st International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS'19) [2].

2 Experimental Study and Results

For greater control, our study used a Wizard-of-Oz method, in which the underlying technology is mimicked [8]. [2] describes our web-based prototype that simulated an ASL dictionary in which users submit a self-video performing a sought-after sign and view search results (Figure 1). Study participants were shown an ASL sign (likely to be unfamiliar to new signers), e.g. RAINBOW or CIGARETTE, and were asked to imagine they encountered this sign and did not know its meaning. Next, participants viewed a screen displaying a live webcam view, where they were asked to performed the sign (based on their memory) to submit a video as a query to search for the sign in the dictionary. Next, participants saw a scroll-able "View results" page displaying 100 results, similar in style to the results page of a video search engine. Six results were visible at a time, prior to the user scrolling to view more. While we had asked the participant to submit a video to simulate a search query, our prototype did not actually use recognition technology. We predetermined the set of results, to seem like realistic matches to the query, sorted based on similarity to the desired sign, following a protocol for creating and sorting the results list described in [2]. Participants were asked to browse the results to find the item that seemed like the best match for the sought-after sign (and to note this on a separate response form). Although the matching sign was always present in the results list, a participant could write "not found" if they did not believe a matching sign was displayed. After each query, participants answered questions adapted from [1], to rate their **satisfaction with the way results are ranked** on a Likert scale from *strongly disagree* to *strongly agree* and **their perceived relevance of the results** on a ternary scale: *highly relevant, relevant, not relevant.*

In a first study (Placement Study), the position rank of the desired result (k) was varied k = 1, 5, 10, 20. When the desired word was closer to the top of the results, users' satisfaction with the results ranking and perceived relevance of the overall results was higher, as illustrated in Figure 2 and 3.

In a follow-up study (Precision Study), the rank position of the sought-after sign was held constant at (k=10 +/- 1) and the precision was varied of the surrounding results (the other 99 signs on the list). In the "high" precision level, the surrounding words were extremely similar to the desired word. In "medium," a random sequence of signs containing an even mix of signs at near-the-face and near-the-torso location were used, and in "low," the surrounding signs had a different handshape and a different location (than the desired sign). Figure 4 and 5 show the effect of precision on users' responses.

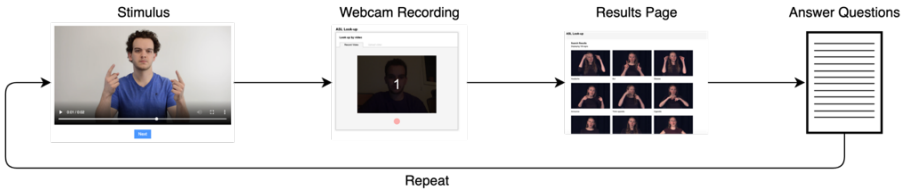

Fig. 1. Flowchart of the procedure our participants followed during the experiment

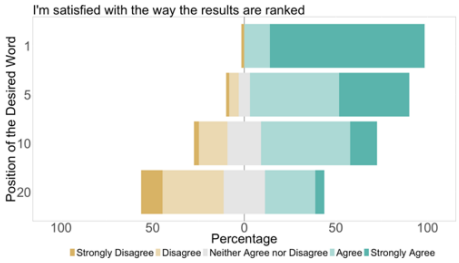

Fig. 2. Users' satisfaction with the way the results were ranked in the placement study.

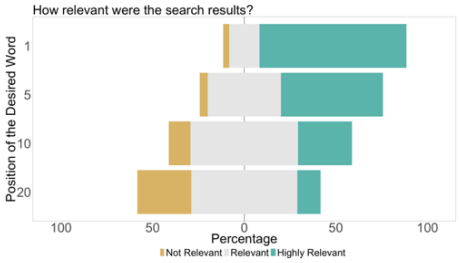

Fig. 3. Users' judgements of the relevance of the results in the placement study.

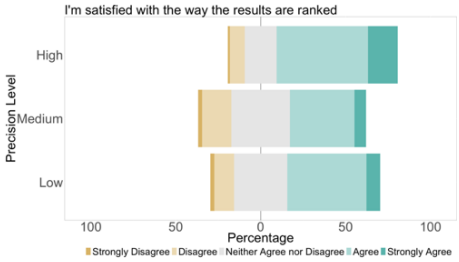

Fig. 4. Users' satisfaction with the way the results were ranked in the precision study.

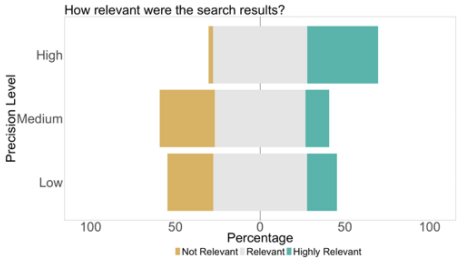

Fig. 5. Users' judgements of the relevance of the results in the precision study.

3 Conclusion and Future Work

Overall, our findings provide guidance for researchers studying sign-language dictionary search systems or for researchers who are developing underlying technologies, e.g. sign recognition from video. Specifically, we investigated whether users' judgements of the quality of an ASL dictionary search system vary depending on the placement of the desired word in the list of search results and the precision of the results list (the similarity of the other words on the list to the desired word). In future work, a two-factor study (placement and precision) with more participants and more realistic search scenarios could allow us to understand any interactions between these variables.

References

1. Al-Maskari, A., Sanderson, M., Clough, P.: The relationship between ir effectiveness measures and user satisfaction. In: 30th Intl. ACM SIGIR Conference. p. 773–774. SIGIR '07 (2007), https://doi.org/10.1145/1277741.1277902
2. Alonzo, O., Glasser, A., Huenerfauth, M.: Effect of automatic sign recognition performance on the usability of video-based search interfaces for sign language dictionaries. In: ACM Conf. on Comp. and Accessibility. pp. 56–67. ASSETS '19 (2019), http://doi.acm.org/10.1145/3308561.3353791
3. Elliott, R., Cooper, H., Glauert, J., Bowden, R., Lefebvre-Albaret, F.: Search-by-example in multilingual sign language databases (2012)
4. Lapiak, J.: Handspeak. https://www.handspeak.com/
5. Looney, D., Lusin, N.: Enrollments in languages other than english in united states institutions of higher education, summer 2016 and fall 2016. 2018 Mod. Lang. Assoc. of America (2018)
6. Marschark, M., Leigh, G., Sapere, P., Burnham, D., Convertino, C., Stinson, M., Knoors, H., Vervloed, M., Noble, W.: Benefits of sign language interpreting and text alternatives for deaf students' classroom learning. J. of deaf studies and deaf edu. (2006). https://doi.org/10.1093/deafed/enl013
7. Mitchell, R., Karchmer, M.: Chasing the mythical ten percent: Parental hearing status of deaf and hard of hearing students in the united states. Sign Lang. Studies (2004). https://doi.org/10.1353/sls.2004.0005
8. Pettersson, J.S., Wik, M.: The longevity of general purpose wizard-of-oz tools. In: OzCHI '15 (2015), https://doi.org/10.1145/2838739.2838825
9. Reddy, K.K., Shah, M.: Recognizing 50 human action categories of web videos (2013), https://doi.org/10.1007/s00138-012-0450-4
10. Yanovich, P., Neidle, C., Metaxas, D.: Detection of major ASL sign types in continuous signing for ASL recognition. In: Conf. on Lang. Res. and Eval. https://www.aclweb.org/anthology/L16-1490