Designing and Experimentally Evaluating a Video-based American Sign Language Look-up System

Saad Hassan
Golisano College of Computing and Information Sciences
Rochester Institute of Technology
New York, USA
sh2513@rit.edu

ABSTRACT

Despite some prior research and commercial systems, if someone sees an unfamiliar American Sign Language (ASL) word and wishes to look up its meaning in a dictionary, this remains a difficult task. There is no standard label a user can type to search for a sign, and formulating a query based on linguistic properties is challenging for students learning ASL. Advances in sign-language recognition technology will soon enable the design of a search system for ASL word look-up in dictionaries, by allowing users to generate a query by submitting a video of themselves performing the word they believe they encountered somewhere. Users would then view a results list of video clips or animations, to seek the desired word. In this research, we are investigating the usability of such a proposed system, a webcam-based ASL dictionary system, using a Wizard-of-Oz prototype and enhanced the design so that it can support sign language word look-up even when the performance of the underlying sign-recognition technology is low. We have also investigated the requirements of students learning ASL in regard to how results should be displayed and how a system could enable them to filter the results of the initial query, to aid in their search for a desired word. We compared users' satisfaction when using a system with or without post-query filtering capabilities. We discuss our upcoming study to investigate users' experience with a working prototype based on actual sign-recognition technology that is being designed. Finally, we discuss extensions of this work to the context of users searching datasets of videos of other human movements, e.g. dance moves, or when searching for words in other languages.

CCS CONCEPTS

• **Human-centered computing** → *Interaction design process and methods*; • **Information systems** → *Search interfaces*; *Multimedia and multimodal retrieval*; *Video search.*

KEYWORDS

Sign Language Dictionary; ASL; Video-based Search; Hybrid Search; Searching Behavior; Search Interface Design

CHIIR '22, March 14–18, 2022, Regensburg, Germany

ACM ISBN 978-1-4503-9186-3/22/03.
https://doi.org/10.1145/3498366.3505804

ACM Reference Format:
Saad Hassan. 2022. Designing and Experimentally Evaluating a Video-based American Sign Language Look-up System. In *ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR '22), March 14–18, 2022, Regensburg, Germany*. ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3498366.3505804

1 MOTIVATION AND PRIOR WORK

American Sign Language (ASL) is used by about 500,000 Deaf or Hard of Hearing (DHH) people in the U.S. [15]. There is also a growing population of hearing second language students who are trying to learn ASL, which is reflected in recent enrollment data reporting 100,000 to 200,000 students enrolled in ASL classes in Fall and Summer 2016 [12]. Other groups, including parents of Deaf or Hard of Hearing children [14], hearing adults who are becoming deaf, and teachers in ASL/English bilingual programs can also benefit from learning ASL [13]. However, compared to educational resources for written languages, ASL dictionaries are currently less usable for students. Unlike written languages, if someone encounters an unfamiliar ASL sign, they cannot type a text string to search for it because there is: (a) no standard ASL writing system, (b) no one-to-one correspondence between English words and ASL signs, and (c) no standard convention for English-based labelling of signs. Thus, to search for a sign, the user must recall linguistic properties of the sign, e.g., physical features such as handshape or movement – which is challenging for people who are learning ASL [11].

Existing sign-language lookup systems can be broadly classified into two types *search by feature selection* (which includes printed ASL dictionaries) and *search by example. Search by feature selection* allows users to create a query by selecting a set of linguistic features of the desired sign [6, 11, 19, 20]. *Search by example* enables users to search for a sign by demonstrating its motion in front of a color or depth camera, or by wearing motion-capture sensors or gloves [6, 19]. Users still face **challenges with both types**. *Search by feature selection dictionaries* remain awkward to use due to their complex interfaces that require users to provide a lot of linguistic features of the sign, which may only be vaguely remembered [21], especially by people who are just learning a sign language. Further, many systems have poor feature-to-sign matching or cumbersome interfaces [4]. Based on computer-vision technologies, researchers have begun to build systems that automatically analyze a video of a sign and seek a match for this sign in a dictionary collection [7]. However, sign recognition from video is difficult and current technology is imperfect [22], especially for videos of non-fluent student signers. Moreover, poor video quality due to lighting, camera motion or angle, occlusions, or cluttered backgrounds can also reduce recognition accuracy [17]. Thus, users may need to browse

a list of possible "matches" to find the word they seek. There is a need to not only improve the underlying sign-matching technology but also human-computer interaction research to identify the most effective user-interface despite current limits. To investigate these issues for a video-based-search ASL dictionary, we created a prototype dictionary to support experimental studies with ASL students or other to investigate settings of key design and performance variables. Contributions of this research include:

- We investigate how the output of the underlying sign-recognition system affects the usability of dictionary search, i.e., how usability is affected by the ranking of the desired sign on the results list, whether it appears "above-the-fold" (visible on-screen before the user must scroll-down), and precision of other list items.
- We then explore how to best present individual result snippets inside the results page and how to make the results page interactive for users, e.g., with sorting and filtering capabilities, so that users can retrieve their desired result quickly even if it does not appear in the top few results.
- We propose a summative study that investigates how users will actually interact with an ASL dictionary system based on an underlying imperfect sign recognition technology.
- Proposed studies also will explore how our findings may inform word look-up for dictionaries of languages with deep orthography (weak correspondence between sounds and writing) or datasets of short videos of human performance.

2 RESEARCH QUESTIONS

Our literature review has revealed the need for better approaches for searching in ASL dictionaries and gaps in knowledge that motivate the following research questions:

RQ1 How does the performance of automatic ASL recognition technology affect the usability of an ASL dictionary search system that depends upon it?

RQ2 How much text or video content should be displayed by default in the search-results list of an ASL dictionary search system for students learning ASL?

RQ3 What linguistic or appearance features of ASL signs should be included in a filtering user-interface in a hybrid-search system, and what terminology for referring to these features would be most understandable to these students?

RQ4 If users are given the opportunity to perform a series of ASL-sign search tasks using a search-by-video or a hybrid-search prototype, is there a difference in user-satisfaction between prototypes, as measured by:
(1) satisfaction with the search experience?
(2) perception of usefulness of results?
(3) satisfaction with the way the results are ranked?
(4) perception of their degree of control with the system?
(5) time taken to complete the search task?
(6) ability to find the correct result?

RQ5 How do users interact with a working hybrid ASL dictionary system when performing tasks using dictionary search?

RQ6 How do our findings from formative studies on ASL dictionaries generalize to search systems for
(1) dictionaries for languages with deep orthography?
(2) datasets of videos of human action?

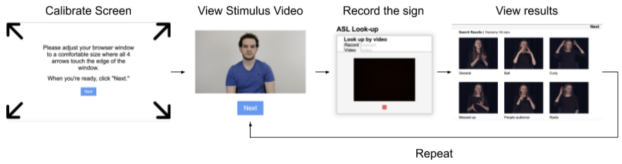

Figure 1: Prototype and Experimental Protocol

3 RESEARCH METHOD COMMON ACROSS STUDIES

Given current limits in accuracy or in vocabulary size of video-based ASL sign look-up technologies, Wizard-of-Oz methods (in which the underlying technology is mimicked) enable quickly testing user-interfaces [16]. Our experimental investigation of RQ1, RQ2, RQ3, and RQ4 used a web-based prototype that simulated users submitting a video of themselves performing a sought-after sign and viewing search results (Figure 1). The prototype was displayed on a computer screen to participants who met with a researcher. Participants were first shown a video of an ASL sign unlikely to be unfamiliar to new ASL learners, e.g., RAINBOW, and asked to imagine they had encountered this sign somewhere and did not know what it meant. They pressed "next" to advance to a screen displaying a live webcam view, where they performed the sign (based on their memory of how it had looked) to submit a video of themselves, which served as a query to search for it in the dictionary. The purpose of making the users record the sign was to make them feel that they are making an actual query, but this prototype did not actually use any automated recognition technology to generate results, which had instead been curated for presentation to participants regardless of the actual video they had submitted. After submitting a video on the "Record the sign" screen (Figure 1), participants were shown a "View results" screen with a scroll-able results page displaying 100 results for their query. This page mimicked the style of video search engines, and exactly six results items were visible on the screen at one time, prior to the user scrolling to view more. The set of results was determined by researchers in our lab [2], and the set of items on the results page was intended to seem like a realistic set of matches to the participant's query, sorted based on similarity to the desired sign, to simulate the output of an actual automatic system, following the protocol for creating and sorting the results list in prior work [8].

The prototype described above has been used in studies to investigate the first four research questions. In upcoming work, for RQ5 and RQ6, we will develop a functioning system and other Wizard-of-Oz prototypes, as discussed below. Participants for the studies described below were recruited primarily from introductory ASL courses and received $40 compensation for participation.

4 PROGRESS MADE

4.1 Study 1: Preferences of ASL Learners

The goal of this study was to guide our design of a video-based ASL dictionary search prototype that of sufficient usability to be

an effective research vehicle. Prior work in information retrieval revealed how important the **design of the search results page** is for the success of video search systems, motivating our investigation of search-results-page design for ASL dictionaries [1, 10, 18]. However, this prior work helped us identify different design factors that need to be optimized for an ASL dictionary system, there are differences between web-search and general video search that do permit directly generalizing the findings to this new context. We therefore conducted semi-structured interviews with 32 users of ASL dictionaries. We showed them the prototype (described above) and then asked questions about how to present text or video content on the search-results list of an ASL dictionary search system. We iteratively refined of the prototype design after the first 16 participants, to adjust the appearance of the search results page. We qualitatively analyzed our data by coding transcripts of participants' responses to the the semi-structured interview questions to identify various themes. Patterns observed from the interview data were investigated using Braun and Clarke's thematic analysis method [5]. We found that users preferred greater information density, i.e., with textual meta-data appearing near items to indicate linguistic properties, and for the video clips of signs to auto-play in a looping manner so that users could quickly browse the set of results without needing to click on each.

4.2 Study 2: Comparing Search-by-Video and Hybrid Search

Prior research mentioned in the first section on sign-language dictionaries had examined *search-by-feature* [20] and *search-by-video* [3, 7] approaches for enabling a user to look up the meaning of an unfamiliar sign they had seen. However, **no prior work had examined the combination of these two approaches**. We therefore proposed a **hybrid-search** approach, in which users search for a sign using video input and then sort the results using filters. To investigate the potential of of a hybrid search approach (as discussed about in our Prior Work section), we conducted a lab-based experimental study with 20 participants, half of whom interacted with a *hybrid-search* version of our Wizard-of-Oz prototype, and half of whom interacted with a **search-by-video** version (which lacked any search-by-feature post-querying filtering interface). We performed statistical difference testing to compare each group of participants' speed, accuracy in finding the desired sign, and their subjective impression of the experience as measured by various question items measuring satisfaction and sense of control. Our study revealed that, compared to search-by-video, users of our hybrid-search approach had higher satisfaction and sense of control, and when the desired word appeared farther down the initial list of results, the filtering capabilities of the interface allowed them to find the desired item more quickly. The findings from both study 1 and study 2 were recently conditionally accepted for publication at CHI 2022.

4.3 Study 3: Effect of Sign-Recognition Performance on Usability of ASL Dictionaries

While the first two studies guided and evaluated the interface design, other factors may affect the usability of a sign-language dictionary search system, e.g., the accuracy of the underlying sign-recognition technology that analyzes the user's submitted video. To investigate this factor, we conducted several experiments to investigate how variations in the performance of of the matching technology would affect users' satisfaction with an ASL dictionary search systems. We made use of the Wizard-of-Oz prototype discussed above, and we engineered various sets of results to display to users of the prototype that mimicked different levels of accuracy in the set of signs returned as search results from the system, i.e., with some sets of results including the desired word near the top of the results list or farther down the list. In some sets of results, the signs surrounding the desired word were relatively similar, i.e., they were "close but incorrect matches," while in other cases, they were less similar to the desired item. Through curation of the results shown to users after their search-by-video, were investigated three properties of the underlying recognition system: rank-placement of the desired result in the search-results list, precision of the surrounding results, and whether the desired result lied on the first page before the users needed to scroll down to view results. All three factors affected users' satisfaction, and we identified numerical thresholds of performance needed to achieve specific levels of user satisfaction in this context. An article on our findings was accepted in the *ACM Transactions on Accessible Computing (TACCESS)* journal [9].

5 FUTURE PLANS

5.1 Proposed Study 4: Evaluation of a Working Prototype

We are currently collecting and annotating individual videos of ASL signs which will be used to train a word-level sign-recognition model. We plan on conducting a user-based evaluation of our recognition technology using our final prototype design, to determine how hybrid-search approaches can enable users to search dictionaries using imperfect sign-recognition video analysis technology. To evaluate user satisfaction and performance more holistically, we plan on designing tasks similar to those performed by new ASL signers when learning the language, e.g., completing a homework assignment consisting of watching a video of ASL and answering comprehension questions. This task-based evaluation will enable us to understand how ASL learners use these dictionaries and the extent to which they are successful. Moreover, it would also be interesting to explore how users respond to the system when it does not return their desired word, when it is based on fully automatic sign-recognition technology.

5.2 Proposed Studies: Hybrid Search for Other Scenarios

We also plan to investigate whether hybrid-search approaches are useful in other scenarios, beyond ASL dictionaries.

5.2.1 Orthographically deep spoken languages. Searching for the meaning of an unfamiliar word is also a challenge for orthographically deep spoken languages, in which there is little sound-to-writing correspondence, e.g., Mandarin Chinese. Fundamentally, hybrid-search consists of (1) *performing* a word to search for it using AI-based matching technology and (2) *filtering* the set of results. While, for ASL, this first step used video-based recognition,

for spoken languages, it could consist of search by voice (through automatic speech recognition of the user's pronunciation) or search by handwriting (through automatic character recognition, e.g., for written Mandarin Chinese)—followed by a filtering step. It would be interesting to evaluate whether users learning these other languages would benefit from hybrid-search approaches through a study evaluating a Wizard-of-Oz prototype for a spoken/written language, similar to our ASL prototype.

5.2.2 Datasets of human-movement. Our findings for ASL dictionaries motivate research on enabling users to search within other collections of relatively uniform-duration videos of human-movement, such as martial arts moves, sports moves, or dance moves. As discussed in related work, there has been some prior work on search-by-video approaches in these contexts, i.e., in which a user would submit of video in which they perform the movement they partially remember. However, no prior work had examined hybrid-search approaches, and it is unknown whether our findings for ASL would directly apply, e.g., in regard to interface design, supporting users in submitting a good video, or combining a search-by-performance query with a search-by-feature filtering step. To investigate this, we will design a Wizard-of-Oz prototype for searching an existing dataset of short videos of human movement, and we will conduct a study with novice learners to understand whether hybrid-search approaches are beneficial in this context.

6 CONCLUSION

This doctoral consortium submission motivates research on users searching for the meaning of unfamiliar signs in a hybrid manner that blends search-by-video and search-by-feature approaches, contributing guidance about the interface design, performance thresholds of underlying technology, and potential extensions of this search approach to other contexts. This research will lead to scalable and searchable ASL dictionary systems, which could transform the ASL learning experience and support greater communication between Deaf and hearing individuals. More broadly, this work may inform designers of interfaces for video-search engines, especially for short videos of human movement. The CHIIR Doctoral Consortium would be a great opportunity to receive feedback from researchers with expertise beyond what I can obtain at my university, especially in the area of IR. I am planning to present my dissertation proposal at the end of Summer 2022, so discussing my plans and receiving feedback at this event would be well-timed, so that I may incorporate that feedback into my dissertation proposal. I am excited to network with other researchers working at the intersection of HCI and IR, and engaging with this community would motivate me in conducting studies on broader domains beyond ASL dictionary search.

ACKNOWLEDGMENTS

I would like to thank my advisor Dr. Matt Huenerfauth. This material is based upon work supported by the National Science Foundation under award No. 1763569.

REFERENCES

[1] Dan Albertson. 2010. Analyzing user interaction with the ViewFinder video retrieval system. *Journal of the American Society for Information Science and Technology* 61, 2 (2010), 238–252. https://doi.org/10.1002/asi.21257 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/asi.21257

[2] Oliver Alonzo, Abraham Glasser, and Matt Huenerfauth. 2019. Effect of Automatic Sign Recognition Performance on the Usability of Video-Based Search Interfaces for Sign Language Dictionaries. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility* (Pittsburgh, PA, USA) *(ASSETS '19)*. Association for Computing Machinery, New York, NY, USA, 56–67. https://doi.org/10.1145/3308561.3353791

[3] V. Athitsos, C. Neidle, S. Sclaroff, J. Nash, A. Stefan, A. Thangali, H. Wang, and Q. Yuan. 2019. Large Lexicon Project: American Sign Language Video Corpus and Sign Language Indexing/Retrieval Algorithms. (2019).

[4] Danielle Bragg, Kyle Rector, and Richard E. Ladner. 2015. A User-Powered American Sign Language Dictionary. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work and Social Computing* (Vancouver, BC, Canada) *(CSCW '15)*. Association for Computing Machinery, New York, NY, USA, 1837–1848. https://doi.org/10.1145/2675133.2675226

[5] V. Braun and V. Clarke. 2006. Using Thematic Analysis in Psychology. *Qualitative research in psychology* 3 (01 2006), 77–101. https://doi.org/10.1191/1478088706qp063oa

[6] Fabio Buttussi, Luca Chittaro, and Marco Coppo. 2007. Using Web3D Technologies for Visualization and Search of Signs in an International Sign Language Dictionary. In *Proceedings of the Twelfth International Conference on 3D Web Technology* (Perugia, Italy) *(Web3D '07)*. Association for Computing Machinery, New York, NY, USA, 61–70. https://doi.org/10.1145/1229390.1229401

[7] R. Elliott, H. Cooper, J. Glauert, R. Bowden, and F. Lefebvre-Albaret. 2012. Search-By-Example in Multilingual Sign Language Databases. (2012).

[8] Saad Hassan, Oliver Alonzo, Abraham Glasser, and Matt Huenerfauth. 2020. Effect of ranking and precision of results on users' satisfaction with search-by-video sign-language dictionaries. In *Sign Language Recognition, Translation and Production (SLRTP) Workshop-Extended Abstracts*. Computer Vision – ECCV 2020 Workshops.

[9] Saad Hassan, Oliver Alonzo, Abraham Glasser, and Matt Huenerfauth. 2021. Effect of Sign-Recognition Performance on the Usability of Sign-Language Dictionary Search. *ACM Trans. Access. Comput.* 14, 4, Article 18 (oct 2021), 33 pages. https://doi.org/10.1145/3470650

[10] L. Hollink, G. P. Nguyen, D. C. Koelma, A. T. Schreiber, and M. Worring. 2004. User Strategies in Video Retrieval: A Case Study. In *Image and Video Retrieval*, Peter Enser, Yiannis Kompatsiaris, Noel E. O'Connor, Alan F. Smeaton, and Arnold W. M. Smeulders (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 6–14.

[11] J. Lapiak. 2021. *Handspeak*. https://www.handspeak.com/

[12] D. Looney and N. Lusin. 2018. 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).

[13] M. Marschark, G. Leigh, P. Sapere, D. Burnham, C. Convertino, M. Stinson, H. Knoors, M. Vervloed, and W. Noble. 2006. 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

[14] R. Mitchell and M. Karchmer. 2004. 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

[15] R. Mitchell, T. Young, B. Bachleda, and M. Karchmer. 2006. How Many People Use ASL in the United States? Why Estimates Need Updating. *Sign Lang. Studies* 6 (2006). https://doi.org/10.1353/sls.2006.0019

[16] J. S. Pettersson and M. Wik. 2015. The Longevity of General Purpose Wizard-of-Oz Tools. In *OzCHI '15*. https://doi.org/10.1145/2838739.2838825

[17] K. K. Reddy and M. Shah. 2013. Recognizing 50 Human Action Categories of Web Videos. *Mach. Vision Appl.* https://doi.org/10.1007/s00138-012-0450-4

[18] Akari Shimono, Yuki Kakui, and Toshihiko Yamasaki. 2020. Automatic YouTube-Thumbnail Generation and Its Evaluation. In *Proceedings of the 2020 Joint Workshop on Mult. Artworks Analysis & Attractiveness Comput. in Multimedia* (Dublin, Ireland) *(MMArt-ACM '20)*. Association for Computing Machinery, New York, NY, USA, 25–30. https://doi.org/10.1145/3379173.3393711

[19] Nazif Can Tamer and Murat Saraçlar. 2020. Improving Keyword Search Performance in Sign Language with Hand Shape Features. In *Computer Vision – ECCV 2020 Workshops*, Adrien Bartoli and Andrea Fusiello (Eds.). Springer International Publishing, Cham, 322–333.

[20] R. A. Tennant and M. G. Brown. 2010. *ASL handshape dictionary*. Gallaudet University Press.

[21] Haijing Wang, Alexandra Stefan, Sajjad Moradi, Vassilis Athitsos, Carol Neidle, and Farhad Kamangar. 2012. A System for Large Vocabulary Sign Search. In *Trends and Topics in Computer Vision*, Kiriakos N. Kutulakos (Ed.). Springer Berlin Heidelberg, Berlin, Heidelberg, 342–353.

[22] P. Yanovich, C. Neidle, and D. Metaxas. 2016. 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