Design and Evaluation of Hybrid Search for American Sign Language to English Dictionaries: Making the Most of Imperfect Sign Recognition

Saad Hassan*
Akhter Al Amin*
sh2513@rit.edu
aa7510@rit.edu
Computing and Information Science
Rochester Institute of Technology
Rochester, NY, USA

Alexis Gordon
School of Information
Rochester Institute of Technology
Rochester, NY, USA
aag7593@rit.edu

Sooyeon Lee
School of Information
Rochester Institute of Technology
Rochester, NY, USA
slics@rit.edu

Matt Huenerfauth
School of Information
Rochester Institute of Technology
Rochester, NY, USA
matt.huenerfauth@rit.edu

ABSTRACT

Searching for the meaning of an unfamiliar sign-language word in a dictionary is difficult for learners, but emerging sign-recognition technology will soon enable users to search by submitting a video of themselves performing the word they recall. However, sign-recognition technology is imperfect, and users may need to search through a long list of possible results when seeking a desired result. To speed this search, we present a hybrid-search approach, in which users begin with a video-based query and then filter the search results by linguistic properties, e.g., handshape. We interviewed 32 ASL learners about their preferences for the content and appearance of the search-results page and filtering criteria. A between-subjects experiment with 20 ASL learners revealed that our hybrid search system outperformed a video-based search system along multiple satisfaction and performance metrics. Our findings provide guidance for designers of video-based sign-language dictionary search systems, with implications for other search scenarios.

CCS CONCEPTS

• **Human-centered computing** → *Accessibility systems and tools*;
• **Information systems** → *Search interfaces.*

KEYWORDS

Sign Languages, American Sign Language (ASL), Dictionary, Search Interfaces, Video Search, User Satisfaction, IR Effectiveness, Search Evaluation, Search System Design

ACM Reference Format:
Saad Hassan, Akhter Al Amin, Alexis Gordon, Sooyeon Lee, and Matt Huenerfauth. 2022. Design and Evaluation of Hybrid Search for American Sign Language to English Dictionaries: Making the Most of Imperfect Sign Recognition. In *CHI Conference on Human Factors in Computing Systems (CHI '22), April 29-May 5, 2022, New Orleans, LA, USA.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3491102.3501986

*Both authors contributed equally to this research.

CHI '22, April 29-May 5, 2022, New Orleans, LA, USA

ACM ISBN 978-1-4503-9157-3/22/04…$15.00
https://doi.org/10.1145/3491102.3501986

1 INTRODUCTION

Over 70 million Deaf and hard of hearing (DHH) people use one of the over 150 recognized sign languages throughout the world [22, 57]. In the U.S. alone, American Sign Language (ASL) is used by about 500,000 people as a primary form of communication [56]. Many hearing and DHH individuals are also motivated to learn sign languages in school or as adults; for example, to provide important language exposure to DHH children [32], many hearing parents or teachers of DHH children are motivated to learn ASL [66, 80]. More broadly, learning ASL can promote interactions between DHH and hearing people, to support greater inclusion, mutual understanding, and participation across society. ASL has one of the fastest growing enrollments among foreign-language classes [28], with 200,000 students [26].

Dictionaries that translate an unknown language to a known language are an important tool for language learners, and searching is easier for written languages, as the user can use text-search or alphabetical listing. In contrast, if someone encounters an unfamiliar sign in ASL or other sign languages, they cannot type a text string to search for it, as sign languages often lack a standard writing system [7, 37]. To search for a sign in most sign languages, a user must recall linguistic properties of the desired sign (e.g., hand configuration, orientation, location, movement, facial expressions

[12]) and specify a query, which is difficult for someone still learning ASL [10]. Thus, there are challenges with **search-by-feature** approaches to ASL dictionary search.

There have been recent technical advances in automatically analyzing a video of a sign to seek a match in a dictionary collection [15, 16, 24, 40, 61, 63, 83]. However, users also face challenges in such **search-by-video** dictionary systems, as these technologies are still imperfect, given: (a) the technical difficulty of recognizing linguistically complex 3D signs from 2D video [83]; (b) poor lighting, camera motion, or cluttered backgrounds [64]; or (c) learners who may struggle to accurately perform a sign that they are attempting to recall. As a result, the system may not return the desired word at the top of a list of results, and current search-by-video systems lack post-query refinement options to narrow the list. Users must browse a long list of possible "matches" to find the word they seek—if the word appears on the list at all.

Our aim in this paper is not to improve the artificial intelligence (AI) technology for analyzing a video of a user, but rather to enable users to **make better use of existing recognition technology** to accomplish their search task. We propose a new **hybrid-search** approach, which begins with a search-by-video step in which the user submits a video performing a sign as best as they can recall it, to produce a long list of potential results. A search-by-feature step is next, in which users use a filtering interface to select visual or linguistic properties of the desired word, to reduce the set of results displayed. This hybrid approach may mitigate the difficulty users face in browsing through a long list of results from a search-by-video system since the filtering can reduce the items shown. Further, since users do not need to formulate a search-by-feature query until after the set of results has already been reduced through the search-by-video step, users may not need to produce as specific a query in order to yield a manageable number of search results.

While post-results filtering and post-query refinement have been used in a variety of search systems, e.g., [29, 31, 42, 85], this has not been previously explored in the context of sign-language dictionary search. Thus, there is a need for human-computer interaction research to understand which design factors support effective search. While prior search-by-feature systems included query interfaces with linguistic features, since hybrid-search filtering operates on a smaller set (the video-query results, rather than the entire dictionary), it is unknown how to best select and present features in a filtering interface. In addition, the visual and text presentation of items on the results page must support users in making quick comparisons of results and informed decisions when filtering.

We present two studies with likely users of ASL dictionary search systems: students learning ASL. An interview-based study, with 32 participants, revealed users' preferences for the appearance of the search-results listing and filtering interface. These findings guided our iterative design of a Wizard-of-Oz prototype ASL dictionary hybrid-search system, which we then evaluated in a study with 20 participants, in comparison to a search-by-video system (i.e., without post-query filtering). Across various measures of task performance and user satisfaction, we observed significant benefits from hybrid search. We also qualitatively compared the experiences of users interacting with both systems and reflect on challenges that users faced, the extent to which hybrid search helped with overcoming them, and the generalizability of our findings.

The **contributions** of our work are empirical, based on understanding the preferences of potential users of ASL dictionary systems and evaluating their experience with our prototype:

(1) We identify users' preferences for how much text or video content should be displayed by default on the search-results list of an ASL dictionary search system, as well as users' preferences for the set of linguistic or appearance characteristics of ASL signs to include in a filtering user-interface in a hybrid search system.
(2) We present empirical evidence of improved user satisfaction with a hybrid-search system, as compared to a search-by-video system, as well as qualitative analysis of users' discussion of factors that affected their experience.

2 RELATED WORK

2.1 Prior Approaches to Searching for an Unfamiliar Sign

A recent survey gathered open-ended responses from ASL students about their use of ASL dictionaries, revealing that a usable and reliable web-based resource for ASL-sign lookup would be beneficial, as the few existing resources for seeking the English translation of an ASL word were rarely used [10].

Prior electronic dictionaries have been proposed for finding written-language translations of individual words, e.g., with users searching by features like handshape, location, or movement of the desired sign [81]. For ASL and other sign languages used internationally, several *search-by-feature* systems have been designed [1, 13, 27, 49, 55, 68, 74]. Research has investigated feature selection approaches [13], specialized notations or GUI elements for constructing a query [21, 71], or learning common ways in which users make mistakes when searching by feature so that the results of a query may be broadened [10]. Despite these advances, users face **challenges with search-by-feature ASL dictionaries**, which require them to specify a query of linguistic features of a sign, which they may only vaguely remember [77], with particular challenges for people who are just learning a sign language, who are less familiar with such features. Many systems may have poor feature-to-sign matching or cumbersome interfaces [10]. A hybrid-search approach mitigates these concerns by replacing the burdensome query-specification step with the user simply submitting a video; then the user may optionally filter the results based on features. Since users are filtering only within the search results, rather than within the entire dictionary, the user does not need to specify as many linguistic features in order to obtain a reasonably small number of results to browse.

Recent advances in sign-recognition technology, e.g., [9, 15, 16, 24, 40, 61, 63, 81, 83], have enabled **search-by-video** dictionaries, which can be simpler to use, since users do not need to select linguistic features of a desired sign. Compared to search-by-feature dictionaries, there have been **relatively fewer search-by-video systems**, e.g., [13, 74]. While some work has required users to wear specialized sensors or gloves [25, 35], most require only a color or depth camera, into which the user performs a sign, and then the system returns a list of videos of similar-looking signs [7, 18, 24, 47, 88]. Recent research with user studies has investigated how users' satisfaction with a search-by-video dictionary for Croatian Sign

Language was affected by variations in the design of the search-results page, navigation through search results, and the ease with which users can narrow down the search results [53]. Other recent work on search-by-video ASL dictionaries has examined how users' satisfaction is affected by aspects of the search results list, e.g., where the desired sign appears on the list or the overall precision of the list items [4, 33, 34]. Methodologically, our research draws upon this final study closely, as those authors granted us access to their initial prototype as a starting point for our own designs and had disseminated their Wizard-of-Oz study protocol [4, 33, 34], which we used in our final study.

The output of sign-recognition technology is not perfectly accurate, especially if a user cannot properly perform the sign or submits a low-quality video. One prior study reported on a system in which the correct sign appeared in top 20 search results up to 67% of the time [7]; however, the set of videos within which users were searching was relatively smaller than the cardinality of an entire ASL dictionary. Because of variation in the video quality and vocabulary size, it is difficult to compare performance across systems in prior work. If the desired sign is even on the results list in a search-by-video dictionary, users typically need to browse through a long list of possible matches, increasing the chance a user will **give up on a search** or incorrectly conclude that an item is the match, before reaching their desired sign. While other researchers investigate technical improvements to the underlying AI technology for sign matching [17, 18, 41, 47, 50], we focus on improving the user's experience through HCI research, to make it easier to **use imperfect AI technology** during this task. Unlike a search-by-video systems, our hybrid-search approach would not require users to perfectly reenact the performance of the unfamiliar sign, nor for the AI to be perfect in its accuracy, as users can filter the search results to more easily find their desired sign. However, no prior research had examined whether users would actually find hybrid-search beneficial during ASL dictionary search, thereby motivating our research on the design and evaluation of this new approach.

2.2 Related Work on Searching Language, Videos, and Human Movement

The lack of a standard writing system with a clear relationship to apparent sub-elements of signs contributes to the challenge in searching for unknown ASL signs in a dictionary. For spoken languages with **deep orthography**, i.e., writing systems with little or no obvious correspondence to the sounds of words, users of cross-lingual dictionary systems face similar challenges. Researchers have investigated various methods of advanced querying, e.g., specifying length of a word, rhyming words, vowel sounds, or descriptions [8, 54, 62, 87]. However, there is little research on user satisfaction with these systems, and we found no prior research on post-query filtering methods.

We therefore broadened our literature review to prior research on **searching video collections**, e.g., [3, 38]. Research has examined how users' satisfaction relates to the design of the "results page" listing videos matching a query [20, 36, 52, 65]. For instance, the design of YouTube thumbnails (photos representing a video before it is played) and the accompanying text affects how many users click on the video [67]. Displaying appropriate descriptions, titles, and tags on the results page has been found to help users find relevant video results [48]. Overall, this prior work reveals the importance of the design of the results page for video search, motivating our investigating this for ASL dictionaries. Prior work has also revealed that enabling users to combine multiple categories of parameters when constructing a query for video search yields better results and higher user satisfaction [5, 78, 82]. When searching for something they vaguely remember, users benefit from describing the content of the desired video rather than text within the video title [86]. While prior work on video search has examined adding filtering and post-query refinement capabilities [14, 31, 42], no prior work had examined filtering of a results list in the context of search-by-video for ASL dictionaries.

Analogous to video collections of ASL, recent work has constructed datasets of **human movement**, such as dancing [44, 89], martial arts [45, 89], everyday indoor and outdoor human actions [60, 84], and sports [59, 89, 90]. Searching for specific human movements within such datasets poses similar challenges to searching sign-language dictionaries. Novice learners can find it difficult to search for a matching result among a set of similar dance moves [6], and research has revealed that creators of digital libraries of dance videos should provide textual meta-data on the dance steps, to assist dance learners in searching for content easily [23]. Some research has investigated which features users are most likely to incorporate into their queries when searching for dance moves [6], and other work has examined how users struggled when the search results page displayed only static photos of a dance move [75], suggesting that displaying a list of moving video clips may facilitate users browsing a list of videos matching a query. In fact, some work has investigated search-by-video systems, in which users submit video-based queries in which they attempt to perform the desired movement [30, 69, 76], with a list of matching videos displayed as search results.[1] Overall, this prior work on searching videos of human movement has provided insights regarding designing an ASL look-up system, and this motivates our investigation among novice ASL learners. However, no prior work has explored how search-by-video can be improved in the context of a language-search task, e.g., sign language, nor has prior work specifically considered how hybrid-search approaches could be useful in this context.

3 RESEARCH QUESTIONS AND OVERVIEW OF STUDIES

In our first study, 32 students learning ASL interacted with an ASL dictionary search prototype and participated in an interview about design factors that may affect their search experience, to investigate:

RQ1: After experiencing an ASL dictionary prototype, what were users' preferences in regard to:

- **(a) how to present text or video content in the search-results list of an ASL dictionary search system?**
- **(b) the set of linguistic or appearance features of ASL signs and understandable terminology for referring to these**

[1] Beyond the realm of video-based search, recent systems have made use of machine learning models to allow a user to search-by-sound-performance, e.g., to search for a song using humming, whistling, or singing [46].

features, within a filtering user-interface in a hybrid-search system?

Our second study was a summative experimental evaluation with 20 students learning ASL, who performed search tasks with either our hybrid-search prototype (which incorporated findings from study 1) or with a search-by-video prototype (identical but without post-query filtering capabilities), to investigate:

RQ2: In a comparison between the experience of users who performed a series of ASL-sign search tasks using a search-by-video or a hybrid-search prototype:

- **(a) Is there a difference between users' satisfaction and task performance between prototypes?**
- **(b) What challenges did users face, and how did they make use of the systems' features?**

4 STUDY 1: OPTIMIZING THE DESIGN OF A HYBRID-SEARCH SYSTEM

4.1 Study Design and Methodology

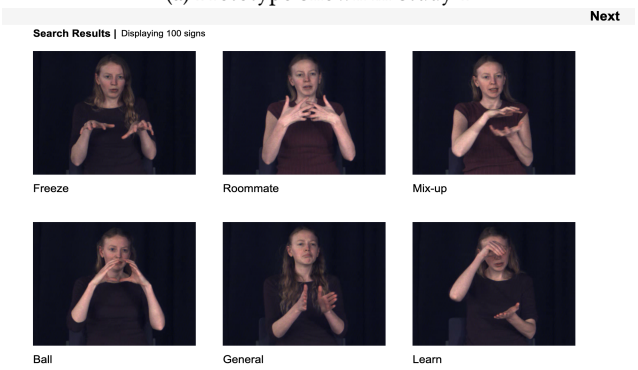
This study was conducted remotely due to the need to maintain social distancing during the COVID-19 pandemic. We started this IRB-approved study by sending an informed consent form to the participant through email, which the participant read and reviewed, prior to a video conference meeting between the researcher and the participant. We first asked participants to imagine a scenario in which they encountered an unfamiliar sign and then to describe how they would look up its meaning. We asked about their prior experiences when encountering difficult to understand ASL signs or using ASL dictionaries. Next, we introduced our ASL dictionary web-prototype (described in the next sub-section) which was deployed online. A link was provided to each participant, and a calibration screen appeared at the beginning of each web-based prototype, to ensure that the size and aspect ratio of the web browser window was consistent across participants. We guided our participants through the process of performing a desired sign into a webcam and viewing a page of results, as shown in Figure 1(a).

After using the prototype, participants then engaged in a semi-structured interview for approximately 35 minutes. Participants were asked about their impression of the overall appearance of the prototype and the text or video information on the page. They were also asked whether video results should play in a looping manner or only on demand, which linguistic properties of ASL to use as filters, and what terminology to use. Participants were also invited to offer any other suggestions on the design.

4.2 Prototype

The authors of a prior study [4, 33, 34] had granted us access to their initial prototype of a search-by-video ASL dictionary, which we used as a starting-point for our own. This prototype consisted of a series of web-pages: The user was first shown a stimulus video of an ASL sign performed by a native signer (not the same person who appeared within the videos in the dictionary); stimuli consisted of vocabulary that new ASL learners would not be familiar with. Additional details about the stimuli and composition of the search results are described in that prior work [4, 33, 34]. After viewing the stimulus, users visited a screen where they could record their own

(a) Prototype Shown in Study 1



(b) Prototype Shown in Study 2

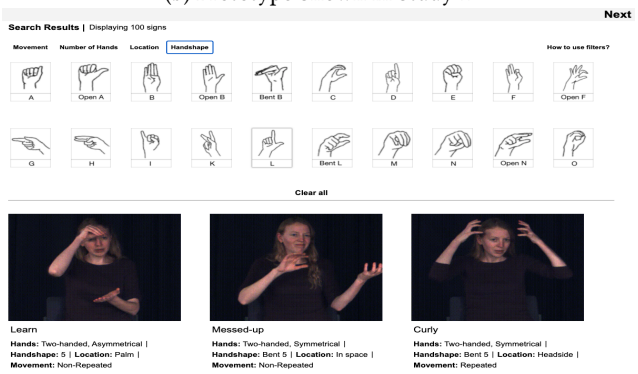


Figure 1: Prototype screenshots: (a) After users submit a video performing a desired word, this search-results page appeared to participants in study 1 (videos played automatically in one version, or only upon hovering in the other). (b) Based on study 1 feedback, a final design for this filtering interface with linguistic terminology, e.g., "Handshape," and graphics illustrating each option.

performance of the sign into a webcam, to initiate a search for the closest matches for their performance within the dictionary. After pressing submit, users saw a results page, consisting of a grid-like listing of short videos of ASL words hand-selected by a native ASL signer, from Boston University's American Sign Language Lexicon Video Dataset (ASLLVD) [58] using the same approach as in [4, 33, 34], each labeled with a one-word English translation. The ASL words chosen in the stimuli consisted mostly of advanced vocabulary that students enrolled in introductory ASL courses would unlikely be familiar with. To avoid the results appearing too perfect, the results list occasionally included some less-similar items, especially near the bottom of the results list.

4.3 Participants and Recruitment

Participants were recruited by contacting professors of introductory ASL courses, who shared an advertisement by email with their students, containing two screening questions: "Are you currently taking an introductory or intermediate course in American Sign Language?" or "Have you completed an introductory or intermediate ASL course in the past five years?" Participants were recruited

if they responded with yes to at least one question. Because we had anticipated some diversity of opinion about aspects of the design of a prototype system, we selected to recruit 32 participants for this study, which included 17 females, 11 males, and 4 non-binary individuals; the mean age was 21 years. A majority of participants were enrolled in introductory ASL courses: fourteen reported studying ASL for one year or less, 15 participants reported between 1 and 3 years, and 3 participants reported studying it more than 3 years.

4.4 Analysis and Findings

We employed both deductive and inductive approaches in our qualitative data analysis. First, based on our research aims and interview questions, we constructed a deductive coding framework with the main categories (e.g., presentation of the video results, textual meta-data, presentation of filtering options). Next, two researchers independently performed an inductive thematic analysis [11] and developed initial codes, reflective of various sub-categories (e.g., for presentation of video results the sub-categories were requiring user to hover over the videos, all videos playing at once, and individual rows animating, etc), through the process of collating and grouping. Codes were reviewed and consolidated in a discussion with all authors, and then agreed-upon sub-category codes were used to extend the original framework. Then two researchers performed the next iteration of the thematic analysis and organized resulting themes into the framework, and they checked the inter-rater reliability of how individual participant responses were labeled with sub-category codes (Cohen's $\kappa = 0.78$). Finally, all authors discussed, synthesized, and extracted the final themes (emboldened design takeaways) presented in the next section.

4.4.1 Findings: Appearance of the Search-Results Page. Most participants were satisfied with the graphical appearance of the prototype, e.g., the number of results per single screen (which was 6) before needing to scroll down the page, the font size of the text, and the overall layout.

ASL video results should play automatically in a simultaneous and looping manner: Seventeen participants preferred videos to play automatically in a looping manner all at once, rather than only when the user clicks or hovers on each item. The main theme that emerged in participants' comments about this issue were in regard to efficiency. For instance, P12 explained that this enabled quicker browsing of the items: *"I prefer auto-play just because I could quickly glance and look for the movement that I was looking for. And then based off the movement, I could look at the handshape. So I think it was a little quicker."*

Seven participants expressed preference for videos being played only upon hovering the mouse over each. P3 explained that viewing all the videos playing at once was "a little overwhelming," and went on to explain how a static view of the initial pose of an ASL sign might facilitate someone quickly browsing the handshapes in the image, which could be used for deciding if a sign is the desired item. Thus, while providing looping-all-at-once as a default setting may be reasonable, users should be given an option to switch to a play-when-hovering mode.

ASL sign result items should include textual meta-data of linguistic properties: Our participants asked for various text information to appear alongside the video items on the search-results page. Fifteen participants requested for text to appear next to each item listing linguistic properties, e.g., P3 mentioned another ASL dictionary website they had used [49], which provided text next to items on the results page indicating, e.g., *"compounds, where you find movement, hand-shape, location."* P1 suggested including text next to each item to describe movement, e.g., *"basic words to describe whether like circular motions or repetitive movements."*

To identify the type of linguistic elements to include in this text as well as terminology to use for describing each, we also asked participants how they usually describe an ASL sign to other people. Twenty-seven participants indicated that they use traditional ASL linguistic features: handshape, location, orientation, movement, and relative position [70].

4.4.2 Findings: Filtering Options and Terminology. We asked our participants whether it would be useful to be able to filter items displayed on the results page, and 23 of our participants expressed strong interest in this feature—mentioning how filters would help reduce the number of search results, help disambiguate similar-looking signs so users could more confidently identify the desired sign, and speed up the process of finding their desired result. For example, P9 said, *"I think the filter would be really nice, especially if you didn't necessarily have to do all of the boxes. So, like, if you had the option of, like, location, handshape or something else, but you only had to fill out one that would make it easier to just trim down some of the ones that you already know are irrelevant."*

Handshape and non-manual-marker filter options should have graphic and text labels: Although some participants admitted to occasionally struggling to recall the exact handshape when thinking of a sign, 26 participants requested we include handshape as a filter and mentioned common handshapes they would like to filter upon. Participants were uncertain how to speak about handshapes in English, unless the handshape corresponded to a specific letter of the ASL fingerspelling alphabet, and two participants suggested including both an image and the conventional English name of each handshape. P32 explained, *"If there was a graphical with it, like a shape like this [performs an open five handshape], or a claw hand [performs a C handshape], or one [performs a numeral 1 handshape], or Y [performs a Y handshape], or whatever. I think that would definitely- being able to sort by handshape would make things a lot easier."* Notably, P16 produced these handshapes while speaking to clarify which ones he was referring to, further illustrating the benefits of dual text and graphic representation. Six participants suggested adding a filter for non-manual markers, which are linguistically important facial expressions, eyebrow movements, and mouth movements in ASL. Similar to handshapes, participants requested inclusion of graphical representations of each, e.g., as P19 explained, *"Facial movement might be a little harder to try and narrow down through check boxes."*

Users should filter on the body-relative location where the hand spends the most time: Seventeen participants suggested location as a filtering criterion and suggested how to specify a sign's location, typically as body-relative locations, e.g., near the head, chest, waist, etc. However, three participants discussed how it can be difficult to ascribe a single location for some signs in which the hands move through space extensively during the sign, and

they offered possible solutions: Two suggested enabling users to filter based on the location where the hand spends the most amount of time, e.g., as P31 explained, *"I think it should be where it spends the most time because if it just passes through an area, if just where it starts and ends, oftentimes, it's just there for a very short period of time... I'd mostly just put it in where it is for the majority of the time."*

Filter signs by 1-handed, 2-handed symmetrical, or 2-handed asymmetrical: Eighteen participants suggested filtering on whether a sign is one-handed or two handed, although others shared some potential concerns: One participant suggested that people often sign two-handed signs using one hand, and if users incorrectly select the one-handed option, they might not find the desired result. P18 suggested that it may be useful to sub-divide two-handed signs even further, i.e., enabling users to specifically filter for two-handed signs in which the movement of the hands is symmetrical or not, as this was a relatively easy distinction to perceive visually.

Filter signs by movement according to whether repeated or non-repeated: Sixteen participants also asked to filter on a sign's movement. Of these, one participant suggested filtering signs based on circular or linear movement, and the remaining 15 participants suggested simply providing two filter options, whether the sign is repeating or non-repeating, e.g., P28 wanted to filter for signs with *"One solid movement? Or is it repeated?"*

Finally, while 23 participants expressed a strong interest in filters, three participants were skeptical. Some worried that they might not be able to narrow down a large set of signs on the basis of filters, and others wondered whether it would require a great deal of time or effort to do so, e.g., with P18 commenting *"for the amount of work that would require, plus like the benefit of that, I don't know if that would rationalize that."* Noting that some participants were uncertain of the benefits of filters further motivated us to investigate whether there would be measurable differences in the usability between an ASL dictionary system with or without filters, in Study 2 below.

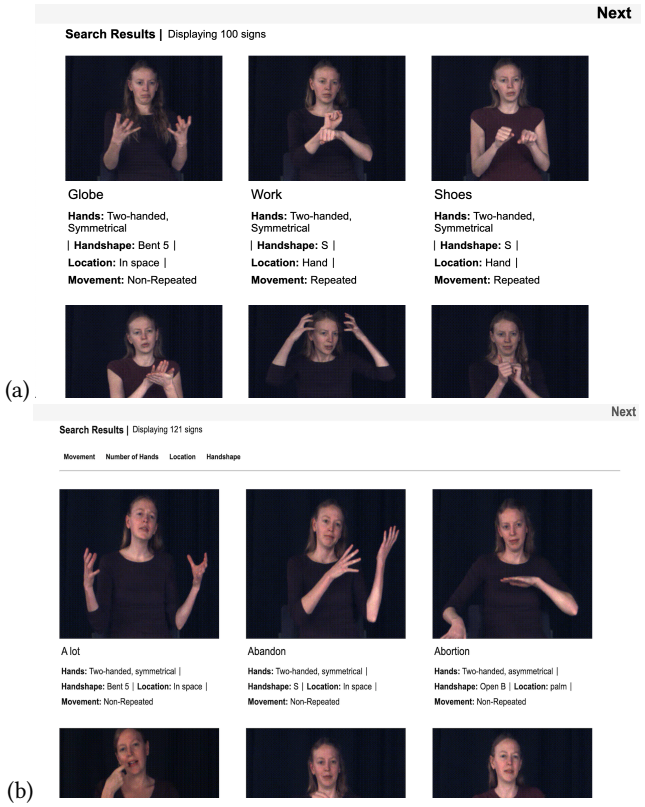


Figure 2: Example screenshots of the search-results pages displayed during Study 2 to: (a) participants in the group who used the search-by-video prototype (i.e. without filtering) and (b) participants in the group who used the hybrid-search prototype.

5 STUDY 2: COMPARISON BETWEEN SEARCH-BY-VIDEO AND HYBRID-SEARCH APPROACHES

Study 1 had provided useful guidance about how elements of a filtering interface could be designed and whether participants were interested in filtering capabilities when asked to imagine their use; however, that study was not able to establish whether including filtering capabilities would actually be beneficial. To address this, it would be necessary for participants to have the experience of actually conducting ASL search tasks with a prototype system; therefore, Study 2 consisted of an experiment to investigate whether there was a measurable advantage to hybrid-search ASL dictionary systems. (This corresponds to research question 2 in this paper.)

To select an **appropriate baseline for comparison**, we considered comparison against search-by-feature or search-by-video systems, ultimately selecting search-by-video as our baseline, for several reasons: Much prior work, e.g., [10, 81], had already evaluated search-by-feature approaches, identifying shortcomings discussed in section 2.1. In contrast, recent research on ASL dictionaries has focused on search-by-video systems, in which the query formation is easier for students, yet accuracy limitations in the technology may require users to browse too long a list of results before finding a desired sign. As our hybrid-search directly builds upon search-by-video with an additional filtering step, it seemed most elucidating to compare our new approach to a search-by-video baseline.

5.1 Study Design and Methodology

This IRB-approved experimental study followed the methodology of prior work [4, 33, 34] but was conducted by videoconference due to COVID-19. In this **between-subjects design**, each participant shared their screen while interacting with one of two versions of our web-based Wizard-of-Oz prototype, where they submitted a video of themselves performing a sought-after sign and viewed search results. A between-subjects design was selected since we were comparing two prototypes in which one had a subset of the features of the other, i.e., with one having filtering capabilities after the search results were displayed. If a within-subjects design had been selected, then each participant would use both systems, and there may be a predisposition to believe that a prototype obviously having an additional capability is superior.

All of the findings from study 1 (as itemized in the boldface text in section 4.4) were incorporated into the design of our prototype,

before it was used in study 2. Two versions of the prototype were compared:

- Our **hybrid-search** prototype included 4 filters: Movement, Number of Hands, Location, and Handshape (with graphics for each handshape). These four filter options are shown in Figure 2(a); if the user clicks on the word "Handshape" on the user-interface, a selection palette displaying handshape names and graphics appears, as shown previously in Figure 1(b). On the interface, we also added a "How to use filters?" link to a page containing a guide for users on how to use the filter interface.
- Our **search-by-video** prototype was identical to the hybrid-search prototype above, except that no filtering interface was provided on the search-results page, as shown in Figure 2(b).

The design of these two prototypes is also explained in the attached video figure file.

Before performing each of 32 search queries, participants viewed a stimulus video of a native ASL signer performing an ASL sign likely to be unfamiliar to novice ASL learners [4, 33, 34]. The person who appeared in this original stimulus video was not the same person who appeared in the dictionary results. Next, participants performed the sign from memory into their webcam, and then the Wizard-of-Oz prototype displayed a search-results page. The desired sign appeared at one of 32 different rank-positions on the page, following a Latin Square per query per participant (details of this schedule are included in Electronic Supplementary Files). The set of rank-positions reflected the likely accuracy of state-of-the-art sign-recognition technology [18, 24, 47, 88] on the quality of videos that users of an ASL dictionary system may submit. Participants were asked to look through the results page and identify (click on) the item that matched their sought-after sign. We recorded how much time participants took to search for the desired item in the results list and whether the sign that they thought best matched their query was correct. After each query, participants responded to subjective questions, which, for brevity, are enumerated in the Findings section below.

5.2 Participants and Recruitment

The recruitment criteria were the same as for the prior two studies. The 20 participants in this between-subjects study were divided into two groups of 10, with each participant doing 32 searches: The **hybrid-search group** included 4 males, 5 females, and 1 non-binary person, with a mean age of 21.5 years. Three participants reported taking their first ASL course less than 12 months ago, 5 between 1 and 3 years ago, and 2 between 3 to 5 years ago. The **search-by-video group** (i.e. without filtering) included 1 male and 9 females, with mean age of 21 years. Two participants reported taking their first ASL course less than 12 months ago, 6 between 1 and 3 years ago, and 2 between 3 to 5 years ago.

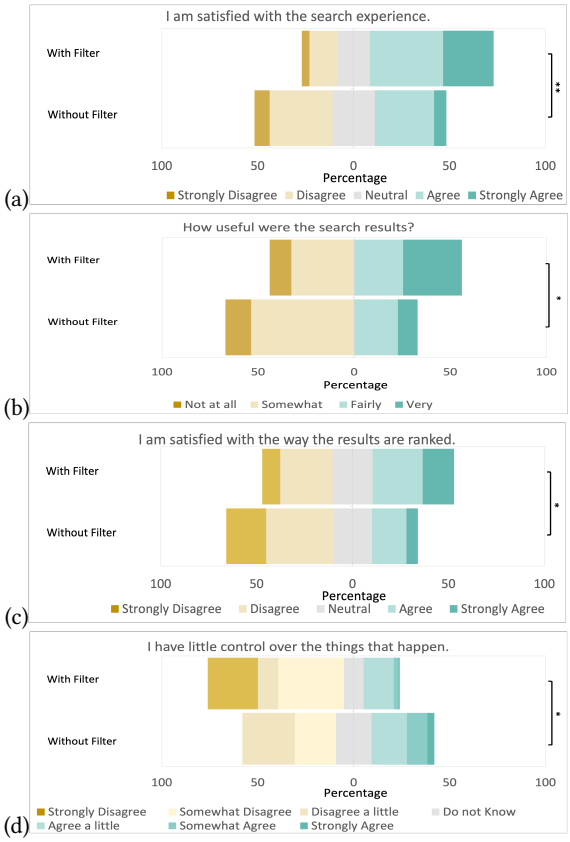


Figure 3: Subjective user responses comparing hybrid-search and search-by-video prototypes in study 2, for: (a) whether "satisfied" with the experience, (b) how "useful" search results were, (c) satisfaction with how "results are ranked," (d) "control" over things (Significant differences are marked with: ** p<0.01, or * p<0.05).

5.3 Findings: Quantitative Analysis of Search Experience

5.3.1 User Satisfaction Metrics. A Mann-Whitney U test was used to determine whether average responses differed between participants who had used the hybrid-search prototype and those who used the search-by-video prototype.

- **Q1: I am satisfied with the search experience.** This 5-point Likert question ("Strongly Disagree" to "Strongly Agree") was adapted from prior work on search systems [39]. As shown in Figure 3(a), hybrid-search participants (Mean=3.685, Median=3.697) reported significantly higher satisfaction than search-by-video participants (Mean=2.952, Median=2.940): $U(N_{Filter}=10, N_{NonFilter}=10) = 11.5$, $z = 2.873$, $P = .0041$ ($< 0.01^{**}$).
- **Q2: How useful were the search results?** This question adapted from from a prior work on search satisfaction [51]

used a 4-point scale from "Not at all" to "Very Useful." Figure 3(b) shows hybrid-search participants (Mean=2.755, Median=2.864) reported significantly higher usefulness of results than search-by-video participants (Mean=2.3, Median = 2.318): U(N_{Filter}=10, $N_{NonFilter}$=10) = 20 , z = 2.230, P = 0.02574 (< 0.05*).

- **Q3: I am satisfied with the way the results are ranked.** This 5 point Likert question ("Strongly Disagree" to "Strongly Agree") was adapted from [2] and was used in prior work on ASL dictionary search [4, 33]. Figure 3(c) displays how hybrid-search participants (Mean=3.127, Median=3.288) reported higher satisfaction with the way the results were ranked than did search-by-video participants (Mean=2.533, Median=2.530): U(N_{Filter}=10, $N_{NonFilter}$=10) = 23, z = 2.419, P = .01552 (< 0.05*).
- **Q4: I have little control over the things that happen.** This 7-point Likert question was adapted from [79], which was a psychological scale that measures an individual's sense of control. As shown in Figure 3(d), there was significant difference in responses between hybrid-search (Mean=2.652, Median = 2.258) and search-by-video participants (Mean=3.803, Median = 4.076): U(N_{Filter}=10, $N_{NonFilter}$=10) = 21, z = -2.154, P = 0.03 (<0.05*). Hybrid-search participants reported that they had greater control.

5.3.2 Time to Complete Search. We recorded the time it took our participants to complete all searches using both the prototypes, measuring from the time when the search-results page loaded until the participant clicked to indicate they had found their desired result. Figure 4 reveals that when the desired item appeared farther down the search results page, more time was required, regardless of whether using a hybrid-search or search-by-video prototype. Across all searches in this study, participants took more time when using the hybrid-search prototype (t-test, p<0.0001). However, Figure 4 reveals that when the desired item was closer to the top of the search-results list (i.e., for all positions below 50), we observed lower median search-times for the hybrid-search prototype.

5.3.3 Task Success. We also recorded whether the sign participants clicked on was a match to the original sign they had been asked to search for. Participants using the hybrid-search prototype found the correct sign in 84% of cases, and participants using the search-by-video prototype found the correct sign in 79% of cases. However, a chi-square test did not reveal any significant difference.

5.4 Findings: Qualitative Analysis of Search Experience

At the end of the study, participants were asked to provide some open-ended comments about their impression of the system, what factors had affected their experience, and how they had used the system. Based on the constructivist grounded theory [19], we performed a thematic analysis on this qualitative data, generated codes, and identified emerging themes by categorizing the codes. Participants discussed how they made use of the functions given, the types of challenges that they faced from the start of search to the end, and suggestions for improvement of the system.

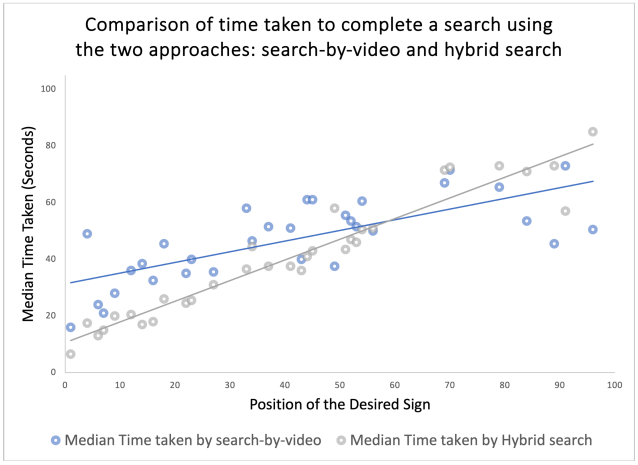


Figure 4: Scatterplot displaying median search times, for each rank-position on the search-results page, for the hybrid-search and search-by-video prototypes, and with best-fit lines added for each group.

5.4.1 Lack of Confidence when Producing Sign Videos. Participants from both the hybrid-search and search-by-video groups discussed their struggles and concerns about trying to remember the sign that they had seen and to re-create its performance to produce a video query. They worried that their insufficient skill or accuracy in producing the sign may negatively impact the quality of the search results. For instance, P20 explained that their main difficulty arose, *"when I'm copying the sign. My way of signing might be different than the way that they produce it. I'm trying to copy them as much as I can. And then go to the video, it might be a little bit different."*

5.4.2 Challenges Faced during Browsing a Long Search-Results List. Most participants who used the search-by-video prototype described their unsatisfactory experiences with perusing the search-results page, including the inefficient and lengthy process of seeking a desired sign, e.g. P2, *"I think the main issue is just, you had to search for a long time to actually find the word."* While browsing the result page, 16 participants expressed how they would have benefited from signs being sorted or grouped to facilitate finding a desired sign, e.g. P18, *"So if similar handshapes were all grouped together in one area I that would have helped me a lot."* Others wished for the ability to see signs similar to a current selection, e.g., P5 said, *"I think it could be helpful if there were an option to view similar signs. So if I find a sign that is close to what I was looking for, I could view similar signs to that."* In fact, 15 search-by-video participants suggested how a filtering feature would have been useful, e.g. P1, *"you could, like, narrow the search results... you know that's not a two handed sign so you can get rid of all of those ones...that's like something I was doing when I was looking through them and be like okay well I know it's not this one because they're only doing like one hand."*

5.4.3 How Participants Made Use of Filtering during Hybrid Search. Most participants with the hybrid-search system did not use the

filters immediately upon viewing the search results but instead used them after an initial scan of the first few rows if no desired sign was found. P4 described the process: *"The filter is... incredibly helpful because sometimes the signs just wouldn't show up right there at the top like the only one that showed up right at the top for me was 'subway,' 'mom'...I would scroll like down a couple unless I didn't know what the sign was, and then I would scroll down a few more. And then I would start going through the filters to try and elucidate out some."* P15 added: *"The search tool was great, and I used it for most of my searches unless what I was looking for was in the first 3 rows."*

During their search, most participants made use of the "one handed or two handed" filter, as well as the "handshape" filter. However, participants described the "symmetrical/asymmetrical" and the "location" filters as being less helpful, due the distinction being unclear between them and the filter for "hand movement." P4 expressed the challenge in using the symmetrical/asymmetrical filter: *"Whether if it was symmetrical or asymmetrical I really had a hard time kind of deciding that one. It, um, unless it was clearly obvious like 'wheelchair' like that's obviously symmetrical..I found myself using the handshape more a lot."* P19 also shared their confusion: *"I wasn't sure if that meant symmetrical and both hands are doing the exact same thing or if that meant like one hand can be doing something like they're both doing the same thing but at different times. They're not like mirror images, they're like offset images."* As for the location feature, P14 explained why this filter was less useful and suggested a possible way to make it better: *"the location, I only really used if I do it in space, because I wasn't quite sure...if it's face or if it's like chin...does that count as his head, or face, or other? So I think if you had like pictures of the handshapes...that would be helpful."*

6 DISCUSSION

6.1 The Perspective of Novice Signers

Our findings from study 1 revealed insights for future researchers and designers of ASL dictionaries, especially in regard to browsing the results list and selecting features on an interface to filter the results, for a hybrid-search system. Participants' preferences may have related their being novice ASL learners. For instance, participants preferred videos on the results page to play automatically, and interview responses suggested this was due, in part, to novice ASL signers being less familiar with many items listed; they struggled to find their desired sign without playing each ASL video. In contrast, experienced ASL users were able to imagine most signs based on their English gloss labels without viewing each video, and such users may find auto-play videos overwhelming, a concern raised by a few participants. The best design trade-off may depend upon the skill-level of the intended users. Our participants also preferred having linguistic properties listed next to each item on the results list, as these suggest features upon which they could filter. Again, experienced ASL learners may already know these features for signs, and we speculate they would benefit less from including this information on the page.

Our findings also revealed specific linguistic properties and terminology for inclusion in a filtering interface. Again, our interviews revealed that our users' novice status influenced which features were familiar and understandable. Our findings suggested the need for filter options to be depicted with both text and graphical representations, e.g., images of handshapes alongside their conventional name. While envisioning a designed sign, ASL learners may have more difficulty recalling the name of a handshape; novices especially may benefit from graphical elements.

6.2 Design Choices Affecting Perceptions of System Performance

Searching a dictionary using video is a collaborative human-AI task, reliant upon both the sign-recognition technology and the user considering the output to navigate to their desired sign. Our findings in section 5.4.1 revealed that users felt that the accuracy of their performance in their video submission contributed to the overall success of the sign-recognition technology. While improvements to such technology, pursued by AI researchers, may further benefit the quality of the results items displayed after a video query, our study has focused on how the design of the search interaction can enable users to make better use of existing, imperfect recognition technology.

Our findings revealed that adding filtering capabilities to a video-based dictionary (to make it a hybrid-search) significantly improved users' satisfaction with the search experience, **even when performance of the underlying sign recognition technology was the same**. In fact, even though the ranking of the results from the video query was the same across prototypes, hybrid-search users were more satisfied with how the results were ranked, suggesting subjective benefits to **users' perception of the system's performance**. Our findings are good news for designers of ASL dictionary systems, who can not only anticipate advances in AI sign-recognition technology performance in coming years, but who can improve user satisfaction through design choices. It should be noted that the performance of the Wizard-of-Oz sign-recognition in our study was kept slightly higher than the current state-of-the-art [7, 18, 24, 47, 88]; so, our findings should remain relevant even as the accuracy of performance of sign-recognition systems improves.

Prior work on human-AI collaborative systems has also revealed the importance of enabling users to successfully do their part of the task, as well as for **instilling users with a sense of control**, and our findings revealed that hybrid-search users were not only more satisfied with the search experience but had higher scores on a metric of their sense of control. Prior research on ASL dictionaries has also discussed how a user's sense of control can help avoid them quitting a search too soon, before finding their desired result [10].

6.3 Implications of Hybrid-Search for Other Linguistic Contexts

Our findings have implications for dictionary systems for **other sign languages**—beyond ASL. Many sign languages used across the world have similarities in their phonetic structure and linguistic properties, e.g., incorporating handshape, hand movements, and non-manual expressions [12]. To use a feature-based dictionary of these other sign languages, a user would also rely on their ability to recall these properties. Therefore, findings from our study are likely generalizable to dictionary search for many other sign languages.

Our findings also motivate hybrid-search for **orthographically deep spoken languages**, i.e., those for which it is harder to guess how to write a word based on its sound. For languages with shallow orthography, it is relatively easier for novice learners to guess how to "sound out" a word to guess its spelling, to query for a partial match in a dictionary. 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 this first step used video-based recognition for ASL, for orthographically deep spoken languages, it could consist of search by voice (through automatic speech recognition of the student's pronunciation) or search by handwriting (through automatic character recognition, e.g., for written Chinese)—followed by a filtering step. Some spoken-language dictionary systems that already use advanced querying options e.g., [8, 54, 62, 87], could use post filtering as well. Our findings may inform this work, e.g., how displaying additional meta-data alongside results items may benefit novice learners in formulating filtering queries.

6.4 Differences from Prior Research on General Video Search

In related work, we mentioned prior research on browsing results in video search engines [3, 20, 36, 38, 52, 65, 67]. While that prior work helped us identify which features to optimize in a search results of an ASL dictionary, there were unique elements of the ASL dictionary context, and the preferences of novice learners of ASL have also motivated **differences from typical designs in general video-search**. Items on the results page of video search engines, e.g., YouTube, do not typically auto-play by default, unlike our recommendation to do so (in section 4.4.1) in dictionaries for novice ASL learners. General video-search users may rely on both video and audio content to determine whether a search was successful, while ASL dictionaries typically contain video without audio content; moreover, the duration of ASL dictionary entry videos is relatively uniform and short. All of these elements support simultaneous and automatic video-looping, rather than playing on demand. We suspect auto-playing also helps users of ASL dictionaries make comparisons across phonemically similar results.

Our participants also requested text content alongside each result item conveying linguistic features. Prior work on video search engines stressed the importance of using appropriate video descriptions and tagging. However, for general video search this is often limited to displaying video names or time duration alongside result items to benefit users [48]. In our case, the uniformity of short videos of a single ASL sign, and their components of their performance, e.g., handshape or location, supports inclusion of substantial additional linguistic meta-data with each item.

6.5 Implications of Hybrid Search for Human-Movement Search

Our research, in the context of ASL dictionaries, on combining a search-by-performance query with a search-by-feature filtering step also motivates research on hybrid search in other scenarios, e.g., when users are seeking a video of human movement they vaguely remember. As discussed above, our findings may be especially applicable in contexts in which the video being sought is short and the database in which users are searching contains relatively uniform items, e.g., dance steps, martial art moves, sports moves etc. In this way, our findings speak to the broader literature on searching using human performance, which currently makes use of search-by-video or search-by-feature approaches, e.g., [6, 23, 30, 69, 76].

6.6 Design Recommendations Based on Our Findings

Prior research on dictionary systems for written languages has examined design features to improve the quality of dictionary search and proposed design frameworks [72, 73]. For ASL dictionaries specifically, prior work has identified the need for developing research-driven design guidelines for sign language dictionary systems [4, 34, 43], a gap which our studies have addressed.

The findings of our studies provide design guidance for sign-language dictionary search systems and reveal some remaining open questions, which may be a basis for future research studies. Our findings relate to several phases of a user's interaction with a video-based dictionary-search system:

(1) When initiating their search, our interviews revealed how users did not feel confident when asked to reproduce a sign from memory into a webcam; future designers and researchers may explore additional interventions to address this concern, such as providing more detailed instructions or tips for users when they begin this initial video query.
(2) For the portion of the interaction when initial query results are displayed to users, our first study motivated two specific design recommendations (section 4.4.1), i.e., for videos to play automatically in a looping manner and for textual metadata with linguistic properties to appear.
(3) During the filtering phase of the interaction, our first study also motivated four design recommendations (section 4.4.2) relating to the specific linguistic filters that should appear and how some, such as handshape, should be accompanied by graphic elements. More fundamentally, our second study provided quantitative evidence that including a filtering capability in the interface led to subjective benefits for users (section 5.3)—e.g., satisfaction, perceived usefulness, sense of control—even when the overall accuracy of the ranking of results returned from the initial query was the same. Observations of our participants revealed that many waited to use the filtering capabilities until after they had visually scanned the first few rows of results; future research may investigate whether providing additional tips or prompts to users about the filtering tool may lead to quicker utilization of the tool and improved usability of the dictionary search system.

7 LIMITATIONS AND FUTURE WORK

There were several limitations in our study which suggest avenues for future research. For instance, our study examined a prototype which participants interacted with using a web-browser on their computer; future research would be needed to investigate the design of prototypes for ASL dictionary search on other devices or form factors, e.g., smartphones.

While our experimental study investigated the specific task of seeking a sign immediately after having viewed a video of someone

performing that single word, this does not capture all **use-cases** of dictionary systems. We did not explore how asking users to recall a sign from farther in the past nor seeking an unfamiliar sign that had appeared within a video of a longer ASL message may affect users' experience of dictionary search. Further, we did not examine how learners might use this tool as part of a larger task, e.g., completing a homework assignment or translating an ASL message, nor how long-term use of this tool might support overall learning of ASL among students. Further studies on this use in context (with measures of overall task success or user satisfaction) or longitudinal research among users of a deployed system (with measures of ASL learning) could shed light on these issues. In fact, most findings in our study have relied on users' subjective judgments. Thus, a future study collecting **objective measures** of success on tasks or analysis of behavioral measures, e.g., eye-tracking, would be valuable for understanding whether hybrid-search approaches for ASL dictionary search affect these other aspects.

Our research has not fully explored **the presentation of textual meta-data**, **the design of video thumbnails**, nor **the entire design space of filtering** in this ASL dictionary context. The filtering method explored in our hybrid-search prototype reduced the number of items on the results list; however, future studies could investigate alternatives, e.g., filters that re-sort results rather then removing them. Future work could seek to address some of the confusion about particular filters which participants reported, as discussed in section 5.4.3. Future work could investigate more deeply the design space of how filtering controls are positioned (e.g., whether at the top or along the left of the page) or formatting of the text containing linguistic properties for each results item.

Some participants had mentioned that they perceived the use of a filter as an extra step that required further effort. While our experimental study revealed that filtering led to improved user satisfaction and sense of control, a future study could explore further participants' view of this trade-off and seek ways to reduce the **participants' perception of effort**, e.g., through alternative filter designs. Further qualitative research could specifically investigate the reasons why users may abandon a search, as this may guide design of systems that support users' success.

Our study focused on understanding the needs of ASL learners who have only been studying the language for a few years, and the participants we recruited tended to be university students enrolled in ASL courses. Future studies could recruit a more **diverse range of potential users**, e.g., individuals who have not participated in formal ASL classes or younger students in secondary school, to determine whether those groups would also benefit from the proposed designs. More broadly, as discussed above, the preferences and requirements of novice ASL learners may differ from those of users with more advanced ASL skill. Furthermore, future research would be essential to understand the use-cases and requirements of DHH users of ASL dictionaries, both among early ASL learners, as well as more advanced signers.

8 CONCLUSION

While the accuracy of sign-recognition technologies is still improving on the videos of diverse quality that ASL learners are likely to submit, the desired word is unlikely to be returned as the first item of search results in search-by-video sign-language dictionary systems. While prior work had explored the relative benefits of search-by-video dictionaries in comparison to earlier search-by-feature approaches, **no prior work had examined the combination of the two**. To address this gap in the literature, we investigated a **hybrid-search** approach, in which users search for a sign using video input and then sort the results using filters. Our study 1 provided several research-driven guidelines for improving the functionality and usability of these systems, with specific needs of novice ASL learners in mind, that future designers of sign-language dictionaries can incorporate in their designs. Further, our study 2 has revealed that users prefer this hybrid approach, which led to higher measures of user satisfaction. Finally, our findings contribute more broadly to research on performance-based search approaches, especially for human-movement datasets.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under award No. 1763569. We are grateful for the contribution of Abhay Rajendra Dixit and Sunnihith Bharadwaj Manduva in the designing of prototypes for this work.

REFERENCES

[1] Alikhan Abutalipov, Aigerim Janaliyeva, Medet Mukushev, Antonio Cerone, and Anara Sandygulova. 2021. Handshape Classification in a Reverse Dictionary of Sign Languages for the Deaf. In *From Data to Models and Back*, Juliana Bowles, Giovanna Broccia, and Mirco Nanni (Eds.). Springer International Publishing, Cham, 217–226.

[2] Azzah Al-Maskari, Mark Sanderson, and Paul Clough. 2007. The Relationship between IR Effectiveness Measures and User Satisfaction. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (Amsterdam, The Netherlands) *(SIGIR '07)*. Association for Computing Machinery, New York, NY, USA, 773–774. https://doi.org/10.1145/1277741.1277902

[3] 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

[4] 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

[5] Robin Aly, Roeland J.F. Ordelman, Maria Eskevich, Gareth J.F. Jones, and Shu Chen. 2013. Linking inside a Video Collection: What and How to Measure?. In *Proceedings of the 22nd International Conference on World Wide Web* (Rio de Janeiro, Brazil) *(WWW '13 Companion)*. Association for Computing Machinery, New York, NY, USA, 457–460. https://doi.org/10.1145/2487788.2487967

[6] Vasiliki Arpatzoglou, Artemis Kardara, Alexandra Diehl, Barbara Flueckiger, Sven Helmer, and Renato Pajarola. 2021. DanceMoves: A Visual Analytics Tool for Dance Movement Analysis. In *EuroVis 2021 - Short Papers*, Marco Agus, Christoph Garth, and Andreas Kerren (Eds.). The Eurographics Association, Norrköping, Sweden, 5 pages. https://doi.org/10.2312/evs.20211057

[7] Vassilis Athitsos, Carol Neidle, Stan Sclaroff, Joan Nash, Alexandra Stefan, Ashwin Thangali, Haijing Wang, and Quan Yuan. 2010. Large lexicon project: American sign language video corpus and sign language indexing/retrieval algorithms. In *Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies (CSLT)*, Vol. 2. European Language Resources Association (ELRA), Valletta, Malta, 11–14.

[8] Doug Beeferman. 2021. OneLook Dictionary Search. https://www.onelook.com/

[9] Danielle Bragg, Oscar Koller, Mary Bellard, Larwan Berke, Patrick Boudreault, Annelies Braffort, Naomi Caselli, Matt Huenerfauth, Hernisa Kacorri, Tessa Verhoef, Christian Vogler, and Meredith Ringel Morris. 2019. Sign Language Recognition, Generation, and Translation: An Interdisciplinary Perspective. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility* (Pittsburgh, PA, USA) *(ASSETS '19)*. Association for Computing Machinery, New York, NY, USA, 16–31. https://doi.org/10.1145/3308561.3353774

[10] 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

[11] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[12] Fabian Bross. 2015. Chereme. In: Hall, T. A. Pompino-Marschall, B. (ed.): Dictionaries of Linguistics and Communication Science (Wörterbücher zur Sprach- und Kommunikationswissenschaft, WSK). Volume: Phonetics and Phonology. Berlin, New York: Mouton de Gruyter. *Proceedings of the XVIII EURALEX International Congress* 1, 1 (01 2015), 9.

[13] 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

[14] Carrie J. Cai, Emily Reif, Narayan Hegde, Jason Hipp, Been Kim, Daniel Smilkov, Martin Wattenberg, Fernanda Viegas, Greg S. Corrado, Martin C. Stumpe, and Michael Terry. 2019. Human-Centered Tools for Coping with Imperfect Algorithms During Medical Decision-Making. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3290605.3300234

[15] Necati Cihan Camgöz, Simon Hadfield, Oscar Koller, Hermann Ney, and Richard Bowden. 2018. Neural Sign Language Translation. In *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018*. IEEE Computer Society, New York, New York, US, 7784–7793. https://doi.org/10.1109/CVPR.2018.00812

[16] Necati Cihan Camgöz, Oscar Koller, Simon Hadfield, and Richard Bowden. 2020. Sign Language Transformers: Joint End-to-End Sign Language Recognition and Translation. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020*. IEEE, New York, New York, US, 10020–10030. https://doi.org/10.1109/CVPR42600.2020.01004

[17] Alvaro Leandro Cavalcante Carneiro, Lucas de Brito Silva, and Denis Henrique Pinheiro Salvadeo. 2021. Efficient sign language recognition system and dataset creation method based on deep learning and image processing. arXiv:2103.12233 [cs.CV]

[18] Christopher Conly, Zhong Zhang, and Vassilis Athitsos. 2015. An Integrated RGB-D System for Looking up the Meaning of Signs. In *Proceedings of the 8th ACM International Conference on PErvasive Technologies Related to Assistive Environments* (Corfu, Greece) *(PETRA '15)*. Association for Computing Machinery, New York, NY, USA, Article 24, 8 pages. https://doi.org/10.1145/2769493.2769534

[19] Juliet Corbin and Anselm Strauss. 2021. Basics of Qualitative Research (3rd ed.): Techniques and Procedures for Developing Grounded Theory - SAGE Research Methods. https://methods.sagepub.com/book/basics-of-qualitative-research [Online; accessed 19. Mar. 2021].

[20] Paolo Cremonesi, Mehdi Elahi, and Franca Garzotto. 2017. User interface patterns in recommendation-empowered content intensive multimedia applications. *Multimedia Tools and Applications* 76, 4 (2017), 5275–5309.

[21] Tirthankar Dasgupta, Sambit Shukla, Sandeep Kumar, Synny Diwakar, and Anupam Basu. 2008. A Multilingual Multimedia Indian Sign Language Dictionary Tool. In *Proceedings of the 6th Workshop on Asian Language Resources*. Asian Federation of Natural Language Processing, Hyderabad, India, 53–57. https://www.aclweb.org/anthology/I08-7008

[22] Eberhard, David M., Gary F. Simons, and Charles D. Fennig (eds.). 2021. Sign language. https://www.ethnologue.com/subgroups/sign-language

[23] Katerina El Raheb and Yannis Ioannidis. 2013. Dance in the World of Data and Objects. In *Information Technologies for Performing Arts, Media Access, and Entertainment*, Paolo Nesi and Raffaella Santucci (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 192–204.

[24] Ralph Elliott, Helen Cooper, John Glauert, Richard Bowden, and François Lefebvre-Albaret. 2011. Search-By-Example in Multilingual Sign Language Databases. In *Proceedings of the Second International Workshop on Sign Language Translation and Avatar Technology (SLTAT)*. SLTAT, Dundee, Scotland, 8 pages. http://personal.ee.surrey.ac.uk/Personal/H.Cooper/research/papers/SBE_SLTAT.pdf

[25] Kirsten Ellis and Jan Carlo Barca. 2012. Exploring Sensor Gloves for Teaching Children Sign Language. *Adv. in Hum.-Comp. Int.* 2012, Article 12 (Jan. 2012), 1 pages. https://doi.org/10.1155/2012/210507

[26] National Center for Education Statistics (NCES). 2018. Digest of Education Statistics Number and percentage distribution of course enrollments in languages other than English at degree-granting postsecondary institutions, by language and enrollment level: Selected years, 2002 through 2016. https://nces.ed.gov/programs/digest/d18/tables/dt18_311.80.asp

[27] José L. Fuertes, Ángel L. González, Gonzalo Mariscal, and Carlos Ruiz. 2006. Bilingual Sign Language Dictionary. In *Computers Helping People with Special Needs*, Klaus Miesenberger, Joachim Klaus, Wolfgang L. Zagler, and Arthur I. Karshmer (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 599–606.

[28] David Goldberg, Dennis Looney, and Natalia Lusin. 2015. Enrollments in Languages Other than English in United States Institutions of Higher Education, Fall 2013.

[29] Google. 2021. Google Scholar. https://scholar.google.com/intl/en/scholar/help.html

[30] Hong Guo, Zou Shanchen, Chuying Lai, and Hongxin Zhang. 2020. PhyCoVIS: A visual analytic tool of physical coordination for cheer and dance training. *Computer Animation and Virtual Worlds* 32 (11 2020). https://doi.org/10.1002/cav.1975

[31] Michael Gusenbauer and Neal R Haddaway. 2020. Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources. *Research synthesis methods* 11, 2 (2020), 181–217.

[32] Wyatte C Hall, Leonard L Levin, and Melissa L Anderson. 2017. Language deprivation syndrome: A possible neurodevelopmental disorder with sociocultural origins. *Social psychiatry and psychiatric epidemiology* 52, 6 (2017), 761–776.

[33] 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*, Vol. 4. Computer Vision – ECCV 2020 Workshops, Virtual, 6 pages.

[34] 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

[35] Jose L. Hernandez-Rebollar and Erick Mendez. 2004. Interactive American Sign Language Dictionary. In *ACM SIGGRAPH 2004 Sketches* (Los Angeles, California) *(SIGGRAPH '04)*. Association for Computing Machinery, New York, NY, USA, 26. https://doi.org/10.1145/1186223.1186256

[36] Alex Hindle, Jie Shao, Dan Lin, Jiaheng Lu, and Rui Zhang. 2011. Clustering web video search results based on integration of multiple features. *World Wide Web* 14, 1 (2011), 53–73.

[37] Robert J Hoffmeister. 2000. *A piece of the puzzle: ASL and reading comprehension in deaf children*. Mahwah, N.J. : Lawrence Erlbaum Associates, New Jersey, USA. 143–163 pages.

[38] 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.

[39] Scott B. Huffman and Michael Hochster. 2007. How Well Does Result Relevance Predict Session Satisfaction?. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (Amsterdam, The Netherlands) *(SIGIR '07)*. Association for Computing Machinery, New York, NY, USA, 567–574. https://doi.org/10.1145/1277741.1277839

[40] Kabil Jaballah and Mohamed Jemni. 2010. Toward Automatic Sign Language Recognition from Web3D Based Scenes. In *Proceedings of the 12th International Conference on Computers Helping People with Special Needs* (Vienna, Austria) *(ICCHP'10)*. Springer-Verlag, Berlin, Heidelberg, 205–212.

[41] Vanita Jain, Achin Jain, Abhinav Chauhan, Srinivasu Soma Kotla, and Ashish Gautam. 2021. American Sign Language recognition using Support Vector Machine and Convolutional Neural Network. *International Journal of Information Technology* 13 (2021), 1–8. https://doi.org/10.1007/s41870-021-00617-x

[42] Michelle Jenkins. 2004. Evaluation of methodological search filters—a review. *Health Information & Libraries Journal* 21, 3 (2004), 148–163.

[43] Michael D. Jones, Harley Hamilton, and James Petmecky. 2015. Mobile Phone Access to a Sign Language Dictionary. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers and Accessibility* (Lisbon, Portugal) *(ASSETS '15)*. Association for Computing Machinery, New York, NY, USA, 331–332. https://doi.org/10.1145/2700648.2811364

[44] Will Kay, João Carreira, Karen Simonyan, Brian Zhang, Chloe Hillier, Sudheendra Vijayanarasimhan, Fabio Viola, Tim Green, Trevor Back, Paul Natsev, Mustafa Suleyman, and Andrew Zisserman. 2017. The Kinetics Human Action Video Dataset. *CoRR* abs/1705.06950 (2017), 5 pages. arXiv:1705.06950 http://arxiv.org/abs/1705.06950

[45] Ksenia Kolykhalova, Antonio Camurri, Gualtiero Völpe, Marcello Sanguineti, Enrico Puppo, and Radosław Niewiadomski. 2015. A multimodal dataset for the analysis of movement qualities in karate martial art. In *2015 7th International Conference on Intelligent Technologies for Interactive Entertainment (INTETAIN)*. IEEE, IEEE, Piscataway, NJ, 74–78.

[46] Krishna Kumar. 2020. Song stuck in your head? Just hum to search. https://blog.google/products/search/hum-to-search/

[47] Pradeep Kumar, Rajkumar Saini, Partha Pratim Roy, and Debi Prosad Dogra. 2018. A position and rotation invariant framework for sign language recognition (SLR) using Kinect. *Multimedia Tools and Applications* 77, 7 (2018), 8823–8846.

[48] Patricia G Lange. 2007. Publicly private and privately public: Social networking on YouTube. *Journal of computer-mediated communication* 13, 1 (2007), 361–380.

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

[50] Carman KM Lee, Kam KH Ng, Chun-Hsien Chen, Henry CW Lau, SY Chung, and Tiffany Tsoi. 2021. American sign language recognition and training method with recurrent neural network. *Expert Systems with Applications* 167 (2021), 114403.

[51] Mengyang Liu, Yiqun Liu, Jiaxin Mao, Cheng Luo, Min Zhang, and Shaoping Ma. 2018. "Satisfaction with Failure" or "Unsatisfied Success": Investigating the Relationship between Search Success and User Satisfaction. In *Proceedings of the 2018 World Wide Web Conference* (Lyon, France) *(WWW '18)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 1533–1542. https://doi.org/10.1145/3178876.3186065

[52] Kai Lukoff, Ulrik Lyngs, Himanshu Zade, J Vera Liao, James Choi, Kaiyue Fan, Sean A Munson, and Alexis Hiniker. 2021. How the Design of YouTube Influences User Sense of Agency. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (to appear)*. Association for Computing Machinery, New York, NY, USA, 17 pages.

[53] Klara Majetić and Petra Bago. 2017. Proposing an Instrument for Evaluation of Online Dictionaries of Sign Languages. *INTEGRATING ICTIN SOCIETY* 1, 1 (2017), 189.

[54] MDBG. 2021. English to Chinese dictionary. http://www.mdbg.net/

[55] Daniel Mitchell. 2021. British Sign Language BSL Dictionary. https://www.signbsl.com/

[56] Ross Mitchell, Travas Young, Bellamie Bachleda, and Michael Karchmer. 2006. How Many People Use ASL in the United States? Why Estimates Need Updating. *Sign Language Studies* 6 (03 2006). https://doi.org/10.1353/sls.2006.0019

[57] J Murray. 2020. World Federation of the deaf. http://wfdeaf.org/our-work/

[58] Carol Neidle and Christian Vogler. 2012. A new web interface to facilitate access to corpora: Development of the ASLLRP data access interface (DAI). In *Proc. 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon, LREC*. Citeseer, OpenBU, Istanbul, Turkey, 8 pages. https://open.bu.edu/handle/2144/31886

[59] Aiden Nibali, Joshua Millward, Zhen He, and Stuart Morgan. 2021. ASPset: An outdoor sports pose video dataset with 3D keypoint annotations. *Image and Vision Computing* 111 (2021), 104196.

[60] Asanka G Perera, Yee Wei Law, Titilayo T Ogunwa, and Javaan Chahl. 2020. A multiviewpoint outdoor dataset for human action recognition. *IEEE Transactions on Human-Machine Systems* 50, 5 (2020), 405–413.

[61] Junfu Pu, Wengang Zhou, and Houqiang Li. 2018. Dilated Convolutional Network with Iterative Optimization for Continuous Sign Language Recognition. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*. International Joint Conferences on Artificial Intelligence Organization, Stockholm, Sweden, 885–891. https://doi.org/10.24963/ijcai.2018/123

[62] Fanchao Qi, Lei Zhang, Yanhui Yang, Zhiyuan Liu, and Maosong Sun. 2020. WantWords: An Open-source Online Reverse Dictionary System, In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations* 57, 175–181. https://doi.org/10.18653/v1/2020.emnlp-demos.23

[63] Razieh Rastgoo, Kourosh Kiani, and Sergio Escalera. 2021. Sign Language Recognition: A Deep Survey. *Expert Systems with Applications* 164 (2021), 113794. https://doi.org/10.1016/j.eswa.2020.113794

[64] Kishore K Reddy and Mubarak Shah. 2013. Recognizing 50 human action categories of web videos. *Machine vision and applications* 24, 5 (2013), 971–981.

[65] Tânia Rocha, J Martins, F Branco, and R Gonçalves. 2017. Evaluating youtube platform usability by people with intellectual disabilities (a user experience case study performed in a six-month period). *Journal of Information Systems Engineering & Management* 2, 1 (2017), 5.

[66] Jerry Schnepp, Rosalee Wolfe, Gilbert Brionez, Souad Baowidan, Ronan Johnson, and John McDonald. 2020. Human-Centered Design for a Sign Language Learning Application. In *Proceedings of the 13th ACM International Conference on PErvasive Technologies Related to Assistive Environments* (Corfu, Greece) *(PETRA '20)*. Association for Computing Machinery, New York, NY, USA, Article 60, 5 pages. https://doi.org/10.1145/3389189.3398007

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

[68] ShuR. 2021. SLintoDictionary. http://slinto.com/us

[69] Sotirios Stasinopoulos and Petros Maragos. 2012. Human action recognition using Histographic methods and hidden Markov models for visual martial arts applications. In *2012 19th IEEE International Conference on Image Processing*. IEEE, Piscataway, NJ, 745–748. https://doi.org/10.1109/ICIP.2012.6466967

[70] William C Stokoe. 1980. Sign language structure. *Annual Review of Anthropology* 9 (1980), 365–390.

[71] P Kumar Sugandhi and Sanmeet Kaur. 2018. Online multilingual dictionary using Hamburg notation for avatar-based Indian sign language generation system. *International Journal of Cognitive and Language Sciences* 12, 8 (2018), 120–127.

[72] Piet Swanepoel. 2001. Dictionary quality and dictionary design: a methodology for improving the functional quality of dictionaries. *Lexikos* 11 (2001), 160–190.

[73] Piet Swanepoel. 2008. Towards a framework for the description and evaluation of dictionary evaluation criteria. *Lexikos* 18 (2008), 207–231.

[74] 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.

[75] Damien Tardieu, Xavier Siebert, Barbara Mazzarino, Ricardo Chessini, Julien Dubois, Stéphane Dupont, Giovanna Varni, and Alexandra Visentin. 2010. Browsing a dance video collection: dance analysis and interface design. *Journal on Multimodal User Interfaces* 4, 1 (2010), 37–46.

[76] Shuhei Tsuchida, Satoru Fukayama, and Masataka Goto. 2019. Query-by-Dancing: A Dance Music Retrieval System Based on Body-Motion Similarity. In *MultiMedia Modeling*, Ioannis Kompatsiaris, Benoit Huet, Vasileios Mezaris, Cathal Gurrin, Wen-Huang Cheng, and Stefanos Vrochidis (Eds.). Springer International Publishing, Cham, 251–263.

[77] 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.

[78] Zhe Wang. 2012. *Similarity Search with Multimodal Data*. Ph.D. Dissertation. Princeton University. https://ezproxy.rit.edu/login?url=https://www.proquest.com/dissertations-theses/similarity-search-with-multimodal-data/docview/963697731/se-2?accountid=108 Copyright - Database copyright ProQuest LLC; ProQuest does not claim copyright in the individual underlying works; Last updated - 2021-05-26.

[79] Michael M Ward. 2013. Sense of control and self-reported health in a population-based sample of older Americans: assessment of potential confounding by affect, personality, and social support. *International journal of behavioral medicine* 20, 1 (2013), 140–147.

[80] Kimberly A. Weaver and Thad Starner. 2011. We Need to Communicate! Helping Hearing Parents of Deaf Children Learn American Sign Language. In *The Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility* (Dundee, Scotland, UK) *(ASSETS '11)*. Association for Computing Machinery, New York, NY, USA, 91–98. https://doi.org/10.1145/2049536.2049554

[81] S. Wilcox, J. Scheibman, D. Wood, D. Cokely, and W. C. Stokoe. 1994. Multimedia Dictionary of American Sign Language. In *Proceedings of the First Annual ACM Conference on Assistive Technologies* (Marina Del Rey, California, USA) *(Assets '94)*. Association for Computing Machinery, New York, NY, USA, 9–16. https://doi.org/10.1145/191028.191031

[82] Shicheng Xu, Huan Li, Xiaojun Chang, Shoou-I Yu, Xingzhong Du, Xuanchong Li, Lu Jiang, Zexi Mao, Zhenzhong Lan, Susanne Burger, and Alexander Hauptmann. 2015. Incremental Multimodal Query Construction for Video Search. In *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval* (Shanghai, China) *(ICMR '15)*. Association for Computing Machinery, New York, NY, USA, 675–678. https://doi.org/10.1145/2671188.2749413

[83] Polina Yanovich, Carol Neidle, and Dimitris Metaxas. 2016. Detection of Major ASL Sign Types in Continuous Signing For ASL Recognition. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*. European Language Resources Association (ELRA), Portorož, Slovenia, 3067–3073. https://www.aclweb.org/anthology/L16-1490

[84] Yuya Yoshikawa, Jiaqing Lin, and Akikazu Takeuchi. 2018. STAIR Actions: A Video Dataset of Everyday Home Actions. *CoRR* abs/1804.04326 (2018), 17 pages. arXiv:1804.04326 http://arxiv.org/abs/1804.04326

[85] YouTube. 2021. Advanced search (Filters) - Computer - YouTube Help. https://support.google.com/youtube/answer/111997

[86] Jin Yuan, Yi-Liang Zhao, Huanbo Luan, Meng Wang, and Tat-Seng Chua. 2014. Memory Recall Based Video Search: Finding Videos You Have Seen before Based on Your Memory. *ACM Trans. Multimedia Comput. Commun. Appl.* 10, 2, Article 21 (Feb. 2014), 21 pages. https://doi.org/10.1145/2534409

[87] NA Yurko, IM Styfanyshyn, UM Protsenko, and OV Romanchuk. 2020. English dictionary apps: the features and benefits.

[88] Zahoor Zafrulla, Helene Brashear, Thad Starner, Harley Hamilton, and Peter Presti. 2011. American Sign Language Recognition with the Kinect. In *Proceedings of the 13th International Conference on Multimodal Interfaces* (Alicante, Spain) *(ICMI '11)*. Association for Computing Machinery, New York, NY, USA, 279–286. https://doi.org/10.1145/2070481.2070532

[89] Weichen Zhang, Zhiguang Liu, Liuyang Zhou, Howard Leung, and Antoni B Chan. 2017. Martial arts, dancing and sports dataset: A challenging stereo and multi-view dataset for 3d human pose estimation. *Image and Vision Computing* 61 (2017), 22–39.

[90] Weichen Zhang, Zhiguang Liu, Liuyang Zhou, Howard Leung, and Antoni B. Chan. 2017. Martial Arts, Dancing and Sports dataset: A challenging stereo and multi-view dataset for 3D human pose estimation. *Image and Vision Computing* 61 (2017), 22–39. https://doi.org/10.1016/j.imavis.2017.02.002