Support in the Moment: Benefits and use of video-span selection and search for sign-language video comprehension among ASL learners

Saad Hassan[*]
Akhter Al Amin[*]
Caluã de Lacerda Pataca[*]
Computing and Information Science
Rochester Institute of Technology
Rochester, NY, USA
{sh2513,aa7510,cd4610}@rit.edu

Diego Navarro
National Institute for the Deaf
Rochester Institute of Technology
Rochester, NY, USA
don6763@rit.edu

Alexis Gordon
Sooyeon Lee
Matt Huenerfauth
School of Information
Rochester Institute of Technology
Rochester, NY, USA
{aag7593,slics,matt.huenerfauth}@rit.edu

ABSTRACT

As they develop comprehension skills, American Sign Language (ASL) learners often view challenging ASL videos, which may contain unfamiliar signs. Current dictionary tools require students to isolate a single sign they do not understand and input a search query, by selecting linguistic properties or by performing the sign into a webcam. Students may struggle with extracting and re-creating an unfamiliar sign, and they must leave the video-watching task to use an external dictionary tool. We investigate a technology that enables users, in the moment, i.e., while they are viewing a video, to select a span of one or more signs that they do not understand, to view dictionary results. We interviewed 14 American Sign Language (ASL) learners about their challenges in understanding ASL video and workarounds for unfamiliar vocabulary. We then conducted a comparative study and an in-depth analysis with 15 ASL learners to investigate the benefits of using video sub-spans for searching, and their interactions with a Wizard-of-Oz prototype during a video-comprehension task. Our findings revealed benefits of our tool in terms of quality of video translation produced and perceived workload to produce translations. Our in-depth analysis also revealed benefits of an integrated search tool and use of span-selection to constrain video play. These findings inform future designers of such systems, computer vision researchers working on the underlying sign matching technologies, and sign language educators.

CCS CONCEPTS

• **Human-centered computing** → *Accessibility systems and tools*; *Graphical user interfaces*; *Empirical studies in interaction design*; *User interface programming*.

KEYWORDS

American Sign language, Sign Languages, Continuous Signing, Sign Language Videos, Sign Look-up, Video Selection, Search Interface, Integrated Search, Sign Language Learning, ASL Learning

ACM Reference Format:
Saad Hassan, Akhter Al Amin, Caluã de Lacerda Pataca, Diego Navarro, Alexis Gordon, Sooyeon Lee, and Matt Huenerfauth. 2022. Support in the Moment: Benefits and use of video-span selection and search for sign-language video comprehension among ASL learners. In *The 24th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '22), October 23–26, 2022, Athens, Greece.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3517428.3544883

1 INTRODUCTION

Over 70 million Deaf and Hard of Hearing (DHH) people worldwide use one of the over 300 sign languages recognized by the World Federation of the Deaf [12, 48]. In the U.S., increasing numbers of DHH and hearing individuals are motivated to learn American Sign Language (ASL), which is used by about 500,000 people as a primary form of communication [46]. There are nearly 200,000 students in ASL classes [15] at schools or universities, and ASL has one of the fastest growing enrollments among language classes [17]. Learning ASL can promote interactions between DHH and hearing individuals, to support greater inclusion, mutual understanding, and participation across society. Further, if DHH children cannot access spoken language nor learn sign language during critical developmental years, they may experience language deprivation [20]. As most DHH children are born to hearing parents, their family or teachers are motivated to learn sign languages [53, 61].

Students trying to understand a challenging video is part of sign language education, to develop comprehension skills [18, 29, 38]. While there have been advances in machine translation to automatically convert an ASL video into an English text, such technology is still under development [51], and its use would bypass the educational activity of students working to understand a video themselves. Technology is needed to support learners during a video-comprehension task without fully automating the process.

In foreign-language-learning contexts, dictionaries are a valuable tool for students when faced with an unknown word in a text or audio recording. However, students learning sign languages face challenges when encountering a sign whose meaning they do not know, given the lack of a standard writing system or an easy way

[*] These authors contributed equally to this research.

ASSETS '22, October 23–26, 2022, Athens, Greece
© 2022 Association for Computing Machinery.
ACM ISBN 978-1-4503-9258-7/22/10…$15.00
https://doi.org/10.1145/3517428.3544883

for students to search for the meaning of a sign based on its visual appearance. Existing search tools are not sufficient [3, 24, 25], and it is difficult for students to use websites that ask them to enter linguistic features of the sign and browse a list of results to find the sign that matches what they see [1, 6, 7, 40, 45, 56, 59]. Based on automatic ASL-recognition technology for video matching, recent tools have been researched to enable students to submit a video of a single sign to conduct a search within an ASL dictionary [3, 5, 11, 13, 26, 37, 59, 63]. However, students trying to understand an ASL video may not be able to accurately extract just one sign nor replicate the sign themselves into a webcam to initiate a search [3, 26]. For this reason, we investigate technologies for enabling users to quickly select a span of a video of ASL signing that contains one or more signs that they do not understand, and then trigger a video-analysis search that will return a set of dictionary results with likely matches to the signs contained within the span.

Rather than considering separately the tasks of watching a video and looking up the meaning of an unknown sign, we instead focus on the overall task. Based on feedback from ASL students during an interview study, we investigate an integrated video-player and sign-search tool. In our second study, some students watched videos containing some signs that they were not familiar with, while using a Wizard-of-Oz prototype for playing videos, selecting spans of video, and performing searches for signs. Other students used a baseline prototype without the searching functionality. This study shed light on how ASL learners interact with such a system. This study also investigated the benefit of providing integrated dictionary search to ASL learners (hearing university students) for video-comprehension, in comparison to their use of an existing dictionary website. Our contributions include:

- We present an interview-based study with ASL learners about their experiences watching challenging ASL videos. Our novel findings reveal their desire to view videos of various genres from multiple platforms, factors that lead to challenges in video comprehension, and their current workarounds when facing unfamiliar signs.
- In the task context of ASL learners translating difficult ASL videos, we present the first comparative study between our video-player prototype with integrated dictionary-search, in comparison to an existing search-by-feature dictionary website. Our findings reveal benefits in terms of quality of translations produced and reduced workload.
- We present the first observational study of ASL learners engaged in the task of translating difficult ASL videos while using search technology, specifically a Wizard-of-Oz prototype of our proposed system. Our analysis revealed how users selected sub-spans and conducted searches, and we characterize how users benefited from an integrated tool that presented search results alongside the video (enabling checking of results in context). We found that usage varied depending on the genre of signing video, and we observed unexpected use of the sub-span selection tool for the purpose of constraining the video play-head.

2 BACKGROUND AND RELATED WORK

Background on sign-language linguistics is provided below to explain key terminology, and prior work on sign-language pedagogy is discussed, to contextualize our work within that domain. Next, section 2.2 discusses the current state of sign-language look-up technologies, to illuminate key limitations of existing resources.

2.1 Sign Language Comprehension

ASL linguistic phenomena contribute to challenges students may face in comprehending an ASL video. Although students can browse dictionaries that show videos of an ASL sign's **citation form**, i.e., the standard way in which a sign may appear when produced in isolation, when signs are produced during sentences in a continuous manner, the appearance may differ. For instance, there is natural diversity in the production of signs across individual signers, which may be based on demographic or geographic regional variation. Further, two or more ASL signs may linguistically combine into a **compound sign** [41]; novice ASL learners may have difficulty segmenting them appropriately to look up a meaning in an ASL dictionary [8]. **Coarticulation**, broadly, refers to how the production of one sign may affect the way in which other nearby signs are produced in continuous signing [19, 54], e.g., the ending location or handshape of one sign may affect the location or handshape of the next. Coarticulation effects may lead to the production of a sign in context to differ from its citation form. When ASL signers produce rapid sequences of handshapes during **fingerspelling**, i.e., when specific words are spelled alphabetically, coarticulation effects are also possible [35], leading to the fingerspelled word not being a simple concatenation of the individual alphabet handshapes. Finally, ASL signing may include **depiction**, in which particular linguistic constructions, often referred to as "classifiers," convey spatial information about the position, movement, or shape of entities [58].

Most prior work on sign-language video comprehension has focused on Deaf users. Little prior work–and no prior observational studies–have investigated the behavior of hearing people when watching a challenging ASL video nor their workarounds for unknown signs, e.g., using sign look-up tools. A recent review of prior eye-tracking studies with DHH participants [2] discussed a study that examined differences in gaze patterns between Deaf and hearing individuals when looking at a live signer [55]. Although understanding sign language in person is different then watching a video (and no sign look-up technologies had been used), that study characterized gaze patterns of hearing people when trying to comprehend sign language. Observational studies, with eye-tracking or other means, may lead to insights about behaviors of ASL learners during video comprehension, especially given limited prior work. In contrast, substantial prior literature exists on non-native learners of various spoken languages engaging in video comprehension, and there has been work on spoken/written language translation tasks while learners use various electronic resources [4, 16, 27, 44, 57, 64].

2.2 State of Sign Language Lookup Resources

Dictionaries are an important tool used by second language learners when looking up an unfamiliar word. However, when someone encounters a sign whose meaning they do not know when viewing sign language, it is more difficult to look up the word, since sign languages lack a common writing system and users cannot use a text-search or alphabetical listing to search for a sign [5, 28].

Some sign-language dictionary systems expect users to recall linguistic properties of the sign that they are looking for–e.g., hand configuration, orientation, location, movement–and enter these properties into a search-query interface to obtain a list of matching signs [1, 6, 7, 40, 45, 56, 59]. Prior work on such search-by-feature systems has revealed that they are challenging for ASL students [6]. Other proposed sign-language dictionary systems expect users to submit a video of a single sign that they have extracted from a longer video – or to recall and perform a sign into a webcam [5, 7, 11, 13, 37, 59, 63]; sign-recognition technology performs a video search against a dictionary to provide potential matches. Even if a student were able to remember and produce a sign into a webcam, there are technical challenges in recognizing signs from video due to various factors [52, 62]. Despite recent advancements [47, 51], state-of-the-art continuous sign-recognition software is still imperfect. To mitigate inaccuracies in the video-to-sign matching, some proposed dictionaries provide users with post-query filtering options, to narrow the set of results that are returned [26]. Overall, existing dictionary systems face several limitations: They expect that the ASL learner can recall linguistic properties of the desired ASL sign or accurately perform the sign from memory. In addition, systems assume that a user is able to precisely identify the starting and ending of a sign they encountered in a video or in a conversation. Fast signing speed or various linguistic factors (section 2.1) make it difficult for ASL learners to precisely select signs in videos. Finally, the user must launch an additional task of querying a dictionary, while engaging in a video-watching and comprehension task. Using separate tool to perform the search may cause users to lose context from the video they were watching.

In contrast to prior work, we investigate a dictionary-search system that enables the user to select a span (of potentially multiple signs) from a video of continuous sign language, as a basis for a query to search for potential matching signs from a dictionary system, with the results presented in an integrated video-player and search-results interface. This approach may mitigate the need for users to recall specific linguistic properties of the unknown sign, mitigate the need to identify the specific start/end of signs in a continuous video, and enable the user to remain in context in their video-watching-and-comprehension task.

Some recent research has investigated ASL learners interacting with Wizard-of-Oz prototype systems for ASL dictionary search, to identify factors that affect users' satisfaction [3, 23–26]. Methodologically, our studies also employ a Wizard-of-Oz prototype of an ASL dictionary-search system to understand users' interaction and potential benefits. However, in those prior studies, users had been shown a stimulus video of native signer performing a single isolated sign (in citation form), and the user was asked to use a dictionary system to identify the sign's meaning. In contrast, our studies examine how ASL learners engage in a search task while in the midst of a video-watching-and-comprehension task.

2.3 Research Questions

There has been limited prior research on the experience of ASL learners who are engaged in the educational activity of watching a sign-language video that is difficult for them, especially how technologies for ASL dictionary look-up may benefit these users. In addition, no prior work has examined how users might benefit from an integrated tool for viewing ASL videos, with users able to select spans of the video as the basis for dictionary search. To address these gaps, we investigate the following research questions:

RQ1 What are the challenges that ASL learners currently experience when trying to understand a difficult sign-language video, and what workarounds do they employ?

RQ2 Comparing the experience of users who used our tool and those who used an existing feature-based ASL-English reverse dictionary, is there a difference between translation quality or perceived workload to produce translation?

RQ3 How do users interact with a Wizard-of-Oz prototype for viewing an ASL video and conducting dictionary-search on selected spans of video, during the video-watching and comprehension task?

3 STUDY 1: INTERVIEW STUDY

This paper presents two studies: The goal of study 1 was to understand ASL learners' challenges with video comprehension and current workarounds they use. The findings from study 1 informed the study design, videos, and prototypes included in study 2.

3.1 Study Design

This IRB-approved study was conducted in person or remotely based on the preferences of participants during the COVID-19 pandemic. After informed consent was obtained, the semi-structured interview began with questions about participants' prior experiences watching ASL videos. They were asked about the type of videos they watch, their experiences when they have difficulty understanding, and any workarounds they use. To provide context for later questions about how difficult it may be for participants to select an individual sign or span of multiple signs they do not understand, we displayed several example videos to participants as a basis for discussion. These videos were taken from advanced ASL or ASL-English interpreting classes, conversational videos between expert signers on YouTube, signing performances at theatres, and interpreted poetry and music. (Video details appear in electronic supplementary files.) Videos contained a variety of linguistic phenomena discussed in section 2.1. Participants were asked how hard it would be to select a sub-span containing one or multiple signs and how they would select a time-range of a video. The average length of each interview was 36.5 minutes (σ=5.46 minutes).

3.2 Participants and Recruitment

Participants were recruited by posting an advertisement on an ASL Reddit channel and by contacting professors of introductory ASL courses, who shared an advertisement by email with their students, containing two screening questions: "Are you currently learning 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. We recruited a total of 14 participants for our first study, which included, 2 men, 11 women, and 1 non-binary individual. The median age was 21 (σ = 3.67). Participants had studied ASL for a mean of 3.4 years, and all participants confirmed that they had taken fewer than 3 years of formal ASL classes.

3.3 Analysis and Findings

We employed a mixture of deductive and inductive approaches in our qualitative data analysis. To become familiar with the interview transcripts, two authors read all 14, then during a subsequent reading, they individually took notes to produce initial codes, which they collated and collapsed into two individual code-books. Each of the authors then investigated underlying patterns among their codes and formed initial categories, and they consulted the interviewers to get feedback on their initial categories and further improved them. The authors then met to review all of their initial categories, to identify similarities and differences. During two three-hour meetings, the authors performed an initial thematic grouping, which lead to final high-level categories. These high-level categories were then presented to the rest of the team to arrive at a final set of themes and sub-themes presented in this section.

3.3.1 Prior experiences and challenges associated with watching signed video content. Participants discussed various **motivations** for viewing signing content. Twelve participants mentioned engaging with ASL videos during classroom-related or homework activities. Ten participants also mentioned watching signed content outside of the classroom for their own enrichment or personal exposure to other types of signing, e.g., Deaf theatre or ASL songs. P11 said, *"It's both in-class, we have different assignments the teacher will give us, and then I also do it on my own time, if I'm looking for a deeper understanding about things, or if I'm looking for specific signs. And I also follow some deaf content creators as well."*

Participants also discussed how their lack of familiarity with **regional or dialectical variation** in signing, such as Black ASL [43] used among some African-American signers in the U.S., led to challenges in understanding videos. P5 described their experience in understanding signing among various communities: *"I know some white people in the community, [but the] black Deaf community and the interpreter community, I still find hard."*

Participants discussed how various **linguistic types of signs** posed comprehension challeges. For instance, P1 described needing to consciously *"switch my brain from a sign to actually each letter"* when encountering fingerspelling. P11 discussed challenges with *"fingerspelling, classifiers, compound words, any of that kind of stuff... fingerspelling is definitely a little tougher for me."* Participants discussed how fingerspelled names were challenging to understand in a video, especially when there were multiple individuals with similar names. Participants also described challenges with understanding numbers, e.g., P5 said, *"numbers are hard for me, for some reason, I don't know why."* Participants also discussed challenges with compound signs, e.g., P13 said, *"I wasn't sure if that was one or two separate signs. So there were definitely points in the video where they were blending together a little bit, and I wasn't sure."*

Overall, participants discussed how **different content sources or genres** pose challenges for comprehension. Participants mentioned viewing signed content on various streaming services, e.g., YouTube and Netflix, as well as on social media, e.g., Instagram and TikTok. P14 said, *"I watch ASL videos when I am going through Instagram because I follow some Deaf creators."* Participants discussed how the signed content on social media is shorter and more unpredictable in nature, with the topic of the video not always well defined, which poses challenges for comprehension. Participants also discussed how factual signing, e.g., in a documentary, was difficult, due to complex vocabulary or increased use of fingerspelling. Other participants mentioned watching videos of ASL poetry and ASL translations of popular songs, contexts in which they described signers as using more depiction and having *"their own flow, and they have their own rhythm"* (P11). Participants described how videos with multiple signers, e.g., Deaf theatre, pose challenges, as P8 described, *"my brain is used to practicing with one signer."* Similarly, participants mentioned how natural conversations were difficult to understand, e.g., P6 discussed how signing in such videos tends to be *"quicker, and they're a little bit more relaxed."*

3.3.2 Workarounds. Participants mentioned several workarounds that were useful in understanding challenging signed video content. For instance, several participants discussed using the **context** of a video to understand unknown signs. Participants would consider the description or title of the video, such as on YouTube, or they would consider what was said before or after any unknown signs. For instance, P3 described a situation in which they figured out the sign for a citrus fruit by considering the context of the surrounding signing, which had mentioned lemons. P3 discussed how understanding later signing may clarify a portion of signing that had not been previously understood, explaining how if they become confused then they *"really focus on the next thing they're saying, so I can piece together what they might have said, so I can understand it."* Participants discussed various strategies that involved controlling the flow of the video player:

- **Periodically pausing** was a strategy among several participants. For instance, participants discussed how they paused videos in-between conversational turns in videos with multiple signers; P8 described how they *"pause in between each speaker... just enough time to grasp"* what had been said.
- **Backtracking and replaying** was another common approach, as P12 explained, *"pausing it and replaying it."* P3 also discussed how they will *"backtrack the video"* if needed.
- **Slowing down the video** was also popular, if possible within the video player. For instance, P7 explained how they will *"slow down the fingerspelling if...it's on YouTube. If I could alter the speed, I might try to slow it down."*

Regarding current strategies for seeking the meaning of an unknown sign, six participants mentioned using **English-to-ASL dictionaries**, i.e., guessing English meanings of the sign they did not understand to look up that English word in the dictionary to see if the sign displayed visually matched the sign that had not been understood. Participants also mentioned using **ASL-to-English "reverse" dictionaries**, i.e., websites that allow someone to enter linguistic properties to search for the English translation of a sign. Participants discussed challenges, e.g., P7 said, *"if I think I have an idea of what the sign is I might use Handspeak, or there's another one I use... [It's] hard to specify handshape in current dictionaries."* P6 discussed struggling to enter linguistic properties when constructing a query: *"I definitely tried using the reverse dictionary stuff online. Usually it doesn't end up being successful, and I have to just end up moving on. Because, the way it's structured, you have the* *handshape, and the movement, and the location. Sometimes it’s a little ambiguous, especially if you don’t actually know what that sign is; so, it’s hard to end up looking [it] up."* Participants also expressed their frustration with having to launch a web-dictionary in another window while trying to understand a video. P11 said, *"It’s pretty frustrating sometimes when I’m trying to find a specific sign, I have to like go to Google and...then go through all the different pages... If I could just scroll and have the source material right there, I think it would be much more efficient."*

Rather than use a specific dictionary website, other participants mentioned typing descriptions of what a sign looked like into a **Google search**, e.g., P11 said, *"I’ve definitely tried to Google it before, but it’s so hard to sometimes describe what it is that you’re looking for. I end up being very vague... It’s very rare that I go to Google and find what I’m looking for as far as trying to describe a sign."* Finally, several participants mentioned that, if other people are available, they may **ask a teacher or a peer**. As P09 said, *"If I’m in class I would ask the teacher. If it’s for a class I would either look it up online or if it’s in a vocabulary learning unit."*

4 STUDY 2: PROTOTYPE STUDY

While study 1 motivated the need for a tool that provided an integrated video-playing and sign-search experience, study 2 investigated how users would interact with such a tool and whether there are benefits as compared to existing systems. In this prototype study, users interacted with a Wizard-of-Oz prototype, at a desktop computer in a lab setting. Findings from study 1 informed both the design of the prototypes used and the selection of videos included in both the prototypes, as described below.

4.1 Prototype Design

4.1.1 Integrated Search. Since the focus of this study was on users’ interaction and behavior, an interactive Wizard-of-Oz prototype was designed (Figure 1), in which the underlying sign-recognition technology was simulated, without any automatic video analysis.

On this web-based prototype, participants entered a participant ID on the first screen. Next, they were provided a calibration screen to ensure that the size and aspect ratio of their browser window was consistent. As shown in Figure 1, the interface displayed an ASL video with a play/pause button at the bottom-left corner of the screen. On a video-timeline at the bottom of the screen, a vertical white line indicated the video playhead (the current position of the video). The users can select a video span by dragging yellow edges of a selection bar on this timeline, and they can press a "Play selection" button to play only the portion of the video in that span. Once satisfied with their span selection, users can click on the yellow "Search selection" button on the top right corner of the screen to search for the signs. The results were displayed in a scrollable window on the right side of the screen. Each result consisted of a video of a sign from the American Sign Language Lexicon Video Dataset (ASLLVD) [50] and a label below showing the closest English gloss for that sign. When clicking on a "more information" icon for each search result, linguistic properties for the item were displayed, as illustrated on the right side of Figure 1.

Study 1 findings informed the prototype design: For instance, participants expressed frustration with needing to leave the context of watching a video to use electronic dictionaries, and this informed our decision to display dictionary results on the same page. In addition, participants’ explanation of workaround strategies they used when encountering difficult video, e.g., re-playing and backtracking a video, led us to provide the "Play selection" button, which played the video while the playhead was constrained to the selected span–to support users in replaying a short segment of the video.

Findings from study 1 also informed the selection of videos shown to participants during the study. Based on participants’ comments about how the genre or linguistic phenomena in a video relates to how challenging it is, we selected three genres of videos to display in this study, including educational videos, conversational videos with at least two signers, with turn-taking between them, and Deaf theatre and poetry videos. Since participants had discussed how they face particular challenges when encountering fingerspelling or compound signs, we ensured that the videos included instances of these types of signing. In addition, we selected videos that included multiple signers engaged in conversational signing, as well as a video with different regional dialects of signing, since both had been discussed by participants in study 1. The videos used in this study were obtained from online sources and material from fourth-year ASL interpreting courses. Videos were an average of 23.7 seconds in length, and details about each video are provided in electronic supplementary files.

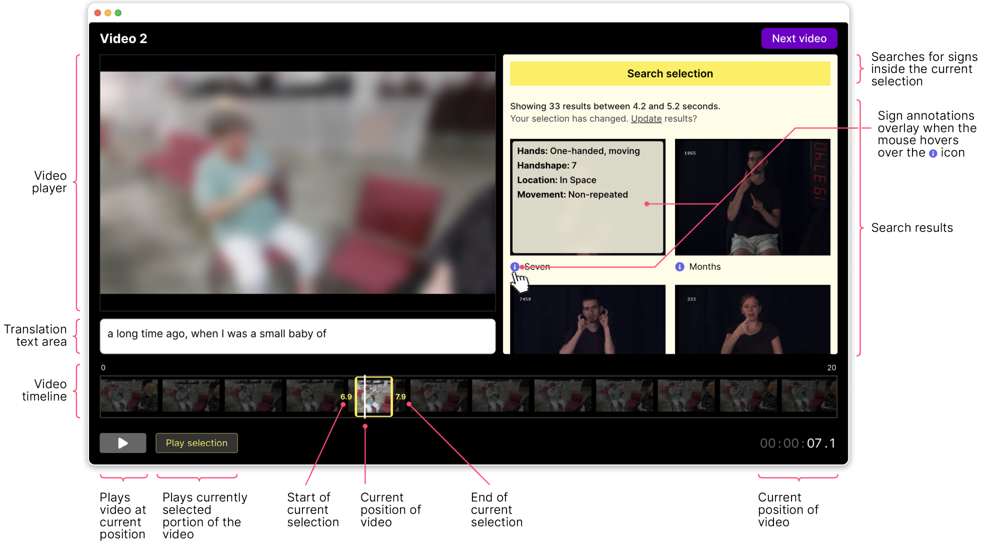
4.1.2 Selection of signs appearing in the results list. Given the rapid pace of advancement in the field of sign-recognition technology, and since the purpose of this observational study was to understand participants’ behavior and interaction, we selected a Wizard-of-Oz approach to simulate an automatic search-recognition system. Therefore, our system returned a pre-determined list of results, in which the actual sign appeared somewhere in the results list. The selection of signs that appeared on the results was based on pre-processing of videos. The protocol is described below:

(1) A Deaf member of our team with native ASL fluency watched all 9 of the ASL videos in advance to identify the sequence of signs appearing in each video, along with the starting and ending time-stamps of each sign.

(2) For each sign, a set of dictionary-search results were manually prepared, to simulate the type of results someone would see if using a real automatic dictionary-search system in the future. Specifically, for each sign, a native signer carefully selected the closest match and 11 other signs that were similar in appearance to the sign, from a collection of signs from the ASLLVD [50]. The researcher prioritized selecting signs for this “match list" with as many properties in common as possible to the given sign, i.e., the same handshape, number of hands, movement, and location.

(3) When a participant selects a span of video, it is possible that the start or end of the span is within the duration of a sign in the video, rather than precisely at a boundary between signs. We established a rule that the prototype would consider a sign to be within a span selected by the participant if at least half of the sign appears within the selected span.

(4) Since a selected span may contain multiple signs, the list of dictionary-search results displayed combined results from the match lists for the all signs within that span, as follows:



Video 2
Next video
Search selection
Video player
Translation text area
Video timeline
Searches for signs inside the current selection
Sign annotations overlay when the mouse hovers over the ℹ icon
Search results
Seven
Months
1865
7458
333
0
20
6.9
7.9
Play selection
00:00:07.1
Plays video at current position
Plays currently selected portion of the video
Start of current selection
Current position of video
End of current selection
Current position of video

Showing 33 results between 4.2 and 5.2 seconds.
Your selection has changed. Update results?

Hands: One-handed, moving
Handshape: 7
Location: In Space
Movement: Non-repeated

a long time ago, when I was a small baby of

Figure 1: Screen image of the prototype, displaying labels for the various interface regions, including the video player at the top left, a text box where translations can be typed below, a video timeline with a span-selection interface along the bottom, and a dictionary-search panel on the right.

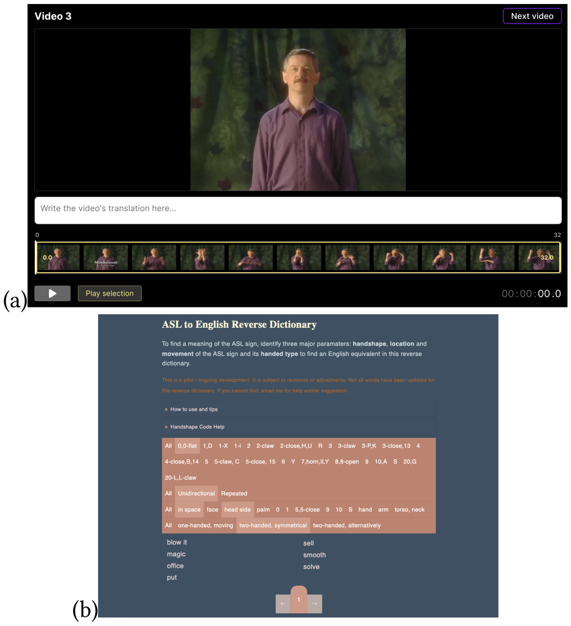
From among the match lists for all signs in the span, one match list was selected randomly, and the top item from that list was taken (without replacement) for inclusion on the combined results list. This process was repeated until all match lists were empty, to produce a combined list of results that contained the union of the original match lists. The first 50 items from this combined list were displayed to users.

4.1.3 Baseline Prototype. We also designed a "baseline" prototype identical to the one described above but without the integrated dictionary-search option, as shown in Figure 2(a). When viewing a video with this baseline, participants were instructed to open a second web-browser window to use the handSpeak reverse dictionary [1], as shown in Figure 2(b). That site enables users to look for an isolated ASL sign by selecting text labels that represent various linguistic properties of a sign, e.g., handshape, hand location, movement, and orientation. Based on the query options selected, results appear as a list of English gloss labels for matching signs. Participants were not allowed to visit other websites or other resources.

4.2 Study Design and Analysis Plan

After providing informed consent in this IRB-approved study, participants were asked to view a video and produce an English translation text for it using the integrated-search prototype or the baseline prototype. Using a sample video, the researcher first demonstrated the prototype (details in section 4.2). After indicating that they understood the prototype, each participant viewed 9 videos. The order of videos was randomized. Their interaction was recorded:

(1) The software prototype was designed to automatically record the starting and ending points on the video timeline of every



Write the video's translation here...

ASL to English Reverse Dictionary

To find a meaning of the ASL sign, identify three major paramaters: **handshape**, **location** and **movement** of the ASL sign and its **handed type** to find an English equivalent in this reverse dictionary.

This is a pilot / ongoing development. It is subject to revisions or adjustments. Not all words have been updated for this reverse dictionary. If you cannot find, email me for help and/or suggestion.

blow it	sell
magic	smooth
office	solve
put	

Figure 2: (a) The prototype for the baseline condition in Study 3, identical to the one from Study 2 but without any dictionary-search ability. (b) The handSpeak ASL-English reverse dictionary website, which participants used in the Study-3 baseline condition. As users click on linguistic properties, the list of English gloss labels at the bottom of the window updates to list matching signs.

[1] https://www.handspeak.com/word/asl-eng/

span the participant selected, the number of signs within each span, whenever the participant triggered a search, and the text of the English translation typed by the participant.

(2) Each participant's face was approximately 65cm from a 19-inch monitor, to which a Tobii Nano [49] 60Hz screen-based eye-tracking device was attached. iMotions (v9.1) [30] software recorded each participant's gaze.

(3) A researcher, who was a fourth-year English-ASL interpreting student at the university, sat 2m away and took observational notes during the experiment. The iMotions software enabled the researcher to monitor the participant's gaze on the user-interface in real-time on a secondary display.

At the end of the entire session, a debriefing interview was conducted to gather participants' impressions of the system, perception of how they interacted with the device, and other recommendations. The interview data was transcribed and coded using the same methodology as in study 1.

Qualitative analysis of the data listed above was performed by two members of our research team who reviewed and coded this data, from the perspective of identifying typical sequences of interaction behavior during each video session. They reviewed recordings of the screen and eye-gaze, plotted eye-gaze patterns, analyzed data captured by the software prototype, and reviewed the observer's notes. The researchers discussed their notes and agreed upon a categorization of the behaviors observed, as presented in Findings section 4.5. After viewing and translating each video for both conditions, participants' English translation texts were saved, and participants completed a NASA TLX [21, 22].

4.3 Participants and Recruitment

We recruited a total of 15 ASL students for study 2, using recruitment criteria and approaches identical to study 1. 8 participants were assigned to the integrated search condition whereas 7 were assigned to the baseline condition.

The median age of participants in the integrated search condition was 20, and this included 7 women and 1 non-binary individual. Participants had studied ASL for a mean of 3.5 years, and all participants confirmed that they had taken fewer than 3 years of formal ASL classes.

The median age of participants who used the baseline prototype was 21, and they included 4 women and 3 men. A single recruitment process was conducted and participants were randomly assigned to either the prototype-with-dictionary-search or baseline-prototype condition. Participants using the baseline prototype had studied ASL for a mean of 3.7 years, and all participants confirmed that they had taken fewer than 3 years of formal ASL classes.

4.4 Findings: Comparative Study

To assess translation quality, we adapted a prior approach [9], in which a human judge looked for translation errors in a text (e.g., wrong or omitted words) and then assigned an overall translation-accuracy score (out of 10). In our study, a fourth-year ASL interpreting student, who had completed a university course on ASL linguistics, analyzed the transcripts to identify errors and assign translation-accuracy scores—without knowing which translations had been produced using the dictionary-search prototype and which had been produced using the baseline prototype. The average translation-accuracy score assigned by the researcher was 8.03 for translations produced using the dictionary-search prototype and 6.67 for the baseline prototype. The distributions in scores between the two conditions differed significantly (Mann–Whitney U = 10, n1=8, n2=7, P = 0.0424 < 0.05 two-tailed, r = 0.24).

Figure 3 shows scaled mean raw NASA-TLX sub-scores (physical demand, temporal demand, performance, effort, and frustration) and results of two-tailed Mann-Whitney U tests comparing sub-scores across conditions. Participants who used the prototype with dictionary-search gave significantly lower values for mental demand (how much mental and perceptual activity was required), temporal demand (how much time pressure was felt), and frustration (how insecure, discouraged, irritated, or stressed they felt). The NASA TLX instrument and details of these scales appear in [21, 22].

4.5 Findings: In-depth Analysis

We present an in-depth analysis of how participants interacted with the integrated search prototype. Since prior work had investigated the experience of students with existing dictionary-search websites, e.g., [6, 26], we do not present a detailed observational analysis of users interacting with the baseline, given the limited novelty of such an analysis, amid prior literature.

4.5.1 Using the span selection to constrain the playhead. Six participants used the span selection tool to constrain the portion of the video that would play at one time, to enable them to progress incrementally through the videos. Participants selected spans of average duration 11.43s for viewing videos in this manner, and they progressively selected spans of video as they typed the English translation text. Figure 4(a) illustrates this trend, by plotting the positions of spans a participant selected over time. Span 1, shown at the bottom of the image, indicates the first span selected.

Comments from debriefing interviews also support this observation. P7 described how they selected a span of a particular width, and then dragged it along the video, to watch portions of video progressively: *"I like how you can just maintain the length and you just drag it over so you're getting the same length of a chunk of the video; that was easy to use."*

4.5.2 Approaching task linearly, sometimes after initial overview. Participants viewed the videos in a linear manner and produced transcripts as they watched short segments of video. In some cases, participants first viewed the entire video, and then they returned to the beginning of the video to progressively view short segments of video in a linear manner, as illustrated in Figure 4(b), which shows a full-video span prior to progressive short spans.

4.5.3 Using dictionary search to inform translation. In 62 out of 72 video sessions, participants made use of the dictionary-search feature to look up the meaning of unknown signs in the video. As illustrated in Figure 5, the results of the search tool informed participants' translation decisions as they linearly progressed through the video. During the debriefing interview, P6 described how the tool helped: *"I knew what he was saying in general, I just couldn't think of the exact English words and that one came up right away."* P7 discussed the benefits of the tool during fast signing, *"It was definitely useful, especially when the signers were going really fast*

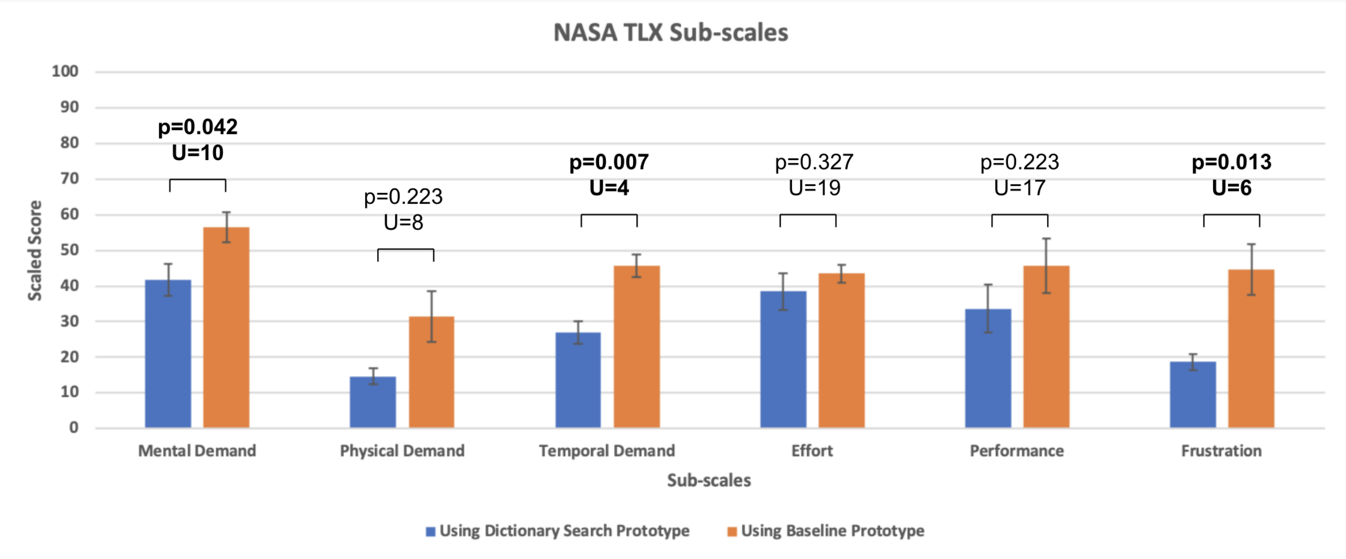


Figure 3: NASA TLX sub-scale scores from participants in the dictionary-search (n=8) and baseline prototype (n=7) conditions, with scores scaled to a 0-to-100 range. For all sub-scales, lower scores are better, i.e., indicating less perceived demand, less effort needed, less frustration, or better sense of performance success. Significance testing results from two-tailed Mann-Whitney U tests are also presented on top of the bars.

because then I could double check to make sure that what I thought I saw was actually what I saw."

4.5.4 Gradually making the span shorter prior to search. When encountering a difficult portion of video, participants often reduced the size of their span before initiating a dictionary search. When selecting a span to view a portion of video (and not to conduct a search), the average span width was 8.17 seconds (10.83 signs), while the length of spans immediately prior to a search request was 2.33 seconds (3.25 signs). If a search result did not enable a participant to identify the meaning of signs in a particular span, some participants progressively narrowed the span width, to more precisely select the specific video portion where they were confused, and then requested additional dictionary searches, as shown in Figure 6(a). While adjusting spans, participants' gaze alternated between the main video region and the span-selection control, as shown in 6(b).

In debriefing interviews, participants described watching longer segments to understand the context and then narrowing in on a shorter span that was difficult to understand. P5 characterized their approach as *"narrowing it down and then pressing search."* Other participants discussed the benefits of beginning with a search of a wider span initially, e.g., as P7 explained, *"if I was just trying to get the general idea of a section, it was helpful that sometimes there were more signs in the up results besides like the specific signs that I had selected because it gave more context and it was easier to understand."*

4.5.5 Using dictionary-search to confirm results after initial translation. Among the 62 video sessions in which participants made use of the dictionary-search tool, in 40 cases, we observed participants using the search tool after they had already completed a full translation of the entire video. As illustrated in Figure 7, after writing a translation for the whole video, the participant reviewed specific earlier portions, using the search tool, to confirm that they had correctly understood specific signs.

During debriefing interviews, participants talked about various ways in the search tool was useful in producing a more accurate translation. For instance, P7 discussed how the search results motivated them to adjust their wording in the translation, e.g., saying, *"I would go back and use the tool to make my translation more precise I guess. So, I could fix the sentences and the wording."* Other times, the search results simply boosted their confidence in how they had understood the video, e.g., with P7 saying *"sometimes that helped to confirm what I thought I saw."*

4.5.6 Participants struggled to find a sign if a different version was being signed in the results. In several instances, participants were confused when the citation-form of the sign displayed in the search results did not match the variation of the sign in the main video, often in the case of compound signs and depiction. In Figure 8, P5 performed a search, but the specific appearance of the sign in the video differed from the citation form shown in the dictionary-search results, which led the participant to glance back and forth between them to compare. Ultimately, the participant did not produce the correct translation for that portion of the video, suggesting that they did not realize this was a match to what they had seen.

In debriefing interviews, participants how matching dictionary-search results to signs in the video was more difficult for some video genres. For instance, P4 discussed how when a sign was produced with great emotion, e.g., in a theater video, then it was difficult to match it to a dictionary-search result with more neutral affect. P4 described their difficulty with a video in which the signer *"was showing emotion and then you would go in the searches and they wouldn't. So it's like, I guess you can get mixed up about the emotion."* Participants suggested that the dictionary-search system could be

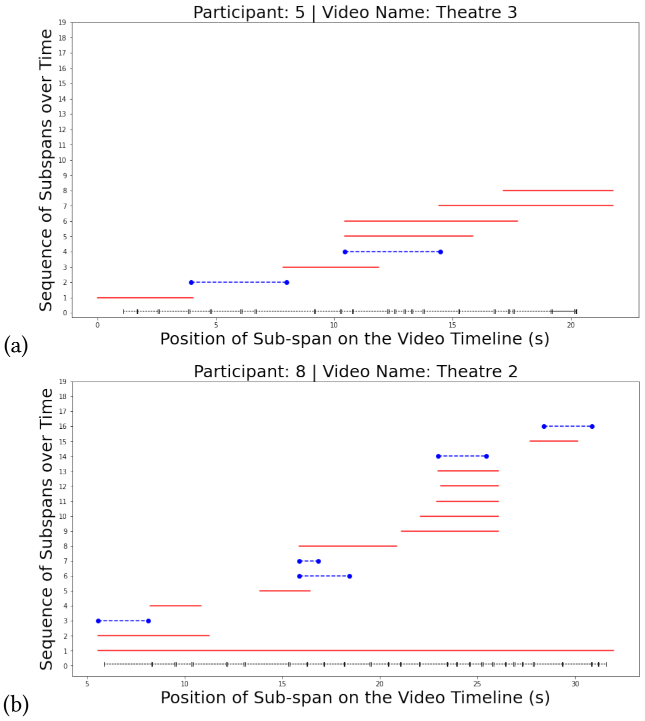
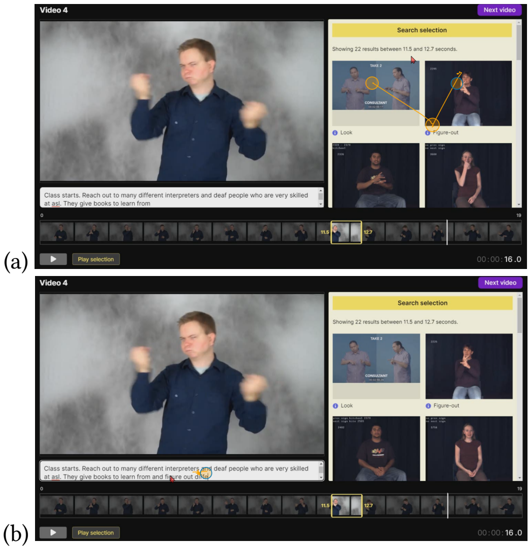


Figure 4: In (a), P5 viewed a video in a linear manner, using the span selection to progressively view short video segments. Horizontal bars indicate spans selected with respect to the total duration of the video. The first span selected is at the bottom of the y-axis, with subsequent spans appearing higher on the y-axis. Spans in blue dotted lines are those for which the participant pressed the "Search" button to conduct a dictionary search. The black lines at the bottom show the actual signs on the video timeline. In (b), P8 first selected the entire video, played it, and then reduced the span to a smaller duration and moved the span forward progressively.

improved by providing dialectical variations of each sign result and an example of each sign's use in a sentence, e.g., P3 said, *"It would be nice to see the sign in a context with more facial expressions or in a sentence like you have in other dictionaries."*

4.5.7 *Differences in span selections across genres of videos.*

As described above, the 9 videos in the study were from three genres: natural conversations, educational videos, and theatre/poetry performances. An analysis of the span-selection data captured by the prototype revealed that participants selected wider spans for theater videos, as illustrated in Figure 9. When users were selecting a span simply to constrain the playhead to view a portion of the video, a Kruskal-Wallis test [H(2)=24.28, p<0.00001, η^2 = 0.043 (Small Effect)] with Mann-Whitney post-hoc testing with Bonferroni corrections, revealed that users selected wider spans for theatre videos. Similarly, when users were selecting a span as input to a search, the testing [H(2)=24.3031, p<0.00001, η^2 = 0.12 (Medium Effect)] revealed that users also selected wider spans for theatre videos.



(a) Showing 22 results between 11.5 and 12.7 seconds.

Class starts. Reach out to many different interpreters and deaf people who are very skilled at asl. They give books to learn from

(b) Showing 22 results between 11.5 and 12.7 seconds.

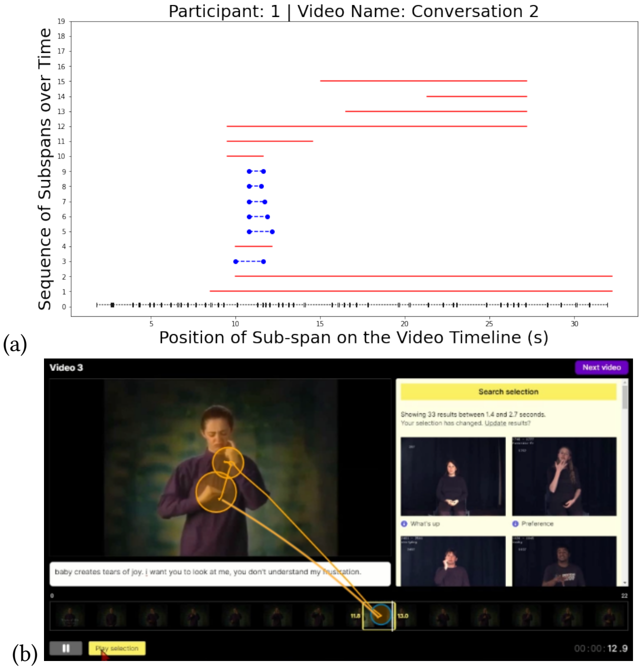
Class starts. Reach out to many different interpreters and deaf people who are very skilled at asl. They give books to learn from and figure out diffe

Figure 5: P5 using the tool to produce a translation: (a) browsing search results for a span and (b) after identifying the meaning of the sign FIGURE-OUT, typing into the translation text box to continue the sentence: "They give books to learn from and figure out..."

In debriefing interviews, participants discussed why they selected spans of different widths for different genres. P6 discussed how theatre/poetry videos were challenging to understand generally, *"This one had more of a poetic meaning and display; so, it did take more focus to understand it for the translation."*. P5 described selecting spans while watching conversational videos in contrast to the theatre/poetry performances: *"the conversational ones, when I chose these subspans were shorter, because they go back and forth a lot [between signers]. But the poetry ones I feel are more conceptual... so you can watch longer pieces. You don't need to cut it down."*

P7 discussed how the width of selected span depended on the overall signing pace, but longer spans were generally needed for theatre/poetry videos due to the style of signing: *"Some of the ones that were very visual, like the mushroom one and the moon one and all those ones that were ASL storytelling type of very figurative language... It's typically slower paced, and sometimes there's a lot of repeated signs. Or there's a lot of just a depiction that's very visual and doesn't have a lot of strictly vocabulary to go with it, but it's more classifiers. I found that I would sometimes need a longer chunk in order to use the tool and actually get relevant results of what was being signed."*

Participants were free to select spans that did not align precisely with the boundaries of when one sign ends and the next begins; however, the results of dictionary-search could be controlled more precisely if spans were selected more accurately. An analysis of the mean error (in seconds), between each span selection boundary and the nearest actual sign boundary, reveled that users were less accurate when selecting spans during theatre/poetry videos. The mean error for natural conversation videos was 0.19 seconds, for educational videos was 0.23 seconds, and for theatre/poetry videos was 0.55 seconds. A Kruskal-Wallis test [H(2)=5.2174, p=0.02236, η^2 = 0.041 (Small Effect)] and Mann-Whitney post-hoc testing with



Showing 33 results between 1.4 and 2.7 seconds.
Your selection has changed. Update results?

baby creates tears of joy. i want you to look at me, you don't understand my frustration.

Figure 6: P1 adjusting span width over time, while performing repeated search requests. (a) Blue dotted lines indicate spans for which dictionary-search was requested, and red solid lines indicate spans for which no search was requested. For spans 7 to 10, the width reduces over time. (b) When fine-tuning span width, participant 8's gaze moves between the video region and the span control. The yellow lines going from the span to the video are showing the gaze pattern.

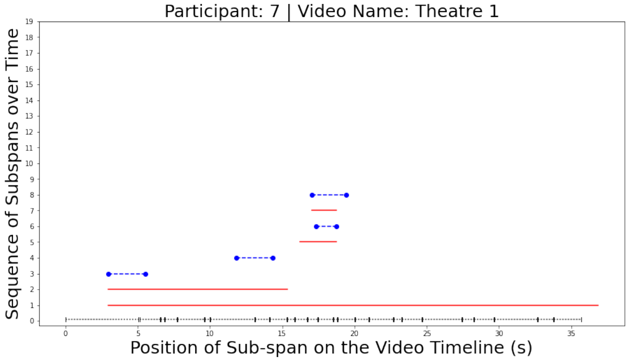
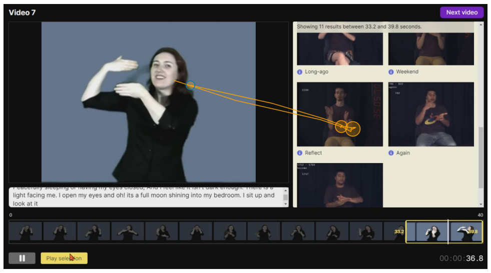


Figure 7: P7 had already completed an English translation for the entire video, and then they returned to a few earlier regions of the video and requested dictionary searches to confirm their translation for specific segments of video.



Showing 11 results between 33.2 and 39.8 seconds.

light facing me. I open my eyes and oh! its a full moon shining into my bedroom. I sit up and look at it

Figure 8: P5 glanced between a dictionary result for REFLECT and corresponding portion of the video, where the sign had been produced in a different manner, leading to P5 to believe they were different signs and ultimately producing an incorrect translation. The yellow lines going from the span to the video are showing the gaze pattern.

Bonferroni correction, revealed that the error in the case of theatre/poetry videos was higher than for the other two genres. In debriefing interviews, P2, P6, and P7 discussed how aligning span selection with actual sign boundaries was more challenging for theatre/poetry videos. P7 explained that the type of signs within such videos was a factor, *"because of depiction... there weren't clear boundaries as much in the signs because it was using classifiers."*

5 DISCUSSION

While prior second-language pedagogical research had investigated the challenges faced by students trying to understand texts with difficult vocabulary [16], no prior study had investigated ASL students' experiences and challenges through firsthand interview and observational methodologies. Our participants described how dialectical variation, linguistic types of signs, and various genres of ASL content led to comprehension challenges. We investigated how students currently approach understanding ASL videos containing unknown signs, revealing how students turn to online video streaming, video sharing, and social media sites in order to gain experience at understanding more diverse and natural examples of signing. While the proliferation of video social media and streaming services has diversified and increased available ASL content [60], most prior ASL pedagogical research on video comprehension predates the ubiquity of such sources, e.g., [14]. Our findings motivate the need for a tool to (a) support students viewing videos they seek from **diverse sources** and (b) with **challenging content** to support developing comprehensions skills.

Study 1 also investigated students' current workarounds for videos with difficult signing, revealing their dissatisfaction with existing ASL dictionary resources. These findings aligned with prior work, e.g., [6], on the need for better tools to enable students in identify the meaning of an unknown sign. A unique focus of our study was that participants considered dictionary-searching challenges within the context of trying to understand a difficult video: We found students were frustrated at needing to leave their video to look up a sign in a separate website and their use of workarounds like repeated pausing and rewinding of videos. These findings specifically motivate HCI research on tools that support: (a) **viewing and repeating** short spans of video and (b) **integrating** the dictionary-search tool into the video-playing experience.

Study 2 investigated the **potential benefit of these tools**. It first compared the full prototype and a baseline prototype in which span selection was still available for constraining the playhead, yet the students had to use an external, existing ASL dictionary website

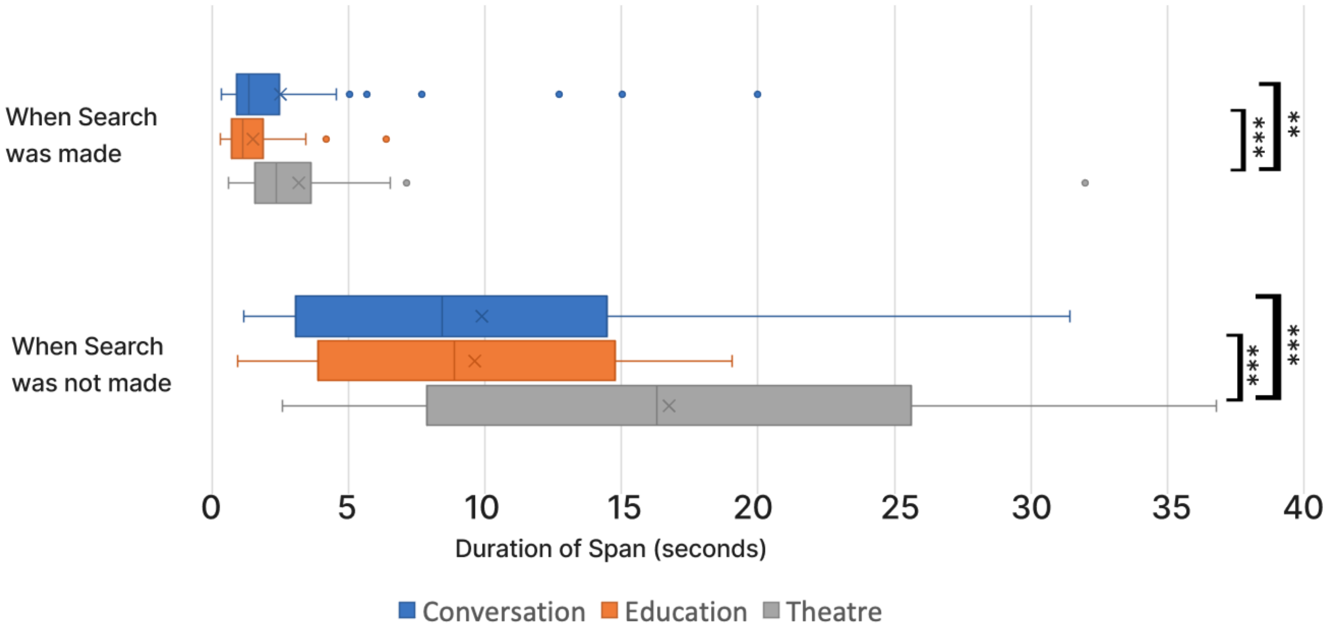


Figure 9: Box and whisker plots illustrating span widths for each video genre, in two cases: spans selected immediately prior to a search (top of graph) and spans selected not immediately prior to a search (bottom of graph). In each case, participants selected wider spans for videos in the theater genre. Median are indicated by vertical bars within each box, means indicated by x, and outliers by dots. Significant pairwise differences are marked with asterisks: $**$ if $p < 0.01$, and $***$ if $p < 0.001$.

as a reference tool. Our findings revealed that the integrated search tool led to translations of greater accuracy and participants rated the workload lower. Prior work had examined at how using in-situ dictionaries improve comprehension for non-native speakers [36], but this was the first study to explore this in the sign-language context. Study 2 also investigated how users would actually interact with a such a system, using a Wizard-of-Oz prototype methodology. While prior HCI research had investigated students interactions with ASL dictionary-search interfaces, e.g., [3, 6, 25, 26], that work had focused on students trying to look up a specific sign from their memory or search using a video of an isolated sign. Our novel observational findings include how students engage with a dictionary search tool **within the context of the overall video comprehension task**, e.g., replaying and comparing the portions of the original video to potential sign matches. These observational findings in Study 2 aligned with the interview-based findings from Study 1, in which students had expressed frustration at losing their video-watching context when using a separate dictionary tool.

Our observational findings also revealed how students engaged in a **dual use** of the video player’s span-selection interface: (a) to constrain the playhead to progressive portions of video and (b) to specify input for dictionary search. Notably, usage (a) served as a probe within our analysis to identify the “window size" of video that students viewed as they worked through the challenging video. While some prior research on video analysis or annotation tools for linguists analyzing ASL videos had incorporated a span-selection interface for labeling regions of videos, e.g., [39, 50], no prior research had investigated span selection in the context of ASL learners viewing video. Our findings therefore motivate further HCI research on span-selection interfaces within this new context.

Our analysis also revealed that video players with span-selection interfaces may be a **useful probe** during studies in which participants use such a tool to view videos may reveal their comprehension strategies. For instance, our analysis revealed differences in span widths that users selected for different genres, and participants discussed how their selection of span width related to the linguistic properties, with wider spans for challenging theater/poetry videos. Similarly, when span-selection boundaries were compared to actual sign boundaries in videos, we observed more error during theater/poetry videos. Thus, post-study analysis of the spans selected by individuals who view ASL videos may reveal insights for ASL linguistics or education researchers, and real-time analysis of spans could enable adaptive educational software capable of identifying when students are currently struggling while viewing a video.

Overall, the findings of this paper are relevant to **several audiences**: The current experience and workarounds of students that motivate research on integrated tools for students viewing challenging ASL videos will inform the work of accessibility and HCI researchers. Students’ perspective on factors that lead to challenges in ASL video comprehension, as well as the potential of span-selection video players as a research tool, will be relevant to ASL linguistics and education researchers. In addition, our findings on the benefits of ASL dictionary search using extracted spans of continuous-signing video informs the work of computer-vision researchers. Specifically, our findings reveal a whole new sign-recognition task of using a sub-segment of a continuous video as input for ASL recognition. Traditionally, when ASL recognition researchers consider processing continuous ASL videos, it is in the context of machine-translation to English text. Rather than providing fully automatic translation of a video, our findings have revealed interest and benefits for students in attempting to understand a challenging ASL video on their own with some integrated sign-searching supports. In addition, compared to recognition of videos of entire utterances, there are unique challenges when the input is a fragment extracted from a longer video. The selected span may not perfectly align with sign boundaries, the signs at the boundary may have been subject to co-articulation effects of signing beyond the span selection, and there will be less contextual information for the recognition system to consider.

More broadly, our findings speak to the literature on users **interacting with videos**, especially when carefully scrutinizing video, e.g., when interacting with specialized video-editing software. While span-selection is less prevalent in video-player systems, several commercial video-editing systems, e.g., [31–33], allow users to select a segment of video. Some prior observational research had investigated users selecting spans while engaged in a video-editing task [34], and our study has extended span-selection interaction to the ASL education and video-search context. Other prior work had investigated educational software tools for students to select a segment of a spoken-language lecture videos while taking notes [10, 42] or integrated approaches to editing, sharing, and controlling spoken-language educational lecture videos [10, 57]. While there are differences between the task of understanding an ASL video and understanding an educational lecture video in spoken language, our findings on the benefits of span-selection interfaces for constraining the playhead and serving as a basis for integrated search tools may be relevant to that domain.

6 LIMITATIONS AND FUTURE WORK

Video stimuli in study 2 varied in frame-rate, bit-depth, compression, and frame scan (i.e., interlaced vs. progressive); these factors can affect video comprehension [29]. Since the videos were consistent in both prototype conditions in study 2, these factors did not affect our results, but future research could investigate the experience of ASL students viewing videos that vary in these dimensions. While our research has focused on ASL learners and videos, future work could be extended to learners of other sign languages.

While the Wizard-of-Oz dictionary-search output in our prototype simulated a single level of sign-recognition output accuracy, future research could examine how variations in the output quality would affect users' experience to inform computer vision researchers on the level of accuracy required. Analogous research has been conducted on isolated-sign search-by-video systems [3, 24, 25].

While our work revealed benefits of integrated, span-based dictionary-search during ASL video comprehension, future work could further explore the design space of span selection or presentation of search results. Further, while study 2 revealed benefits in translation quality and lower workload scores, it did not investigate whether engaging in the task of watching a challenging video led to measurable learning effects among students; future work could examine potential short- and long-term benefits of such activity.

Finally, our studies have focused on ASL learners, but the video-playing and span-searching tool could be useful for other user groups, e.g., linguists annotating videos of ASL, or experienced ASL interpreters translating complex technical videos. Future research could investigate the design of tools for these related tasks.

7 CONCLUSION

Students trying to understand a challenging video is a key part of comprehension-skill development among ASL learners. However, there are limitations with existing tools for looking-up signs. Our interview study with ASL learners revealed users' firsthand perspective on challenges they face during comprehension of such videos and their existing workarounds. These findings motivated and suggested the design of a Wizard-of-Oz video-player prototype with an integrated dictionary-search, based on span selection. Our study revealed differences between this prototype and a baseline (use of an existing ASL dictionary website) revealed benefits for student's accuracy of translation of ASL videos and subjective rating of workload. An analysis of how users interacted with the integrated system also revealed differences in use based on the genre and linguistic complexity of the video, benefits of the integrated design, and a dual use of span selection both as input to dictionary search and for constraining the video playhead.

Our work motivates research and tools to address the needs of ASL learners engaged in comprehension of challenging videos, and our findings inform future designers of ASL systems, computer vision researchers working on sign-matching technologies, and sign-language educators or linguists. These findings may also inform the design of video comprehension tools for other contexts.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. 1763569, 2212303, and 2125362.

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] Chanchal Agrawal and Roshan L Peiris. 2021. I see what you're saying: A literature review of eye tracking research in communication of Deaf or Hard of Hearing Users. In *The 23rd International ACM SIGACCESS Conference on Computers and Accessibility* (Virtual Event, USA) *(ASSETS '21)*. Association for Computing Machinery, New York, NY, USA, Article 41, 13 pages. https://doi.org/10.1145/3441852.3471209

[3] 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

[4] Stavroula Sokoli Athens and Stavroula Sokoli. 2007. Stavroula Sokoli (Athens) Learning via Subtitling (LvS) : A tool for the creation of foreign language learning activities based on film subtitling. In *MuTra 2006 – Audiovisual Translation Scenarios: Conference Proceedings*. MuTra, Copenhagen, Denmark, 8 pages.

[5] 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.

[6] 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

[7] 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

[8] Naomi K. Caselli, Zed Sevcikova Sehyr, Ariel M. Cohen-Goldberg, and Karen Emmorey. 2017. ASL-LEX: A lexical database of American Sign Language. *Behavior Research Methods* 49, 2 (01 Apr 2017), 784–801. https://doi.org/10.3758/s13428-016-0742-0

[9] Sheila Castilho, Stephen Doherty, Federico Gaspari, and Joss Moorkens. 2018. *Approaches to human and machine translation quality assessment.* Springer International Publishing, Cham, 9–38. https://doi.org/10.1007/978-3-319-91241-7_2

[10] Konstantinos Chorianopoulos and Michail N. Giannakos. 2013. Usability design for video lectures. In *Proceedings of the 11th European Conference on Interactive TV and Video* (Como, Italy) *(EuroITV '13)*. Association for Computing Machinery, New York, NY, USA, 163–164. https://doi.org/10.1145/2465958.2465982

[11] 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

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

[13] 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.

[14] Karen Emmorey, Robin Thompson, and Rachael Colvin. 2009. Eye gaze during comprehension of American Sign Language by native and beginning signers. *Journal of deaf studies and deaf education* 14, 2 (2009), 237–243.

[15] 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

[16] Susan M Gass, Jennifer Behney, and Luke Plonsky. 2020. *Second language acquisition: An introductory course* (5 ed.). Routledge, New York. 774 pages.

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

[18] Debbie B Golos and Annie M Moses. 2011. How teacher mediation during video viewing facilitates literacy behaviors. *Sign Language Studies* 12, 1 (2011), 98–118.

[19] Michael Andrew Grosvald. 2009. *Long-distance coarticulation: A production and perception study of English and American Sign Language.* University of California, Davis, 1 Shields Ave, Davis, CA 95616.

[20] 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.

[21] Sandra G Hart. 2006. NASA-task Load Index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 50. Sage Publications Sage CA, Sage publications, Los Angeles, CA, 904–908.

[22] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, Amsterdam, Netherlands, 139–183.

[23] 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* (Regensburg, Germany) *(CHIIR '22)*. Association for Computing Machinery, New York, NY, USA, 383–386. https://doi.org/10.1145/3498366.3505804

[24] 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.

[25] 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

[26] 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* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 195, 13 pages. https://doi.org/10.1145/3491102.3501986

[27] Saad Hassan, Aiza Hasib, Suleman Shahid, Sana Asif, and Arsalan Khan. 2019. Kahaniyan - Designing for Acquisition of Urdu as a Second Language. In *Human-Computer Interaction – INTERACT 2019*, David Lamas, Fernando Loizides, Lennart Nacke, Helen Petrie, Marco Winckler, and Panayiotis Zaphiris (Eds.). Springer International Publishing, Cham, 207–216.

[28] 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.

[29] Simon Hooper, Charles Miller, Susan Rose, and George Veletsianos. 2007. The effects of digital video quality on learner comprehension in an American Sign Language assessment environment. *Sign Language Studies* 8, 1 (2007), 42–58.

[30] iMotions A/S. 2019. *iMotions Biometric Research Platform.* imotions. https://imotions.com/academy/

[31] Adobe Inc. 2008. Adobe Premiere Pro. https://www.adobe.com/products/premiere.html. [Online; accessed 03-March-2022].

[32] Apple Inc. 2008. Apple Finalcut. http://aiweb.techfak.uni-bielefeld.de/content/bworld-robot-control-software/. [Online; accessed 03-March-2022].

[33] Apple Inc. 2008. Apple iMovie. https://www.apple.com/imovie/. [Online; accessed 03-March-2022].

[34] Tero Jokela, Minna Karukka, and Kaj Mäkelä. 2007. Mobile Video Editor: Design and Evaluation. In *Proceedings of the 12th International Conference on Human-Computer Interaction: Interaction Platforms and Techniques* (Beijing, China) *(HCI'07)*. Springer-Verlag, Berlin, Heidelberg, 344–353.

[35] Jonathan Keane, Diane Brentari, and Jason Riggle. 2012. Coarticulation in ASL fingerspelling.

[36] Annette Klosa-Kückelhaus and Frank Michaelis. 2022. The Design of Internet Dictionaries. *The Bloomsbury Handbook of Lexicography* 1 (2022), 405.

[37] 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.

[38] Marlon Kuntze, Debbie Golos, and Charlotte Enns. 2014. Rethinking literacy: Broadening opportunities for visual learners. *Sign Language Studies* 14, 2 (2014), 203–224.

[39] The language archive. 2018. ELAN - The Max Planck Institute for Psycholinguistics. https://archive.mpi.nl/tla/elan

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

[41] Scott K Liddell and Robert E Johnson. 1986. American Sign Language compound formation processes, lexicalization, and phonological remnants. *Natural Language & Linguistic Theory* 4, 4 (1986), 445–513.

[42] Ching (Jean) Liu, Chi-Lan Yang, Joseph Jay Williams, and Hao-Chuan Wang. 2019. NoteStruct: Scaffolding Note-Taking While Learning from Online Videos. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3290607.3312878

[43] Carolyn McCaskill, Ceil Lucas, Robert Bayley, and Joseph Christopher Hill. 2011. *The hidden treasure of Black ASL: Its history and structure.* Gallaudet University Press Washington, DC, Gallaudet University Press, 800 Florida Avenue, NE, Washington, DC 20002-3695.

[44] John Milton and Vivying S. Y. Cheng. 2010. A toolkit to assist L2 learners become independent writers. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Linguistics and Writing: Writing Processes and Authoring Aids* (Los Angeles, California) *(CL&amp;W '10)*. Association for Computational Linguistics, USA, 33–41.

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

[46] 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

[47] Anshul Mittal, Pradeep Kumar, Partha Pratim Roy, Raman Balasubramanian, and Bidyut B Chaudhuri. 2019. A modified LSTM model for continuous sign language recognition using leap motion. *IEEE Sensors Journal* 19, 16 (2019), 7056–7063.

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

[49] Tobii Pro Nano. 2014. *Tobii Pro Lab.* Tobii Technology. https://www.tobiipro.com/

[50] 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

[51] 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

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

[53] 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

[54] Jérémie Segouat. 2009. A study of sign language coarticulation. *SIGACCESS Accessible Computing* 1, 93 (Jan 2009), 31–38. https://doi.org/10.1145/1531930.1531935

[55] Zatorre RJ Shiell MM, Champoux F. 2014. Enhancement of visual motion detection thresholds in early Deaf people. *PloS one* 9, 2 (2014), e90498. https://doi.org/10.1371/journal.pone.0090498

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

[57] Namrata Srivastava, Sadia Nawaz, Joshua Newn, Jason Lodge, Eduardo Velloso, Sarah M. Erfani, Dragan Gasevic, and James Bailey. 2021. Are You with Me? Measurement of Learners' Video-Watching Attention with Eye Tracking. In *LAK21: 11th International Learning Analytics and Knowledge Conference* (Irvine, CA, USA) *(LAK21)*. Association for Computing Machinery, New York, NY, USA, 88–98. https://doi.org/10.1145/3448139.3448148

[58] Ted Supalla. 1982. *Structure and acquisition of verbs of motion and location in American Sign Language*. Ph.D. Dissertation. University of California, San Diego.

[59] 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.

[60] Carolina Tannenbaum-Baruchi and Paula Feder-Bubis. 2018. New sign language new (S): the globalization of sign language in the smartphone era. *Disability & society* 33, 2 (2018), 309–312.

[61] 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

[62] 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

[63] 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

[64] Mikhail A. Zagot and Vladimir V. Vozdvizhensky. 2014. Translating Video: Obstacles and challenges. *Procedia - Social and Behavioral Sciences* 154 (2014), 268–271. https://doi.org/10.1016/j.sbspro.2014.10.149