ASL Sea Battle: Gamifying Sign Language Data Collection

Danielle Bragg
Microsoft Research
Cambridge, MA, USA

Naomi Caselli
Boston University
Boston, MA, USA

John W. Gallagher
Microsoft Research
Cambridge, MA, USA

Miriam Goldberg
Boston University
Boston, MA, USA

Courtney Oka
Microsoft Research
Cambridge, MA, USA

William Thies
Microsoft Research
Bangalore, India

ABSTRACT

The development of accurate machine learning models for sign languages like American Sign Language (ASL) has the potential to break down communication barriers for deaf signers. However, to date, no such models have been robust enough for real-world use. The primary barrier to enabling real-world applications is the lack of appropriate training data. Existing training sets suffer from several shortcomings: small size, limited signer diversity, lack of real-world settings, and missing or inaccurate labels. In this work, we present ASL Sea Battle, a sign language game designed to collect datasets that overcome these barriers, while also providing fun and education to users. We conduct a user study to explore the data quality that the game collects, and the user experience of playing the game. Our results suggest that ASL Sea Battle can reliably collect and label real-world sign language videos, and provides fun and education at the expense of data throughput.

CCS CONCEPTS

• **Human-centered computing** → **Collaborative and social computing systems and tools**; **Accessibility systems and tools**; • **Computing methodologies** → **Machine learning**.

KEYWORDS

sign language; ASL; crowdsourcing; game; data; machine learning

ACM Reference Format:
Danielle Bragg, Naomi Caselli, John W. Gallagher, Miriam Goldberg, Courtney Oka, and William Thies. 2021. ASL Sea Battle: Gamifying Sign Language Data Collection. In *CHI Conference on Human Factors in Computing Systems (CHI '21), May 8–13, 2021, Yokohama, Japan.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3411764.3445416

1 INTRODUCTION

The development of accurate sign language models has the potential to improve access for millions of signers. Many people use sign languages worldwide, including about 70 million deaf people who use a sign language as their primary language [30], as well as sign language students, and friends and family of deaf signers. There are many applications that accurate machine learning models could enable for sign language users, including personal digital assistants that respond to signed commands, automatic sign language transcription services, and automatic translation between a signed and spoken language.

However, no existing sign language recognition or translation systems are robust enough for real-world adoption. The primary obstacle is lack of sufficient real-world training data. Existing datasets primarily consist of videos recorded in-lab, with a small set of homogeneous signers. They are also small in size (e.g., state-of-the-art datasets consisting of about 100,000 signs compared to 5 million spoken or 1 billion written words [1]). Many datasets also suffer from lacking or inaccurate labels. As current sign language recognition software cannot provide accurate labels, labeling is a high-skilled, labor-intensive task. As a result, recognition and translation models trained on these datasets do not work well in real-world settings, which requires generalization to diverse settings and diverse signers.

At the same time, there are few sign language-based entertainment and education resources. Existing digital resources are almost exclusively built for people who use a spoken or written language. For example, informational resources including search engines, encyclopedias, dictionaries, and written publications all almost exclusively support interacting with the material through written queries and read text. Similarly, entertainment and social resources, such as online games or forums typically involve written or spoken language. As sign languages are not spoken, and do not have a standard adopted writing system, these resources generally fail to include sign language users. In rare cases, a sign language interpreter may be hired so that these resources may be accompanied by signed interpretations. Creating more sign language resources is important both for providing access to people whose primary language is a sign language, but also for supporting people who are learning one.

In this work, we present ASL Sea Battle, a sign language smartphone game prototype designed to collect and label real-world sign language videos. The game serves two primary purposes: 1) creating real-world sign language corpora, while also 2) providing a fun and educational resource for sign language users. The game is based on the traditional game of battleship, where two opponents hide ships on two grids, and take turns guessing cells where ships might be hidden. Instead of referencing a cell by row and column number, each cell in ASL Sea Battle is labelled with a sign. A player attacks a cell by recording a video of him/herself executing the desired label. By recording videos on their smartphones in their

CHI '21, May 8–13, 2021, Yokohama, Japan

ACM ISBN 978-1-4503-8096-6/21/05…$15.00
https://doi.org/10.1145/3411764.3445416

daily lives, players provide real-world videos of themselves signing. Their video is sent to their opponent, who unlocks the cell by tapping it to match it with the incoming video, which seamlessly provides a label for the opponent's video. In this way, ASL Sea Battle not only generates videos of people signing, *but also labels them*.

To evaluate the feasibility of using ASL Sea Battle to collect and label real-world datasets, we conducted a user study comparing collection through the game to collection through a traditional, straightforward collection app. We analyzed the quality of the data and labels collected, and feedback from participants on their experience. Our results suggest that ASL Sea Battle can be used to sustainably collect high-quality real-world sign language videos, and provide accurate labels. It can also provide players with entertainment, education, and social connections. However, these benefits come at the cost of slower data throughput rates, as the game incurs delays between turns, for example as players strategize.

2 BACKGROUND AND RELATED WORK

In this section, we provide background on the required qualities of sign language datasets used for training machine learning models, the limitations of existing datasets, and the landscape of existing sign language resources and games with a purpose. To this space, we add the first game designed to collect and label sign language videos, while simultaneously providing the benefits of fun and connection to the signing community.

2.1 Sign Language Datasets

In order to successfully train models for the desired application(s), sign language datasets need to have several properties. There should be many examples of each sign, and signs should be labelled in a machine-readable, consistent format. This is a challenge, as there are no widely-used conventions for notating signs. As a result, lemmatizing–determining whether two sign productions are examples of the same sign or different signs–is not trivial. The training data also needs to match the videos that will be supplied in the final application. This means having a diversity of signers: different people, of different skin tones and body types, who use different regional and sociolinguistic varieties of the sign language, and have different skill-levels in the sign language. The data would also need to be diverse with respect to the filming conditions: including a range of camera angles, lighting conditions, backgrounds. If the end application is a phone app, the dataset will need to include videos of people signing while holding a phone, which can significantly change how signs are produced. Signs produced in isolation are likely insufficient for training models of continuous signing and vice-versa, as sign production in context is markedly different from production of single signs.

While there are datasets of sign languages that meet some of these criteria (e.g., there are large quantities of signing videos that are unlabelled, and there are sets of labelled videos of people signing that are not diverse and/or are relatively small), there remains a dearth of large-scale video datasets that meet enough of these criteria to adequately develop real-world sign language technologies [1]. For example, state-of-the-art sign language datasets include corpora of small numbers of homogeneous sign language interpreters (e.g., [14, 15]), other corpora with small sets of signers (i.e. < 15) (e.g., [29, 46]), and poorly labelled videos of unverified quality scraped from online sources (e.g., [24]). In this work, we explore the possibility of gamifying sign language video collection and labelling, to help address the lack of appropriate real-world data.

2.2 Sign Language Resources

Currently, existing digital resources for ASL and other signed languages are extremely limited, and digital sign language games are virtually non-existent. Existing resources are comprised of a small number of dictionaries (e.g., [5, 16]), educational materials (e.g., [10, 12, 21, 43]), lexical databases ([6, 20]), and mobile vocabulary apps (e.g., [27, 28]). However, the landscape of existing digital sign language resources pales in comparison to the rich landscape of resources for spoken/written language users, who are typically considered by default. Existing research on how well existing apps meet DHH needs also suggests large room for improvement [33].

Tools and resources that involve machine learning or experimental interfaces are almost always research projects with limited real-world viability. For example, attempts have been made to create browser tools that provide signed translations of written content [23], to create signing avatars ([13, 37, 37]), and to more generally creation recognition and translation systems (e.g., [4, 9, 9, 31, 45]). Research projects have also tackled educational aspects, for example proposing a learning game for DHH children and families [8], an app for learning fingerspelling [35], and data-collecting flashcards to help students learn vocabulary and identify sign features [3]. Our game adds a resource for both fun and education to this landscape. The sparse landscape also provides an opportunity for our game to achieve higher adoption and thereby collect a larger corpus, due to reduced competition with other apps and games.

2.3 Games with a Purpose

Games with a purpose have been proposed in other domains, some with great success. For example, researchers have proposed games to crowdsource protein folding [11], to label audio [26, 36] and images (including for accessibility purposes) [38, 39, 41], and to collect common-sense knowledge [40] and user preferences [18]. Several of these games have achieved millions of users, demonstrating the viability of gamification as a technique for scaling human contributions. Accessibility of gaming interfaces has also been explored (e.g., [17, 44]). Most existing games that curate data focus either on collection or labelling, and our game provides both. It is also the first game with a purpose serving sign language users.

More generally, "organic crowdsourcing" [25] refers to genres of data-collection methods that provide non-monetary benefits to users. In addition to gamification (described above), other examples include citizen science [22, 34], where people are incentivized to contribute by a desire to help advance science, and systems that incentivize contributions by providing people with personalized insights [32]. Crowdsourcing provides the opportunity to break tasks down into small "microtasks", which have been shown to take longer to complete, but can result in higher quality results by enabling people to make use of smaller units of available time [7]. ASL Sea Battle is designed as an asynchronous mobile game, to

enable people to take advantage of small units of available time to contribute and have fun.

3 ASL SEA BATTLE

In this section, we present ASL Sea Battle, a mobile app designed to gamify sign language data collection, including both sign videos *and* their labels.

3.1 Design Process

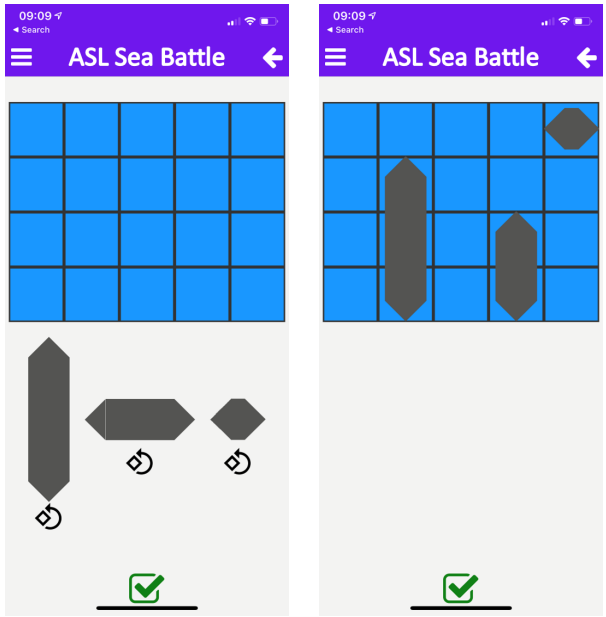
To reach our final design, we first identified criteria a data-collecting sign language game must meet in order to generate datasets that can be used to train real-world AI/ML models. We then engaged in an iterative design process where we prototyped different games and tried them out. We included deaf signers from the beginning of the process, both through iterative testing and refinements and as part of our research team. Our final design was the only design we found that met our criteria, and was fun to play.

To enable training AI/ML models that are viable in real-world settings, we sought real-world videos and accompanying labels.

- Videos: We seek videos of people using ASL, to enable training of sign recognition models. Further, we seek real-world videos, in order to enable training models that will work well for people in real-world settings. This means videos that are collected in real-world settings (including homes, offices, vehicles, and outdoors settings) and by diverse signers (including gender, ethnicity, and geography). This also means collecting videos recorded on devices where sign recognition technology may be deployed (i.e., laptops and smartphones).
- Labels: It is not enough to collect videos of people signing; in order to train sign recognition models, those videos must also be labelled (annotated with their contents). Sign language video annotation typically requires specialized skills, is very time-consuming, and expensive if done after video collection. Hence, collecting labels in the game is highly desirable.

To meet these criteria, we experimented with a set of designs. We started with existing mainstream games and tailored them to ASL, because mainstream games have already been vetted in terms of appeal and adoption potential. Specifically, we designed, created, and played prototypes of ASL versions of: Scattergories (where people think of signs that start with specific handshapes of a given topic, rather than words that start with certain letters of a given topic), hangman (where people guessed the set of features that compose the sign to get hints, rather than letters that spell a word), Flappy Bird (where players make a bird jump up by executing a sign, rather than pressing the keyboard), and battleship (where players guess tiles by executing signs, rather than specifying row and column number).

Except for battleship, each game presented prohibitive barriers to meeting our design criteria: Scattergories provided topic labels rather than exact labels; hangman primarily provided feature labels, had extremely low throughput for videos; and controlling Flappy Bird through signing resulted in distorted fast signing and did not allow for the level of timing precision that the game requires. In contrast, battleship allowed us to collect both videos *and* exact labels without hindering gameplay.



(a) Ships ready for placement. **(b) Ships placed on board.**

Figure 1: Screenshots of placing ships on the board in ASL Sea Battle.

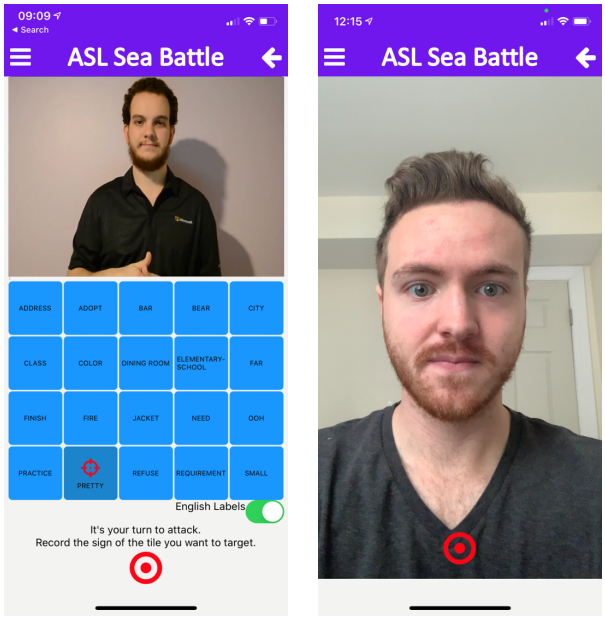
3.2 ASL Sea Battle Game Design

Battleship is one of the most popular board games of all time; fans claim that over 100 million copies have been sold [42]. During this classic game, each of two players hides a set of "battleships" in a grid. Then, two players take turns guessing where their opponent has hidden their battleships by guessing one square at a time. The player who guesses where all of their opponent's ships are hidden first wins.

ASL Sea Battle is inspired by this traditional game, but incorporates sign language in guessing squares and revealing whether the guess is a hit or miss. Specifically, to execute a guess, the player records a video of themselves executing a sign that matches the label on the square they would like to guess. The opponent then views the recorded video, and taps on the corresponding square. At that point, the game reveals what is under the square that has been guessed either part of a battleship, or nothing. The game ends when one player has found all of the other player's battleships.

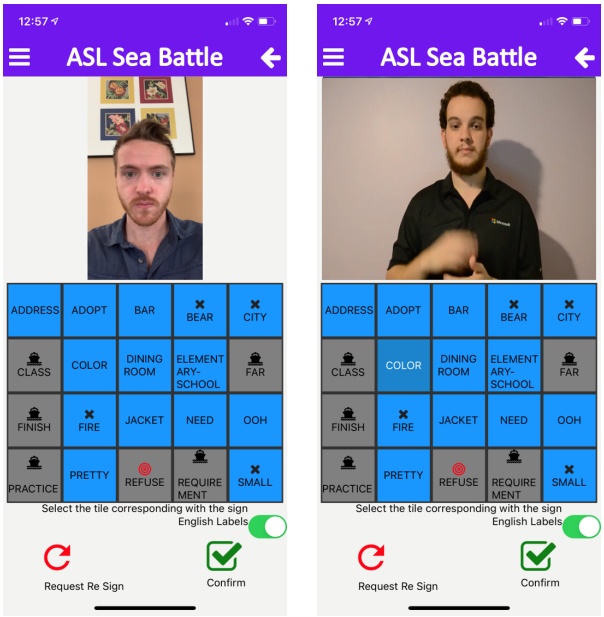
3.2.1 Game Board. The ASL Sea Battle game board is a grid that, to fit comfortably on a smartphone screen, is 5x4 in dimension. Each player has their own board where they hide a fleet of ships. The fleet consists of three ships, all one tile wide, and one, two, or three tiles long. This set of ships was chosen to fit on the small game board, while making game play challenging but not impossible (they cover 6/20 squares, so initially each player has about a 1/3 chance of hitting a ship). During placement at the beginning of the game, the ships may be rotated, and placed anywhere on the grid, so long as they do not overlap and do not extend past the end of the board (shown in Figure 1).

Each tile is represented by an ASL sign, which enables the embedding of signing into the game play. The sign that corresponds



(a) Targeting a square. (b) Recording a sign video.

Figure 2: Screenshots of the recording process in ASL Sea Battle. After recording, the player can view their recording and decide whether to submit it or re-record (not shown).



(a) Matching opponent's video. (b) Match selected.

Figure 3: Screenshots of labeling process in ASL Sea Battle. A redo can also be requested of the opponent (not shown).

to a tile is represented in two ways: with an English label, and with an expert video of the sign. English labels on the board tiles may be toggled on or off by users. At any time, users may tap tiles to view a video of the sign with which that tile is labeled.

Each player has a view of their own board (which contains their own fleet placement), and a view of the opponent's board (which shows which squares have been attacked, and whether they were hits or misses). A player's own board consists of blue squares (representing water), and grey squares with black ship icons (representing intact ship tiles), and grey squares with a fire icon (representing a hit ship tile). The view of the opponent's board similarly contains blue squares (represented water covering unknown contents), grey squares with a fire symbol (representing a hit ship), and blue squares with an X over them (representing a miss entering the water).

3.2.2 Recording Videos. The game generates sign language videos when players attack squares on their opponents' boards. To target a tile on an opponent's board, a player can tap on a particular square. Once tapped, a model video of the corresponding sign appears, and a target symbol appears over the square (as shown in Figure 2a). This model video helps disambiguate between signs and may also have educational benefits for students learning the language.

Once the player has identified a square they wish to attack, the player can attack the square by recording a video of themself executing sign associated with the square. The player records themself in the app directly and can view their recording before submitting it. The option to re-record is available to players who are unsatisfied with their recordings.

3.2.3 Labeling Videos. The game generates labels for player videos when the opponent views an attack video, and unlocks the corresponding square on the grid. After a player records an attack video of a particular sign, their video is sent to their opponent. The opponent then matches the player's video against squares of the grid. Taping on a square selects the square, and displays the sign that represents the square, executed by an expert signer. When the opponent finds a match, they tap "Confirm." This selection completes a turn. It also unlocks the square that has been attacked, so that the attacker's game board updates so they can view whether their guess was a hit or a miss. Alternatively, if the recording is not labelable (e.g., was mistakenly recorded and does not contain signing, or contains a sign with no match), the player can request a new recording from the opponent by selecting "Request Re-Sign".

3.3 Implementation

The ASL Sea Battle app prototype was built with React Native and deployed via the respective Android and iOS app stores. User-submitted videos are recorded locally on-device using the standard Android or iOS Camera APIs. Game state between users is synchronized by updating and polling a REST API written in Typescript using the Express framework for NodeJS and hosted in Azure Container Instances. Server-side persistence of user and game data is achieved through a MongoDB instance managed by CosmosDB in a cloud provider.Certificates and keys to enable HTTPS for the API were stored in the cloud and automatically fetched by the server.

4 USER STUDY

To explore the feasibility of gamifying sign language data collection (both the quality of data collected, and the user experience),

we ran a remote user study, with IRB approval. During the study, participants used two apps: ASL Sea Battle, and a control app for video collection and labelling. Afterwards, they answered answered several questions about their experience.

4.1 Participants

We recruited participants by contacting relevant email lists, and through snowball sampling. In particular, we reached out to ASL class lists and local Deaf community members. All participants had to have access to a smartphone (to access the apps, since the study was run remotely), and be age 18 years and above. Deaf participants were fluent ASL users, and hearing participants had to have some level of experience with ASL (at least introductory ASL classes).

In total, we recruited 20 participants: 10 deaf or hard-of-hearing (DHH), and 10 hearing. Basic demographics for our participants are:

Table 1: Participant demographics. ASL Fluency was measured on a scale from 1 (I do not use ASL) to 7 (I am fluent).

	DHH	Hearing
Gender	8 Female, 2 Male	9 Female, 1 Male
Age	24-70 (mean 33.5)	20-33 (mean 24.1)
ASL Fluency	5-7 (mean 6.5)	2-7 (mean 4.7)

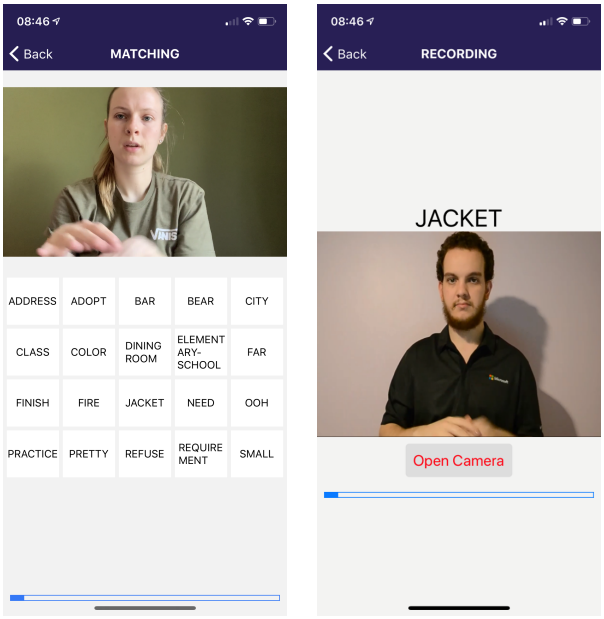
4.2 Procedure

The study was conducted remotely, due to the COVID-19 pandemic making in-person user studies infeasible. During the study, a researcher connected with the participant via a video call to walk the participant through the procedures, observe usage of the apps, and answer any questions. A researcher fluent in ASL connected with DHH ASL users, and a researcher fluent in English connected with hearing English speakers.

The study consisted of three main activities, after consent: 1) playing ASL Sea Battle, 2) using a control app for recording and labelling sign videos, and 3) answering questions about their experience with the apps and their own demographics. The order in which participants experienced the two apps was pseudo-randomized. We counterbalanced for which app was used first, so that a pseudo-random half of DHH participants and a pseudo-random half of hearing participants used each app first. After using both apps, participants used an online form to answer several questions about their experience and enter basic demographics (age, gender, ASL experience, etc.).

To provide a consistent experience across apps, we used the same set of signs (described below) across the two apps. All participants played against the same opponent (a researcher who is Deaf and a fluent signer), who also recorded the videos that participants labelled in the control app. Having the same person provide videos that are labelled in both apps allows for a more systematic comparison of the labels. To minimize time spent searching for an identified sign, we also alphabetized the set of signs on both the ASL Sea Battle board and control app labeling task.

The study took about one hour in total. After completing the study, each participant received an online gift card as compensation.



(a) Labeling **(b) Recording (watching model)**

Figure 4: Screenshots of the control app.

4.3 Control App

As a baseline comparison, we created a control data collection app. In this control app, there are two basic functionalities, with no gamification. Participants simply 1) view model signs and record themselves signing their own version of the sign, and 2) view other people's recordings and label (identify) which sign was executed. Figure 4a shows the labeling interface of the control app, and Figure 4b shows the recording interface of the control app.

To provide a comparable labeling experience across the two experiences, the control app provided the same labeling interface as the ASL Sea Battle app. Specifically, users are presented with a 5x4 grid of possible matches, ordered alphabetically.

4.4 Sign Videos

We used the same set of 20 signs for both apps and for all participants. In order to include signs that even newer ASL learners might know, we selected signs randomly from an instructional text widely used in ASL classes (Signing Naturally Units 1-5). Signs were chosen to be varied in form (both one-handed and two-handed signs, and a range of handshapes, movements, and non-manual markers). The 20 signs were (glossed, in alphabetical order): ADDRESS, ADOPT, BAR, BEAR, CITY, CLASS, COLOR, DINING ROOM, ELEMENTARY-SCHOOL, FAR, FINISH, FIRE, JACKET, NEED, OOH, PRACTICE, PRETTY, REFUSE, REQUIREMENT, and SMALL.

The model videos of these 20 signs in both apps were taken from ASL-LEX [6], with permission. These model videos are viewable in ASL Sea Battle when a participant selects a square prior to recording an attack, when tapping on a square while waiting for the participant to attack (e.g. to review vocabulary), and when tapping on a square in the grid to unlock the square in response to an attack. The model videos are viewable in the control app as

prompts for recording, and also appear when a video is labelled by tapping on a matching square (initially labelled just with English gloss).

In addition to these model videos, the control app used a set of videos pre-recorded by one of the researchers (who is a fluent Deaf signer) for the labelling task in the control app. We chose not to pre-record the opponent videos used in ASL Battlehips, so that participants had the experience of playing with a real live opponent.

5 RESULTS

We analyzed the results of our user study to shed light on 1) the quality of data collected through the apps (both recordings and labels), and 2) the user experience of using the apps for data generation. We analyze DHH and hearing signers separately, due to different baseline levels of language proficiency across groups. Our results suggest that ASL Sea Battle can be used to collect and accurately label real-world sign language video datasets; and that gamification increases willingness to contribute, but decreases speed.

5.1 Recordings

In total, we collected 657 videos through the user study. 400 were collected through the control app (20 per person), and 257 were collected through ASL Sea Battle (fewer than 20 per person, depending on game length). The 400 control app videos are evenly distributed among the 20 signs, since we asked each participant to record each sign exactly once. The 257 battleship app videos are not evenly distributed among the 20 signs, since selection depends on gameplay. The exact breakdown in battleship videos for DHH and hearing participants is provided in Table 5 in the Appendix. The 6 videos without a label are those that players decided to re-record, and thus were never sent to the opponent for labeling. The mean video length was 2.60 seconds (std dev .63 seconds). This consistently short video length aligns with the content that participants were recording (individual signs).

5.1.1 Expert Evaluation Procedure. To understand the quality of the recordings that participants provided, two Deaf linguistics experts viewed and evaluated every participant-recorded video. We made a simple website to facilitate this analysis, which displayed each video along with a set of questions. The exact questions and result tallies for each question are provided in Table 7 in the Appendix. If the answer to question 1 (whether the video contained a recognizable sign) was "Yes," the site proceeded to questions 3-7 (evaluating the recording quality); otherwise, it proceeded to question 2 (identifying the video contents). In addition to these seven questions, the experts could optionally input notes.

5.1.2 Quality for Training Recognition Models. To analyze the fraction of videos that would be useful for training a sign language model (i.e. contained the desired sign or a rough approximation, without major errors), we looked at videos that: 1) contain a single recognizable sign (Q1 answer "Yes"), which is 2) either the same sign or basically the same sign as the target (Q3 answer "It is the same" or "It looks a little different, but is basically the same sign"). All of the other "mistakes" the experts checked for (i.e., repetitive signing, one-handed sign execution, errors in execution, or failure to capture the full signing space) are representative of real-world data that an AI/ML system would need to be able to handle, and so we want to include that variability in the training set. Consequently, we do not filter out these mistakes.

Table 2: Percent of recordings where at least one expert's evaluations indicate the video is appropriate for training real-world recognition models.

	ASL Sea Battle	control app	Total
DHH	96.85%	98.00%	97.55%
Hearing	96.77%	99.50%	98.46%
Total	96.81%	98.75%	98.0%

Of the 651 total videos (657 minus the 6 re-records), 638 (98%) met these criteria for usefulness by at least one expert's evaluation, and 609 (93.55%) met these criteria by both experts. The breakdown in percentages of videos recorded that met the criteria by at least one expert is shown in Table 2.

The percent of useful videos was consistently high (above 96%) for each app and group. The difference in video usefulness was not statistically significant between apps ($\chi^2(1, N = 651) = 2.974, p = .085$) or groups ($\chi^2(1, N = 651) = 0.679, p = .410$). There were only 30 videos that were deemed useful by only one expert's evaluation, and the difference between expert evaluations in such cases were largely subjective. For example, 8 of these videos were judged as basically the same sign by one expert, and to be a different sign with the same/similar meaning by the other. Such disagreements highlight the ambiguity of defining boundaries between signs (unlike the relative clarity in defining boundaries between words in spoken languages).

Only 13 videos failed to meet our criteria by both experts, and the experts had consensus on their exact evaluations. Of these videos, 9 were considered by both to be a different sign with the same/similar meaning, and 2 were considered by both not to be a video containing a single recognizable sign, and 1 was considered to contain a different sign with similar meaning by one, and to contain other content by the other. Of these 13 videos, 8 were recorded in ASL Sea Battle (4 DHH, 4 hearing), and 5 were recorded in the control app (4 DHH, 1 hearing). This accounts for a total of 3.12% of battleship videos, and 1.25% of the control app videos. The predominance of DHH people recording these videos aligns with some DHH participants' comments that their signs for certain topics differed from the model videos.

5.1.3 Re-recorded Videos. In ASL Sea Battle, we gave players the option to request a video redo from their opponent if they could not identify a corresponding board tile. While the game collected these recordings, they are not labeled because the would-be labeler requested a new video instead of providing a label (resulting in the 'no label' row of Table 5). In total, there were six requests to re-record a video (one of a hearing participant, five of one DHH participant). Of these redos, one was due to no sign being recorded, two were due to the sign recorded having the same meaning as the English label, but being a different sign than was shown on the board; and three were due to technical difficulties (e.g., networking issues caused the video to appear frozen in one case).

In both apps, we also gave participants the option to view and re-record their own videos if they were not satisfied with the quality. Because these videos were never collected, and because the study was conducted remotely, we were not able to identify the contents of the videos. However, we were able to track how often this occurred. In total, there were 76 self-initiated re-records: 32 in ASL Sea Battle (4 DHH, 28 hearing), and 44 in the control app (6 DHH, 38 hearing). Most of the re-records were initiated by hearing participants (87%). It is possible that this disparity is due to differences in experience/comfort recording oneself signing. ASL is a central part of Deaf cultures in the U.S. (but not hearing cultures), so it is likely that DHH participants had engaged with ASL more frequently by recording and viewing videos.

5.2 Labels

In total, our study participants provided labels for 400 videos through the control app, and 259 videos through ASL Sea Battle. As with the recordings, participants provided labels for a uniform distribution of the 20 signs through the control app, as each participant labeled exactly one video of each sign. For the battleship game, the exact signs they labelled depended on gameplay, and are summarized in Table 6.

5.2.1 Evaluation Procedure. To analyze the label quality provided in both apps, we compared the true label for each video to the participant's label of the sign. For the control app, we knew the true label of each video that participants recorded, since these videos were pre-recorded by one of our research team members. For the battleship app, we built a simple website for the researcher who played the game to input the true label (i.e. the sign they intended to sign) for each video they recorded in the game. We then compared the true labels to the participants' labels, in both apps.

5.2.2 Quality for Training Recognition Models. In both ASL Sea Battle and the control app, 100% of the labels provided by participants were correct. This result applies to all 659 labels collected through both apps, and to the labels provided by both hearing and DHH participants, with varied levels of ASL.

This high accuracy suggests that a crowd of signers (even non-fluent students) may in some cases be able to serve as accurate labellers or provide quality control checks for sign language videos. It is worth noting that the vocabulary level of the signs used in our study was relatively low, and it is possible that accuracy would drop for more advanced vocabulary, especially if labelled by non-fluent signers. It is also worth noting that the set of signs from which participants picked was quite limited (20 signs), and fairly dissimilar. The most similar pair of signs was JACKET and ADDRESS. Both are two-handed, symmetric signs. This limited set of signs from which to pick may actually be a strength of our design, by limiting the cognitive load of labeling and constraining opportunities for mistakes.

5.3 Data Throughput

To assess the efficiency of collecting and labeling a corpus of videos through ASL Sea Battle, we also compared throughput, or the time it took participants to record and label videos, in each app (shown in Figure 5). The time to record a video in ASL Sea Battle is calculated as the total time spent on the attacking portion of a participant's turn. Recording time in the control app is calculated as the time spent on each recording task. Similarly, labeling time in ASL Sea Battle is calculated as the total time spent on the labeling portion of a participant's turn. Overall, to create a recording took a mean of 17.57 s, median 14.49 s, with standard deviation of 12.52 s.

To evaluate the significance of the differences between apps and populations, we again used the 2-way ANOVA. The results are shown in Tables 3 and 4. The differences in throughput between apps is strongly statistically significant, in terms of both recording time and labeling time ($p<.001$). The difference between groups (DHH and hearing) is not statistically significant, in terms of recording or labeling time. These results suggest that ASL Sea Battle provides data at a slower rate than traditional, straight-forward collection mechanisms, but that both populations groups can use the game with comparable efficiency.

Table 3: Recording throughput (time to record videos) - statistical significance of 2-way ANOVA.

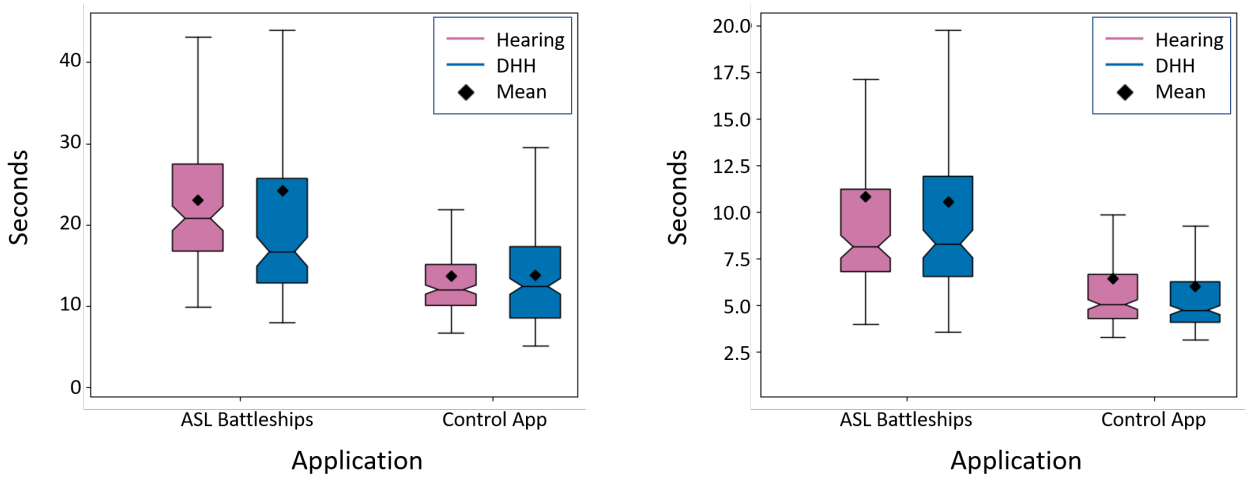
Source	SS	df	MS	F	p	
Hearing	43.14	1	43.14	0.320	.572	
App	14837	1	14837	110.2	<.001	***
Hearing*App	44.64	1	44.64	0.332	.565	

Table 4: Labeling throughput (time to label videos) - statistical significance of 2-way ANOVA.

Source	SS	df	MS	F	p	
Hearing	24.79	1	24.79	0.72	.396	
App	3132	1	3132	91.04	<.001	***
Hearing*App	0.58	1	0.58	0.02	.897	

5.4 Participant Preferences

The vast majority of participants (95%) found ASL Sea Battle more enjoyable than the control app, primarily due to increased fun and interactivity. While playing ASL Sea Battle, many participants commented on their enjoyment, for example commenting "This is fun!", or making happy exclamations when they hit a ship. Participants did not make any such comments or exclamations while using the control app. Specifically, in response to the question "Out of the apps you just used during the study, which was more fun?", 19 participants replied that ASL Sea Battle was more fun, while 1 (hearing) participant replied that the control app was more fun. Participants were asked to explain their preferences. Those who preferred ASL Sea Battle cited a variety of reasons for this preference, including: finding the game more intellectually stimulating and engaging (9), enjoying playing games (6), enjoying competition (6), and enjoying real-time interaction with another person (5). In contrast, the participant who preferred the control app explained that they liked having control over the pace, and found it better for studying vocabulary. One participant summarized this difference between the two apps: "I would choose battleship because it felt like it was geared toward entertaining, whereas the [control app]



(a) Recording **(b) Labeling**

Figure 5: Comparison of data throughput of ASL Sea Battle and the control app, for DHH and hearing participants, in terms of both a) recording time and b) labeling time. The boxplots show the median line within the interquartile range with confidence intervals for those medians as notches and with whiskers extending to the maximum and minimum values within the 95th percentile range.

felt a bit more like a vocab practice session." With this in mind, it is perhaps unsurprising that all the fluent signers (and in particular the DHH participants) preferred ASL Sea Battle, as they had less need for vocabulary practice. This preference is perhaps even more striking, given that participants lost most of their games (80% hearing games, 50% DHH games).

Most participants (85%) also preferred using ASL Sea Battle overall. In response to the question "Out of the apps you used during the study, which did you prefer overall?", 17 participants preferred ASL Sea Battle while 3 participants (1 DDH, 2 hearing including the one who found the control app more fun) preferred the control app. Participants were again asked to explain their preferences. Participants who preferred ASL Sea Battle cited similar reasons, including: fun (7), enjoyment of gameplay (5), interactive connection with others (5), educational benefits (3), and competition (2). All participants who preferred the control app cited timing issues (one stating a preference for control while practicing vocabulary, one stating a dislike of delays between turns in the game, and one preferring that the control app was quicker overall).

5.5 Willingness to Contribute Videos

To assess participants' willingness to contribute videos collected through the apps towards a corpus, we asked: "It is possible that the videos you created in the apps could help improve sign language technologies. Would you be willing to contribute your videos from either app to this cause? (Select all that apply.)" All (20 or 100%) participants were willing to contribute videos from at least one app: the vast majority (17 or 85%) from both apps, two DHH participants only from ASL Sea Battle, and one hearing participant only from the control app. When asked "Why are you willing or not willing to share your videos?", responses differed somewhat between DHH and hearing participants. The primary response from DHH participants (5 or 50%) was that they were willing to contribute their videos in order to provide diversity, while no hearing participants cited this reason. Many even explained the type of diversity they provided, including different types of ASL accents and levels of fluency. In contrast, the primary response from hearing participants was a willingness to help technical advancements (6 or 60%), which was also cited by some DHH participants (2 or 20%). Another major point of commonality across both DHH and hearing participants was a desire to support development more ASL games and educational resources (2 or 20% DHH, 4 or 40% hearing).

To gauge long-term interest, we also asked, "Would you like to be notified when either app is publicly released, so you can use it? (Select all that apply.)" One hearing participant did not want to be notified of a public release of either app. The remaining participants (19 or 95%) all wanted to be notified of a public release of ASL Sea Battle, and most (16 or 85%, all except three DHH participants) also wanted to be notified of a public release of the control app. It is likely that only DHH participants selectively opted out of the control app release, due to differences in benefit. As described above, participants found educational value in both apps, but found entertainment value primarily in ASL Sea Battle. As our DHH participants were more fluent in ASL, this educational benefit may not have been sufficient to warrant use for these participants. One DHH participant summarized the difference in motivating appeal between the two apps: "[ASL Sea Battle] is more fun. The [control app] you'd have to pay people to use."

6 DISCUSSION

In this section, we discuss the higher-level value of gamifying sign language data collection (notably motivation for contributing and

immediate language resources), as well as limitations and opportunities for future work (which are vast, as this work presents only an initial attempt at gamifying sign language data collection).

6.1 Value of ASL Games and Resources

It is possible that gamification will be key to engaging the deaf community in creating scalable sign language datasets. In our user study, our DHH participants strongly preferred the game to the control app (with all but one DHH participant preferring the game), and also had high ASL language fluency. This difference in preference may be explained by the fact that both apps offered vocabulary practice (unlikely to appeal to high-fluency DHH users), but only ASL Sea Battle offered entertainment and social benefits (which can appeal to high-fluency DHH users). More generally, traditional data-collection mechanisms typically involve vocabulary or content repetition (as in our control app), which has educational value, and so may may appeal to novice hearing signers, but not to fluent DHH signers. In the absence of intrinsic appeal, it is likely that data curators would need to pay fluent DHH signers to contribute, which limits scalability. One participant reinforced this notion, commenting 'you'd have to pay me to do this again'. It is possible that gamification could provide an alternative hook for this key demographic.

In addition to generating corpora, games like ASL Sea Battle may help fill a more direct need for ASL resources, which participants highlighted as a problem. When asked "Do you play any ASL games on your phone or computer?", only two hearing participants answered "yes", citing Quizlet flashcards and practice games through ASL dictionary websites. These two participants noted deficiencies in the ASL resource landscape, explaining that there are few options, and the existing options tend to be inaccurate or boring. One participant described the landscape: "Through the few searches I've done for ASL games (whether for practice or pure entertainment), I have never come across a good array of options. It seems as though the existing ASL games are very limited, or not particularly helpful/accurate." The Quizlet user further added: "It's a fast way to learn new signs but is kind of boring." Participants' willingness to contribute videos to help develop more ASL resources (described above) further suggests that they see a need for such resources. Though our study focused on ASL, the game design is generalizable to other signed languages, and may provide similar benefits for other signing communities worldwide.

A number of participants viewed ASL Sea Battle (and to a lesser extent the control app) as valuable games and resources to fill this void. For ASL Sea Battle, participants envisioned educational, social, and entertainment use cases. As most deaf children are born into hearing families who do not know sign language at the time of birth, deaf children are at risk of limited access to both spoken and signed language, which can be extremely problematic [19]. As such, parents and other family members are often eager to learn ASL. As one DHH participant explained, ASL Sea Battle could serve as a group learning tool "beneficial for families that want to learn ASL for their kids/sibling/ or family member". Other participants noted the value of the human connection the game provides, for example stating, "I want to play ASL Sea Battle with my friends", and "It was fun! And a way to connect with someone else from a distance." It is possible that the current pandemic's social distancing environment highlighted the benefit of human-to-human digital connection. Another DHH participant simply noted that it was fun to interact in a game in ASL rather than English, explaining, "It is fun to pick a spot by using sign language rather than saying 'A1'." Participants also noted educational use cases for the control app, which "felt... more like a vocab practice session". Overall, participants noted educational benefit of both apps, and additional social and entertainment benefits of ASL Sea Battle. As one participant summarized: "I think I would choose battleship as my favorite overall, because the game added an extra layer of fun on top of its educational benefit."

6.2 Limitations and Future Work

Some participants noted that their personal execution of signs or concepts differed from that of the model signer in the apps. In particular, some DHH participants thought other signs better matched the English glosses, and wanted instead to use their own signs. For example one participant commented, "I have a different sign for OOH. Do I have to copy [the model]?" In some cases the signers wanted to know how closely their sign had to match the model. For example, one person wanted to know if it was okay that, "[The model] signed ROOM and I did it opposite" (they had used a slightly different movement). Despite such questions, our recording quality analysis suggests that participants still largely submitted recordings that matched the model. Nonetheless, for future such games, it may be beneficial to clarify for users what types of variations are acceptable, or to use machine learning methods to cluster different signs submitted for the same concept or English word. Alternatively, it may be worth creating different modes for fluent and non-fluent players.

In our user study, we only explored collection of a small set of signs, with relatively basic vocabulary. It is possible that more advanced signs would have resulted in less accurate recordings or labels, in particular from less fluent contributors. It is also possible that sets of signs that are more visually similar would result in lower labeling accuracy. However, the design of the game actually enables collectors to prevent such confusions by providing labelers with a small set of labels from which to choose, and giving collectors the capability to intentionally create boards that include a highly distinct set of signs. It is also possible that collectors could position signs for which they need the most data in squares on the board that are statistically attacked most frequently. Future work includes exploring the collection of larger corpora with expanded vocabulary.

We also focused on collection and labeling of individual signs, as opposed to continuous signing. While systems that model individual signs have utility (e.g., dictionaries that support looking up concepts by demonstrating them, or digital assistants that recognize simple commands), models of continuous signing are required for many applications (e.g., end-to-end translation). Modifying the ASL Sea Battle design to collect and label continuous signing, and designing new games that would enable such collection, makes for rich future work.

Building and evaluating AI/ML models trained on real-world data collected by ASL Sea Battle and other sign language games

is another open area for future work. Collecting sufficient data to train models would require long-term deployments, and addressing privacy concerns in sharing videos of oneself signing may require video modifications in order to secure participation [2]. It is also possible that training successful models would require development of new modeling methods. State-of-the-art sign language modeling research is primarily conducted with clean, high-quality data collected in laboratory settings. It is possible that the resulting models will not apply well to the type of real-world data collected by such games.

7 CONCLUSION

The development of sign language machine learning systems has the potential to break down communication barriers for millions of signers worldwide. However, it has not been possible to train models that perform sufficiently well for real-world deployment. The primary barrier is a lack of labeled real-world training data. In this work, we proposed a mechanism for gamifying the collection and labeling of such a real-world dataset, through a game called ASL Sea Battle. We conducted a user study to explore the quality of data and labels that the game can collect, and to better understand players' experiences and willingness to use the game to contribute data. Our results suggest that the game can collect high-quality recordings and labels from a wide range of contributors. They also suggest that by providing benefits of enjoyment and social interactions, gamification may provide a new scalable mechanism for collecting and labeling real-world sign language datasets.

ACKNOWLEDGMENTS

We thank Abraham Glasser for insightful discussions, and for serving as a model signer for this paper. This material is based upon work supported by the National Science Foundation under Grant No. (NSF BCS-1918252 and NSF BCS-1749384).

REFERENCES

[1] Danielle Bragg, Oscar Koller, Mary Bellard, Larwan Berke, Patrick Boudreault, Annelies Braffort, Naomi Caselli, Matt Huenerfauth, Hernisa Kacorri, Tessa Verhoef, et al. 2019. Sign Language Recognition, Generation, and Translation: An Interdisciplinary Perspective. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility*. 16–31.

[2] Danielle Bragg, Oscar Koller, Naomi Caselli, and William Thies. 2020. Exploring Collection of Sign Language Datasets: Privacy, Participation, and Model Performance. In *The 22nd International ACM SIGACCESS Conference on Computers and Accessibility*. 1–14.

[3] 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 & Social Computing*. 1837–1848.

[4] Jan Bungeroth and Hermann Ney. 2004. Statistical sign language translation. In *Workshop on representation and processing of sign languages, LREC*, Vol. 4. Citeseer, 105–108.

[5] B Cartwright. 2017. Signing Savvy. *Access mode: https://www.signingsavvy.com* (2017).

[6] 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 (2017), 784–801.

[7] Justin Cheng, Jaime Teevan, Shamsi T Iqbal, and Michael S Bernstein. 2015. Break it down: A comparison of macro-and microtasks. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 4061–4064.

[8] Ching-Hua Chuan and Caroline Anne Guardino. 2016. Designing smartsignplay: An interactive and intelligent american sign language app for children who are deaf or hard of hearing and their families. In *Companion publication of the 21st international conference on intelligent user interfaces*. 45–48.

[9] Necati Cihan Camgoz, Simon Hadfield, Oscar Koller, Hermann Ney, and Richard Bowden. 2018. Neural sign language translation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 7784–7793.

[10] ASL Clear. 2020. https://aslclear.org/

[11] Seth Cooper, Firas Khatib, Adrien Treuille, Janos Barbero, Jeehyung Lee, Michael Beenen, Andrew Leaver-Fay, David Baker, Zoran Popović, et al. 2010. Predicting protein structures with a multiplayer online game. *Nature* 466, 7307 (2010), 756–760.

[12] ASL Core. 2020. https://aslcore.org/

[13] Ralph Elliott, John RW Glauert, JR Kennaway, Ian Marshall, and Eva Safar. 2008. Linguistic modelling and language-processing technologies for Avatar-based sign language presentation. *Universal Access in the Information Society* 6, 4 (2008), 375–391.

[14] Jens Forster, Christoph Schmidt, Thomas Hoyoux, Oscar Koller, Uwe Zelle, Justus H Piater, and Hermann Ney. 2012. RWTH-PHOENIX-Weather: A Large Vocabulary Sign Language Recognition and Translation Corpus.. In *LREC*, Vol. 9. 3785–3789.

[15] Jens Forster, Christoph Schmidt, Oscar Koller, Martin Bellgardt, and Hermann Ney. 2014. Extensions of the Sign Language Recognition and Translation Corpus RWTH-PHOENIX-Weather.. In *LREC*. 1911–1916.

[16] Ann Grafstein. 2002. HandSpeak: A Sign Language Dictionary Online. *Reference Reviews* (2002).

[17] Dimitris Grammenos, Anthony Savidis, and Constantine Stephanidis. 2009. Designing universally accessible games. *Computers in Entertainment (CIE)* 7, 1 (2009), 1–29.

[18] Severin Hacker and Luis Von Ahn. 2009. Matchin: eliciting user preferences with an online game. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1207–1216.

[19] Wyatte C Hall. 2017. What you don't know can hurt you: The risk of language deprivation by impairing sign language development in deaf children. *Maternal and child health journal* 21, 5 (2017), 961–965.

[20] Julie Hochgesang, Onno Crasborn, and Diane Lillo-Martin. 2020. ASL Signbank. https://aslsignbank.haskins.yale.edu/

[21] Leala Holcomb and Jonathan McMillan. 2020. Home. http://www.handsland.com/

[22] Alan Irwin. 1995. *Citizen science: A study of people, expertise and sustainable development*. Psychology Press.

[23] Søren Staal Jensen and Tina Øvad. 2016. Optimizing web-accessibility for deaf people and the hearing impaired utilizing a sign language dictionary embedded in a browser. *Cognition, Technology & Work* 18, 4 (2016), 717–731.

[24] Hamid Reza Vaezi Joze and Oscar Koller. 2018. Ms-asl: A large-scale data set and benchmark for understanding american sign language. *arXiv preprint arXiv:1812.01053* (2018).

[25] Steven Komarov and Krzysztof Z Gajos. 2014. Organic peer assessment. In *Proceedings of the CHI 2014 Learning Innovation at Scale workshop*.

[26] Edith LM Law, Luis Von Ahn, Roger B Dannenberg, and Mike Crawford. 2007. TagATune: A Game for Music and Sound Annotation.. In *ISMIR*, Vol. 3. 2.

[27] Colin Lualdi. 2020. SignSchool. https://www.signschool.com/

[28] Matt Malzkuhn, Melissa Malzkuhn, Tim Kettering, and Megan Malzkuhn. 2020. The ASL App. https://theaslapp.com/

[29] Aleix M Martínez, Ronnie B Wilbur, Robin Shay, and Avinash C Kak. 2002. Purdue RVL-SLLL ASL database for automatic recognition of American Sign Language. In *Proceedings. Fourth IEEE International Conference on Multimodal Interfaces*. IEEE, 167–172.

[30] World Federation of the Deaf. 2018. *Our Work*. http://wfdeaf.org/our-work/ Accessed 2019-03-26.

[31] Lionel Pigou, Sander Dieleman, Pieter-Jan Kindermans, and Benjamin Schrauwen. 2014. Sign language recognition using convolutional neural networks. In *European Conference on Computer Vision*. Springer, 572–578.

[32] Katharina Reinecke and Krzysztof Z Gajos. 2015. LabintheWild: Conducting large-scale online experiments with uncompensated samples. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*. 1364–1378.

[33] Ryan Lee Romero, Frederick Kates, Mark Hart, Amanda Ojeda, Itai Meirom, and Stephen Hardy. 2019. Quality of Deaf and Hard-of-Hearing Mobile Apps: Evaluation Using the Mobile App Rating Scale (MARS) With Additional Criteria From a Content Expert. *JMIR mHealth and uHealth* 7, 10 (2019), e14198.

[34] Jonathan Silvertown. 2009. A new dawn for citizen science. *Trends in ecology & evolution* 24, 9 (2009), 467–471.

[35] Jorge Andres Toro, John C McDonald, and Rosalee Wolfe. 2014. Fostering better deaf/hearing communication through a novel mobile app for fingerspelling. In *International Conference on Computers for Handicapped Persons*. Springer, 559–564.

[36] Douglas Turnbull, Ruoran Liu, Luke Barrington, and Gert RG Lanckriet. 2007. A Game-Based Approach for Collecting Semantic Annotations of Music.. In *ISMIR*, Vol. 7. 535–538.

[37] Margriet Verlinden, Corrie Tijsseling, and Han Frowein. 2001. A Signing Avatar on the WWW. In *International Gesture Workshop*. Springer, 169–172.

[38] Luis Von Ahn and Laura Dabbish. 2004. Labeling images with a computer game. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 319–326.

[39] Luis Von Ahn, Shiry Ginosar, Mihir Kedia, Ruoran Liu, and Manuel Blum. 2006. Improving accessibility of the web with a computer game. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*. 79–82.

[40] Luis Von Ahn, Mihir Kedia, and Manuel Blum. 2006. Verbosity: a game for collecting common-sense facts. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*. 75–78.

[41] Luis Von Ahn, Ruoran Liu, and Manuel Blum. 2006. Peekaboom: a game for locating objects in images. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*. 55–64.

[42] Battleship Wiki. 2020. https://battleship.fandom.com/wiki/Battleship_(game)

[43] Alicia Wooten and Barbara Spiecker. 2020. Atomic Hands. https://www.atomichands.com/

[44] Bei Yuan, Eelke Folmer, and Frederick C Harris. 2011. Game accessibility: a survey. *Universal Access in the information Society* 10, 1 (2011), 81–100.

[45] 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*. 279–286.

[46] Morteza Zahedi, Philippe Dreuw, David Rybach, Thomas Deselaers, and Hermann Ney. 2006. Continuous sign language recognition-approaches from speech recognition and available data resources. In *Second Workshop on the Representation and Processing of Sign Languages: Lexicographic Matters and Didactic Scenarios*. 21–24.

A APPENDIX

Table 5: Signs recorded through ASL Sea Battle.

Sign	DHH	Hearing
OOH	10	4
SMALL	9	4
PRACTICE	8	6
FIRE	8	6
FINISH	8	5
NEED	7	8
REFUSE	7	8
REQUIREMENT	7	6
ADOPT	7	5
BAR	6	9
COLOR	6	7
CITY	6	7
JACKET	6	6
DINING ROOM	6	5
FAR	6	4
ADDRESS	5	8
CLASS	5	7
ELEMENTARY SCHOOL	4	8
PRETTY	3	7
BEAR	3	4
(no label)	5	1
Total:	132	125

Table 6: Signs labelled through ASL Sea Battle.

Sign	DHH	Hearing
COLOR	9	6
PRETTY	9	6
CLASS	8	10
ADOPT	8	6
BAR	8	9
PRACTICE	8	4
FIRE	7	7
REFUSE	7	5
ELEMENTARY-SCHOOL	7	4
BEAR	7	6
DINING ROOM	6	4
NEED	6	8
REQUIREMENT	6	6
FAR	6	6
SMALL	6	7
ADDRESS	6	4
FINISH	5	6
OOH	4	10
JACKET	4	8
CITY	3	7
Total:	130	129

Table 7: Expert evaluations of the videos collected in our user study, separated by app and hearing status. For each video, two experts answered the listed questions, by selecting a single answer from the given answer choices. For each answer choice, the table provides the number of videos where both experts input that answer. The "disagreement" option indicates the number of videos where they did not agree for that question.

		ASL Sea Battle				Control App			
		Hearing		DHH		Hearing		DHH	
		%	#	%	#	%	#	%	#
1. Does the video contain a single recognizable sign (possibly repeated)?	Yes	92	114	100	127	95	190	98	196
	No	1	1	0	0	0	0	1	1
	Disagreement	7	9	0	0	5	10	2	3
2. What does the video contain?	Multiple distinct signs	0	0	0	0	0	0	0	0
	Unrecognizable signing	0	0	0	0	0	0	0	0
	No signing (e.g. scenery/body shot)	0	0	0	0	0	0	0	0
	Too low quality to tell	0	0	0	0	0	0	0	0
	Other (write-in)	0	0	0	0	0	0	0	0
	Disagreement	100	1	0	0	0	0	100	1
3. Does the sign match this one [video of model sign]?	It is the same.	76	87	80	102	82	156	91	178
	It looks a little different, but is basically the same sign.	4	5	2	3	6	11	2	4
	It has the same/similar meaning, but is a different sign.	2	2	3	4	1	1	1	2
	Disagreement	18	20	14	18	12	22	6	12
4. Was the sign recorded as a one-handed sign when it is typically two-handed?	Yes	5	6	0	0	0	0	0	0
	No	94	107	98	125	97	185	99	194
	Disagreement	1	1	2	2	3	5	1	2
5. Is the sign repeated unnecessarily?	Yes	1	1	0	0	2	3	0	0
	No	89	102	96	122	92	174	98	193
	Disagreement	10	11	4	5	7	13	2	3
6. Are there other errors in sign execution (wrong handshape, movement, or location)?	Yes	2	2	1	1	4	7	0	0
	No	84	96	92	117	84	159	97	190
	Disagreement	14	16	7	9	13	24	3	6
7. Is the full signing space captured in the video (hand(s) involved, torso, face)?	Yes	73	83	76	96	85	161	80	156
	No	3	3	1	1	1	2	0	0
	Disagreement	25	28	24	30	14	27	20	40