A Corpus for Verifying American Sign Language During Game Play by Deaf Children

Helene Brashear[1], Zahoor Zafrulla[1], Thad Starner[1],
Harley Hamilton[1], Peter Presti[2] and Seungyon Lee[1]

[1]Georgia Institute of Technology, College of Computing, GVU, Atlanta, Georgia, USA.
(brashear, zahoor, thad, harley.hamilton)@cc.gatech.edu
[2]Georgia Institute of Technology, Interactive Media Technology Center, Atlanta, Georgia, USA.
peter.presti@imtc.gatech.edu

Abstract

The CopyCat project was designed to develop an interactive educational adventure game to help deaf children acquire language skills. The main goals of the project are to improve the language and memory abilities of deaf signing children, advance basic research in computer-based sign language recognition, and design an efficient language interaction model in order to assist in the language learning of deaf children. The CopyCat project was begun as a collaboration between Georgia Tech and the Atlanta Area School for the Deaf in 2004 and has been collecting ASL (American Sign Language) data since Spring of 2005. Since then we have collected 5829 signed phrases from over 30 children. In this paper we describe the evolution of the CopyCat system design, data collection methodology, and resulting corpus, as well as challenges and successes throughout the process.

1. Introduction

It is important that children are exposed to sufficient language examples during early childhood to aid in the development of life long language skills. Language learning is dependent upon the availability of that language and the opportunities a child (Spencer and Lederberg, 1997) or an adult learner (Krashen, 1980) have for interacting with skilled users of the language. This "critical period" of language exposure is important for both spoken and signed languages (Mayberry and Eichen, 1991; Newport, 1990). Ninety percent of deaf children are born to hearing parents who may not know sign language or have low levels of proficiency with sign language (Gallaudet, 2001). Many of these deaf children of hearing parents remain significantly delayed in language development due to a lack of language exposure at home. For many of these children the first consistent exposure to quality language models will be when they enter school, which can results in lifelong difficulties with communication (Stinson and Foster, 2000).

CopyCat was designed to address these language learning issues by facilitating the development of both expressive language and working memory skills. While computer-child interaction cannot replace high quality adult-child interaction, it can be designed to integrate meaningful authentic communication in order to enhance expressive language and working memory skills that may facilitate the childs ability to make the most of opportunities for acquiring language in a natural way via human interaction.

2. Evolution of CopyCat System

Our ASL data is collected on-site at schools around the Atlanta area. Children play a computer game by wearing colored gloves and signing to characters within the game to accomplish game objectives such as rescuing kittens or defeating villains such as alligators and snakes. Data is collected via wireless accelerometers mounted on the wrists of the gloves and a single video camera. The sensor data is collated and time stamped by the game system and saved as our library for developing our recognition system and linguistic review.

The system has been built in three main design phases: each phase addresses game design, data collection, and the ASL recognition engine. Each iteration has been designed with the ultimate goal of moving towards a fully functional system with live recognition that provides productive feedback for students of varying skill levels.

Our corpus collection methods were designed to attempt to elicit live, natural signing from children as they interact with characters in the game. This approach has resulted in a data set that contains many language modeling challenges including disfluencies, pauses, dominant hand switching, and sign variations. Our research has focused on developing labeling schemes and training models to accurately reflect the children's signing. A prototype live recognizer developed from this corpus has been deployed for testing and has been shown to have a statistically significant educational effect on language learning as compared to a control condition.

2.1. The CopyCat Games

As part of the CopyCat project, several computer-assisted language learning games have been designed. Each game entails some sort of quest by the hero to collect items in order to remediate a problem. In each quest, the children interact with the hero via sign language to tell the hero warn them of a villain or identify where a hidden object is located. If the children know what to tell the hero regarding the guards location they can push a "talk button to turn the hero towards them so they can sign to him/her. They then push the "talk button again when they are finished signing. If the children are uncertain what to say they can click a "help button to see the tutor in the top left corner of the screen tell them what to say. The child may view the tutor repeatedly if (s)he so chooses (see Figure 1).

After the child talks to the hero, the child's signing is classified as correct or incorrect. If the childs utterance is in-

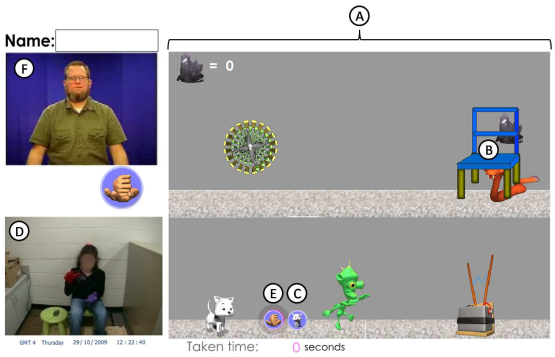

Figure 1: CopyCat screen shot from Mini Quests: A) Animated game characters in their worlds B) The villain (a snake) is hiding under the chair. C) The push to sign button has a picture of the kitten on it. Children push the button to sign to the kitten and warn her about the snake. D) The live video feed allows children to see themselves as they sign. E) The help button has a picture of the sign for help and will bring up ASL video to help the children during game play. F) This window is the help video feed.

correct, a question mark appears above the heros head, to simulate misunderstanding by the hero, and the child must try again to communicate accurately. If the childs sign is correct, the hero, with the wave of a paw, “poofs” the guard, turning it into an innocuous item and the hero continues on the quest

2.2. Language Learning

The video tutor examples in the game were designed to be similar to a communication setting which young children encounter while learning language through interaction with adults. As the childs linguistic and communicative competence and confidence grow, the need for such assistance diminishes and the child can respond appropriately without help. Thus, our tutor performs the role of the good adult language model (Schiefulbusch and Bricker, 1981), always available the child, responding to the childs cue (in this case a press of the “help” button) in an appropriate linguistic manner.

2.3. Educational Evaluation

In order to collect data regarding the language processing abilities of the children and the efficacy of the games language interaction model, pretests and post tests were administered and in-game response data were recorded. These tests consisted of sections to test receptive language skills, expressive language skills, and working memory. The results of the expressive language test indicate that the experimental group made a significant gain in the accuracy of their utterance to describe the video they saw as well as in their length of utterance as measured by mean length of utterance from pretest to post test (Weaver et al., 2010).

3. System Design

3.1. Iterative Design Cycle

The iterative design cycle allows us to adapt to problems as the emerge during the development process and has allowed the CopyCat system to improve rapidly.

3.2. Interface Design

Our user interface for game play uses a video stream of the user, feedback from characters in the game, and help videos in ASL to engage the children. The live video stream allows the children to see their signing and engages them in the signing. The children enjoy “being in the game” and tend to use the feedback to stay in frame.

The game characters have been designed to attempt to elicit natural signing. When the child pushes the signing button, the character will face the child and pay attention while (s)he is signing. If the signing is incorrect, a question mark thought bubble shows above the character. We have found that visual clues such as these help guide the children in their interactions.

The introduction instruction and game help videos are all ASL. We have taken care to synchronize the spatial layout of the game with the spatial constructs in signing to provide consistency. Even simple modifications to the interface such as moving a button require a check of all of the ASL spatial referencing in the videos.

3.3. Wizard of Oz

When the functionality of a system is under development, developers can sometimes replace that functionality with a person, similar to the “Great Wizard of Oz” operating behind the curtain. The system can be tested while the hidden “wizard” controls operations and developers can obtain critical feedback about system design early in the process (Dix et al., 2004).

We divided game development and sign language recognition by using a “Wizard of Oz” setup, shown in Figure 2 (Henderson et al., 2005). The child interacts with the user computer (on the right) by navigating with the mouse and signing to characters. The wizard’s computer (on the left) controls the game’s response to children signing and collects data from the sensors and game logs for future use.

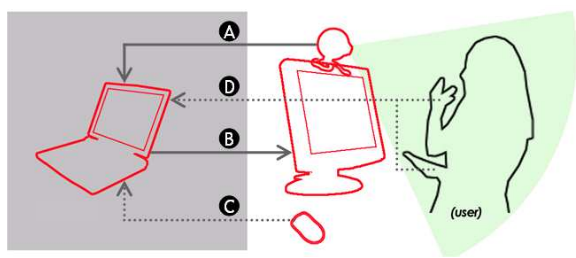

Figure 2: Diagram of the Wizard of Oz system system setup showing a) live camera and sensor feed b) interface output split between wizard and user c) child’s mouse and d) the interface computer

3.4. Sensors

The CopyCat system uses computer vision and three-axis accelerometers to collect data for use in sign language recognition. Our computer vision is processed from video collected on a single camcorder that faces the children. The children wear colored gloves, which contain small accelerometers mounted on the outside of the wrist (shown in Figure 3). These accelerometers provides information on movement acceleration, direction, and rotation of the hand. The distinct color of the gloves helps distinguish the hands from the skin color of the face and cluttered backgrounds. The wizard's computer coordinates the data streams, synchronizes them, and stores them for future use.

Figure 3: Gloves with accelerometer (top). Close up of wrist-mounted accelerometers (bottom).

One key design goal has been to have a portable system that will work in a variety of environments. Our deployment environments at the schools have ranged from classrooms and libraries, to a re-purposed supply closet. Figure 4 shows the construction of a "signing kiosk" and the resulting view from the camera. The kiosk is inexpensive and modular so that it can be transported easily. The kiosk fixes the position of the camera relative to the childs position on the chair. The color of the furniture can be used to help calibrate the video cameras color balance to enable better hand tracking. This kiosk design allows us to move our equipment from area to area with minimal re-calibration.

4. Resulting Corpus

4.1. Overview of data collected

Each phase of the CopyCat project includes on-site deployments to collect data at our two partner schools. We have collected a total of 5829 phrases over four phases, with a total of nine deployments. Table 1 shows a count of phrases collected throughout the CopyCat project. Each phrase is a three, four, or five sign sentence taken from a vocabulary of 22 signs. The phrases are of the format *[adjective1] subject preposition [adjective2] object* .

Figure 4: Kiosk setup

Subject	Object	Adjective	Verb
alligator	bed	black	behind
cat	box	blue	in
snake	chair	green	on
spider	flowers	orange	under
	wagon	white	
	wall		

Table 2: Game vocabulary

4.2. Characterizing the Children's Signing

Most of the sign language databases used for automatic sign recognition are carefully scripted and collected in a controlled environment (Holt et al., 2006). Our data set provides many samples of children signing as they interact with the online characters. This signing contains many of the artifacts of conversational signing such as difluencies like pauses, false starts, hesitations, and sign variations. It also has many examples of non-signing activities such as scratching and fidgeting.

The conversational nature of the children's interaction with characters results in signing samples that contain signing beyond basic game vocabulary. The data set contains many non-game communications towards game characters (including messages such as "wrong", "start again" or "not red I mean blue"), and even gestures that are not ASL such as a wave which is used generally to indicate an error and restart (a kind of "erase" gestures).

The children's signing handedness did not directly correspond to their dominant handedness for other activities and was inconsistent even within the phrases. This hand switching makes it more difficult to group signs and phrases by handedness for modeling purposes. Dominant hand switching is probably a symptom of their low fluency and is common among children (Mandal et al., 1999).

Phase	Game	Date	Participants	Ages	Total Phrases
Pilot	Kitten Escape!	Spring 2005	3	9-11	50
Pilot	Kitten Escape!	Spring 2005	2	9-11	78
				Total	**128**
First Deployment	Kitten Escape!	Spring 2005	5	9-11	627
First Deployment	Castle Quest	Fall 2005	9	9-11	1812
				Total	**2439**
Second Deployment	MiniQuests	Fall 2008	5	6-9	505
Second Deployment	MiniQuests	Spring 2009	5	6-9	503
Second Deployment	MiniQuests	Spring 2009	14	6-9	822
				Total	**1830**
Third Deployment	MiniQuests	Fall 2009	11	6-9	1432
				Total	**1432**
			CopyCat	**Total**	**5829**

Table 1: Table of data collected during the CopyCat project

4.3. Annotation

Our current annotation system is designed for sign classification, recognition, and verification. First we label each sign by its English label (green, cat, etc.). The initial label set has been expanded to included non-game vocabulary from children, as well as some non-ASL gestures such as pauses, fidgets, and waves. Signs are then annotated for handedness by hands used during the sign and the dominant hand: right hand, left hand, both+right hand dominant, both+left handed dominant, both+symmetric. Finally signs are rated for quality as good, ok, or bad.

5. Using the Data

5.1. ASL recognition

Our first task with the data set is automatic sign language recognition. In this process, we collect samples of signs, train up models using the samples, and then use the models for recognition. When the models are trained we use an independent test set for validation results. This means that we divide the data set into one group for training the models and another group for testing the models in order to see how well the models perform against signs examples that are previously unseen to the computer (Brashear et al., 2006).

5.2. ASL verification

Our second task is automatic sign language verification. In this process, we collect samples of signs, train up models using the samples, and then use the models to verify sign samples as a correct match or incorrect match to a baseline phrase. To get the verification we run data for a sample against the expected model and the use a common rejection threshold on the likelihood.

5.3. Tests of live system

In Fall of 2009 we conducted our first pilot tests of the live system. The verification system was based on models built with data from our first deployment. The results of that test are currently being compiled.

6. Challenges of the CopyCat Corpus

6.1. Library Continuity

There is a continued tension between goals for system improvement, expansion of game functionality, and library expansion. Though our upgrades in sensors and configuration have improved the reliability and portability of the system, they also detract from backwards compatibility. This discontinuity results in a larger corpora of children signing, with sub-sets from various deployments that are incompatible with each other.

The library data is stored in both its raw format as well a format that includes post-processing from vision and accelerometer sub-routines. This redundancy in storage requires more disk space, but helps alleviate the continuity problems by allowing for changes in post-processing without losing entire library sets. For example, we have changed our computer vision code several times. The raw data library allows us to experiment different post-processing schemes and choose optimally.

6.2. Sensor Changes

During the design cycle we have changed the sensors several times. Two of our main design priorities are system reliability and system portability. Our long term goal is a system that can be set up at any school and requires minimal maintenance. We started with the explicit goal that our sensors be inexpensive and easy for schools to use.

During the project, we have used both commercially available accelerometer and those we design in-house. We have gone through several iterations of accelerometer collection code in order to address issues that emerged with calibration, output normalization, and sensor drift (Westeyn et al., 2009). These changes, combined with changes to the video frame rate, create incompatibilities with existing data from previous deployments. This is a further challenge to library continuity and such changes must be carefully considered.

6.3. Varied Environments

Each time we visit a school for a deployment, we have no guarantees where they will have space for us to set up. These changes in environment create challenges for computer vision algorithms. Many sign language recognition systems depend on very static environments for their algorithms to work. We have worked to make the system more portable by a combination of choosing more flexible algorithms and creating an environment where visual cues can help keep the algorithms calibrated. The kiosk helps ensure that the camera distances are approximately the same each time. Additionally the colored gloves and furniture help provide reference points for algorithms to track hands, face, and body movement in the video frame.

6.4. Data Integrity

During data collection the system must coordinate data streams from three different sensors. These streams must be saved to disc, logged, and synchronized. One of the challenges of this configuration is keeping the data streams synchronized and providing live feedback for errors in reading, synchronizing, or logging sensor data. Our most recent iteration has focused on creating a subsystem specifically to provide feedback to administrators to prevent problems in game play and data loss.

Post-processing of the libraries can also discover errors in the data stream. These errors must be diagnosed for future prevention and the samples must be catalogued as damaged data.

6.5. Automatic Annotation

We have designed the game to provide as much automatic annotation as possible to help us index and use our data. Each signed phrase contains logs with information on user, session details, wizard feedback, and game information. After the data is stored, our post-processing is also largely automated. These logs provide further information about the content of the signed phrase. All of this data helps us rapidly compile statistics on the data set and pull out subsets by interesting features.

6.6. Maintaining library

As the library increases in size and complexity we have continued to try to address issues with maintaining our data. Maintaining logs and raw data allow us to continue to do retrospective evaluations of many aspects of the process. There are different research and publication cycles for the various topics of the CopyCat project: computer vision, machine learning, human-computer interaction, sign linguistics, and education.

The size of the data has been growing since the beginning of the project. Not only does each deployment add more data instances to the library, but the size of the data per instance has been growing as well. Verifying the integrity of automated process logs is tedious and is time consuming. We have increased sensor sampling rate, as well as the detail and complexity of the game logs. Additionally, we must keep track of data from educational testing which includes a large amount of video of the children's language testing sessions.

6.7. Sign Variation

The machine learning system needs many examples of the same signs across many systems for building representative models that are robust to variations. Thus far we have maintained a fairly small vocabulary, which allows for many examples of a sign. Even with the small vocabulary, we have discovered that there are often many variations on how a sign is performed. Most of these variations are technically correct and we must make allowance for them. If only one or two children perform a specific variation, it can make collecting sufficient examples difficult.

6.8. Developing Annotation Schemes

One of the goals of the machine learning research is develop generalized annotation schemes that will scale with larger data sets and vocabulary. Experimenting on this front can be very challenging since annotation schemes aren't standardized and the conversational nature of the children's signing creates unexpected variations in sign structure and performance. Annotating large sets of data can be time consuming and tedious. We have created an in-house annotation tool that acts like a video editor and can add multiple tags to the same sign sequence to indicate various labels such as the sign name, handedness, and quality. This tool allows for the addition of new tags as the annotation scheme evolves. Additionally, the collection of tags can be used to create different model groups for classification. For example a time sequence could be modeled as ”cat”, ”cat” with both hands, or ”cat” of good quality. We can test these variations in modeling to compare their performance.

6.9. Influencing the Children's Signing

Throughout the iterations of game design, we have continued to create an interface that influences how the children sign. The story line of the game helps restrict vocabulary by limiting the scope of objects and characters on the screen for the children to describe or address. By creating a conversational environment, we can influence how the children sign. The "click to sign" approach to the game provides a dual purpose of segmenting the signing sequences and giving the children pause to focus. We have even found that children will sometimes take a moment to rehearse their signing before clicking to get the character's attention in the game. These techniques have greatly improved the quality and kind of signing we get from the children, but we still face challenges with out of vocabulary signs and the children's difficulties performing the signs correctly.

6.10. Live Testing

As the machine learning research progresses, we will begin to conduct more live tests of the recognition system. We have recently augmented our system so that we can collect data while the live tests are being conducted. This multitasking allows us to continue to catalogue data while we test our machine learning system.

6.11. Privacy Issues

Because our data is collected from children our data is subject to strict privacy requirements. Our long term goal is to make sections of the data available to linguistic and machine learning researchers. Anonymizing the video data compromises the content, since the face is the center of the signing space and facial gestures are a component in ASL. We have been working with our institutional review board and the host schools to create an agreement that would allow us a mechanism to release data to other researchers.

7. Conclusion

CopyCat is a long-term project that has used an iterative development to design an interactive, educational game for deaf children. Designing and deploying the game for user testing has created unique challenges in collecting, storing, and using the large data set of children's signs. We have addressed many of these challenges with strategic game improvements generated from the feedback phase of the iterative cycle.

As CopyCat matures into a commercial-grade system, we are focusing on long-term library collection and management. The success of CopyCat will depend on our ability to easily integrate new data from each deployment into our library. We are focusing on ways to automate the collection and indexing of data for storage in a central library. As we build models off of the central library, each deployment site will get updates to the game recognition system.

8. Future Work

We are currently reviewing data collected from the most recent deployment as well as the results of the live system tests. Our long term goals include expanding the number of students and creating new games. We are working to expanding the vocabulary and language structure in new games. Additionally we will be performing more user testing on the live recognition system to determine its educational efficacy and to further examine the user experience when the Wizard is removed from the loop.

9. Acknowledgements

This work is supported by the NSF, Grants #0093291, #0511900, and the RERC on Mobile Wireless Technologies for Persons with Disabilities, which is funded by the NIDRR of the US Dept. of Education (DoE), Grant #H133E010804. The opinions and conclusions of this publication are those of the grantee and do not necessarily reflect those of the NSF or US DoE.

Special thanks to students and staff at the Atlanta Area School for the Deaf, Georgia School for the Deaf, and Gwinnett ISD for their generous help with this project.

10. References

Brashear, H., Park, K.-H., Lee, S., Henderson, V., Hamilton, H., and Starner, T. (2006). American Sign Language Recognition in Game Development for Deaf Children. In *Assets '06: Proceedings of the 8th International ACM SIGACCESS Conference on Computers and Accessibility*, pages 79–86, New York, NY, USA. ACM Press.

Dix, A., Finlay, J., Abowd, G., and Beale, R. (2004). *Human-Computer Interaction*, chapter 6.4 Iterative Design and Prototyping. Prentice Hall.

Gallaudet (2001). Gallaudet University. Regional and National Summary Report of Data from the 1999–2000 Annual Survey of Deaf and Hard of Hearing Children and Youth. Washington, D. C.

Henderson, V., Lee, S., Brashear, H., Hamilton, H., Starner, T., and Hamilton, S. (2005). Development of an American Sign Language Game for Deaf Children. In *IDC '05: Proceeding of the 2005 Conference on Interaction Design and Children*, New York, NY, USA. ACM Press.

Holt, G. T., Hendriks, P., and Andringa, T. (2006). Why Don't You See What I Mean? Prospects and Limitations of Current Automatic Sign Recognition Research. *Sign Language Studies*, 6(4).

Krashen, S. (1980). The Theoretical and Practical Relevance of Simple Codes in Second Language Acquisiton. In Scarcella, R. and Krashen, S., editors, *Research in Second Language Acquisition: Selected Papers of the Los Angeles Second Language Acquisition Research Forum. Issues In Second Language Research*. Newberry House, Rowley, MA.

Mandal, M. K., Asthana, H. S., Dwivedi, C. B., and Bryden, M. P. (1999). Hand Preference in the Deaf. In *Journal of Developmental and Physical Disabilities*, volume 11.

Mayberry, R. I. and Eichen, E. B. (1991). The Long-Lasting Advantage of Learning Sign Language in Childhood: Another Look at the Critical Period for Language Acquisition. *Journal of Memory and Language*, 30:486–498.

Newport, E. L. (1990). Maturational Constraints on Language Learning. *Cognitive Science*, 14:11–28.

Schiefulbusch, R. L. and Bricker, D. D. (1981). *Early Language: Acquisition and Intervention*. University Park Press, Baltimore.

Spencer, P. and Lederberg, A. (1997). Different Modes, Different Models: Communication and Language of Young Deaf Children and Their Mothers. In Romski, M., editor, *Communication and Language: Discoveries from Atypical Development*, pages 203–230. Harvard University Press.

Stinson, M. and Foster, S. (2000). Socialization of Deaf Children and Youth in School. In Spencer, P. and Marschark, M., editors, *The Deaf Child in the Family and at School: Essays in Honor of Kathryn P. Meadow-Orlans*, pages 151–174. Lea Lawrence Erlbaum Associates, London.

Weaver, K. A., Hamilton, H., Zafrulla, Z., Brashear, H., Starner, T., Presti, P., , and Bruckman, A. (2010). Improving the Language Ability of Deaf Signing Children through an Interactive American Sign Language-Based Video Game. In *Proceedings of 9th International Conference of the Learning Sciences*.

Westeyn, T., Presti, P., and Starner, T. (2009). A Naive Technique for Correcting Time-Series Data for Recognition Applications. In *Proceedings of the Thirteenth IEEE International Symposium on Wea rable Computers (ISWC 2009)*. IEEE Computer Society.