American Sign Language Recognition in Game Development for Deaf Children

Helene Brashear [1], Valerie Henderson [1]

Kwang-Hyun Park [2], Harley Hamilton [3],

Seungyon Lee [1], Thad Starner [1]

[1] Georgia Institute of Technology
GVU Center
College of Computing
Atlanta, Georgia, USA
(brashear, sylee, vlh, thad)@cc.gatech.edu

[2] Korea Advanced Institute of Science and Technology
Daejeon, Republic of Korea
akaii@robotian.net

[3] Center for Accessible Technology in Sign
Atlanta Area School for the Deaf
Clarkston, Georgia, USA
hhamilto@doe.k12.ga.us

ABSTRACT

CopyCat is an American Sign Language (ASL) game, which uses gesture recognition technology to help young deaf children practice ASL skills. We describe a brief history of the game, an overview of recent user studies, and the results of recent work on the problem of continuous, user–independent sign language recognition in classroom settings. Our database of signing samples was collected from user studies of deaf children playing a Wizard of Oz version of the game at the Atlanta Area School for the Deaf (AASD). Our data set is characterized by disfluencies inherent in continuous signing, varied user characteristics including clothing and skin tones, and illumination changes in the classroom. The dataset consisted of 541 phrase samples and 1,959 individual sign samples of five children signing game phrases from a 22 word vocabulary.

Our recognition approach uses color histogram adaptation for robust hand segmentation and tracking. The children wear small colored gloves with wireless accelerometers mounted on the back of their wrists. The hand shape information is combined with accelerometer data and used to train hidden Markov models for recognition. We evaluated our approach by using leave–one–out validation; this technique iterates through each child, training on data from four children and testing on the remaining child's data. We achieved average word accuracies per child ranging from 91.75% to 73.73% for the user–independent models.

Categories and Subject Descriptors

K.4.2 [**Social Issues**]: Assistive technologies for persons with disabilities; I.2.7 [**Natural Language Processing**]: Language models; I.5.4 [**Pattern Recognition**]: Implementation: interactive systems

General Terms

Human Factors, Languages

Keywords

Sign Language, ASL, Recognition, Game

1. INTRODUCTION

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 [5]. Unlike hearing children of English–speaking parents or deaf children of signing parents, these children often lack the access to language at home which is necessary for developing linguistic skills. Often these children's only exposure to language is from signing at school. Linguists have identified a "critical period" for language development - a period during which a child must be exposed to and immersed in a language. It is important that children are exposed to sufficient language examples during this period to aid in the development of life long language skills. Although originally thought to exist only for spoken languages, research has shown that this critical period also applies to ASL acquisition [14, 16].

Hearing children have a multitude of educational software products to enhance the language instruction they receive at school. This software is designed for use both at home and at school. A 1999 Kaiser Family Foundation report estimates that 17% of children ages 2-7 and 37% of children ages 8-13 play computer games on any given day [19]. Interactive ASL software is usually concentrates on students' ability to receive and comprehend language rather than on their ability to generate language independently. Two examples of this software are Con-SIGN-tration [10] and Aesop's Fables: Four Fables [18]. During Con-SIGN-tration, children play a memory game which involves matching cards bearing ASL signs to cards with English words. Aesop's Fables presents several of Aesop's Fables interpreted into sign and then the child is asked a series of comprehension questions in English following the stories. However, to our knowledge, no games currently on the market allow children to communicate with the computer via their native language of ASL. Games that do prompt children to mimic signs have no measure of evaluation to help the child improve the clarity and correctness of their signs. This lack of repetition with feedback prevents children from benefiting fully from the software.

1.1 Sign Language Recognition

Sign language recognition is a growing research area in the field of gesture recognition. Research on sign language recognition has been done around the world, using many sign languages, including American Sign Language [26, 22, 2], Korean Sign Language [11], Taiwanese Sign Language [13], Chinese Sign Language [4, 6], Japanese Sign Language [20], and German Sign Language [1]. Many sign language recognition systems use Hidden Markov Models (HMMs) for their abilities to train useful models from limited and potentially noisy sensor data [6, 22, 26].

Sensor choices vary from data gloves [13] and other tracker systems to computer vision techniques using a single camera[22], multiple cameras, and motion capture systems [25] to hand crafted sensor networks[8].

Starner et. al. demonstrated a continuous sign recognition system that performed at 98% accuracy with a 40 ASL sign vocabulary, in a lab environment using HMMs, using a simple grammar [22]. Further work was done to explore different sensor configurations and increase both the flexibility and mobility of the system [2, 8, 15]. These studies showed that the accuracy of ASL recognition can be increased by combining computer vision techniques with a small number of accelerometers.

2. THE SYSTEM

CopyCat is an educational computer game that utilizes computer gesture recognition technology to develop American Sign Language (ASL) skills in children ages 6-11. CopyCat's goal is to encourage signing in complete phrases and to augment a child's educational environment with a fun and engaging way to practice language skills. CopyCat consists of the hardware necessary for gesture recognition and the game software. The game is interactive – with tutorial videos demonstrating the correct signs, live video (providing input to the gesture recognition system and feedback to the child via the interface), and an animated character executing the child's instructions. The game focuses on the practice and correct repetition of ASL phrases and allows the child to communicate with the computer via ASL.

We are using an iterative design approach for the development of CopyCat [12, 7]. Iterative design is a cyclic process of design work, prototyping, testing and evaluation. This approach has allowed us to continually improve the game design and include the users throughout the entire design process. Throughout the testing process children are asked questions about their game experience. Each cycle results in a testing session at AASD, as well as a post-testing evaluation of the game and our collected data. The evaluation provided us with a checklist of strengths and weaknesses to bring to the next iteration.

Since no suitable previous ASL recognition engine exists for this project, a Wizard of Oz (WOz) approach is used. The Wizard of Oz (WOz) technique is an evaluation method which uses a human "wizard" to simulate the functionality that will eventually be provided by the computer. The Wizard is situated out of sight of the subject, receives the subject's input, and controls the system manually, emulating the missing functionality of the system [3]. During the WOz simulation, a participant performs tasks using the prototype. The subject is not aware of the Wizard's presence and believes that he is using a fully functioning system. In our scenario, while the child plays the game, a human Wizard simulates the computer recognizer and evaluates the correctness of the player's sign with the help of the interpreter.

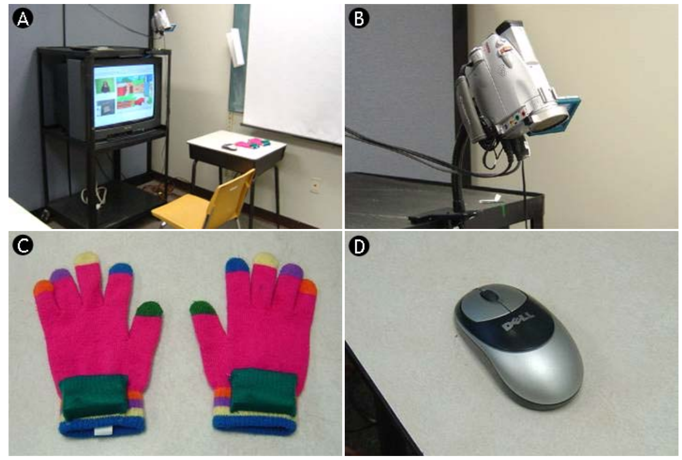

Figure 2: System setup. a) the workspace b) IEEE1394 camera c) colored gloves with wireless accelerometers in green-colored pockets d) wireless mouse

By using a Wizard of Oz technique we allow for iterative, parallel development tracks. Design work can continue in each of the research areas and be evaluated periodically through user testing. The parallel track spans several areas:

- **Game interface design** is a combination of Human Computer Interaction (HCI) design principles and iterative design practices. Research issues include game and character design, presenting ASL information in a game context, collecting data on ASL signing, responding appropriately to signs, and linguistic consistency in the game's spatial layout.
- **Language development** is both an educational and linguistic focus. Research issues include designing principled signed interactions in the game and evaluating student game performance in a meaningful and useful way, as well as longer term evaluation of the developmental impact of the system on the children's language skills.
- **ASL recognition development** focuses on machine learning. Research issues include modeling the children's signs, recognized the children's signs as correct or incorrect, identifying and modeling disfluencies in sign, and identifying and acting on non-scripted communication with the characters in the game.

2.1 System Design

The system's configuration is shown in Figure 2. The workspace area (shown in Figure 2A) consists of a computer monitor, a desk for the mouse, and a chair. Users

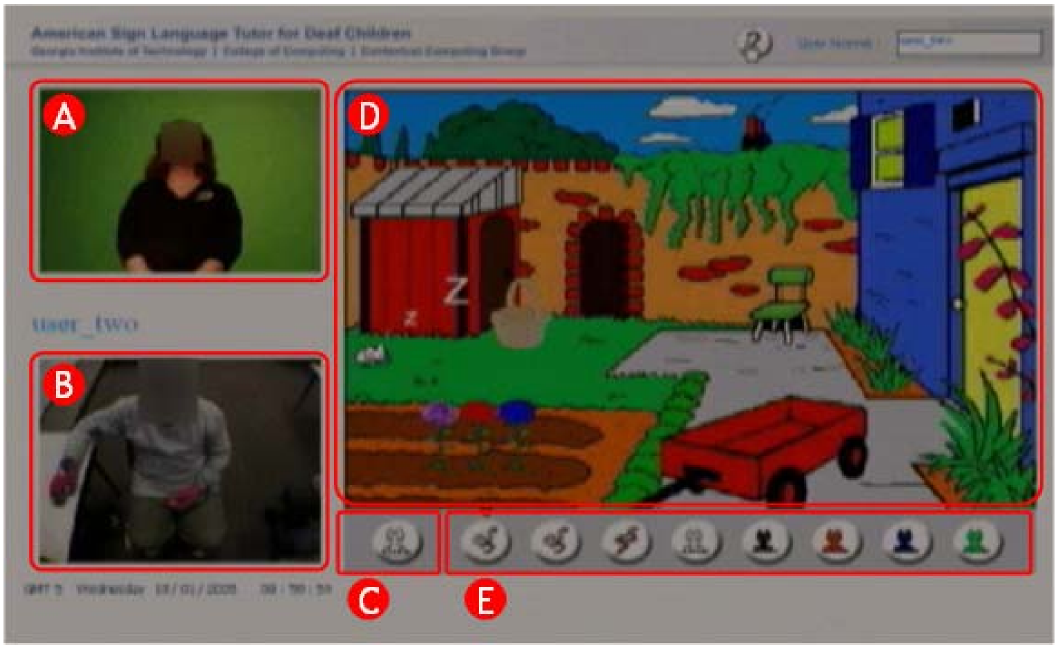

Figure 1: Screen shot of ASL game. a) tutor video b) live camera feed c) attention button d) animated character and environment e) action buttons

Figure 3: Wireless accelerometer board

can navigate parts of the game (such as requesting a tutorial video) using the mouse. Data from the children's signing is recorded using an IEEE 1394 video camera (shown in Figure 2B) and using wireless accelerometers (shown in Figure 3) mounted in colored gloves (shown in Figure 2C). The colored gloves help aid the computer vision algorithms used on the video data and hold the wireless accelerometers in a stable position on the wrist. The wrist–mounted accelerometers provide additional information which can aid the recognition task [2].

The sensor configuration was chosen to satisfy specific requirements. The game should be able to run in a variety of locations and lighting environments. The system should not be expensive and should be easy to configure in a school environment. Computer vision has been successfully used for sign recognition, and cameras are available, inexpensive and durable. Data gloves are instrumented gloves that measure flexion and movement. They are appealing to sign language recognition projects because of the quantity and detail of information, but they tend to be very expensive and are not commonly available in child's sizes. Additionally, they are generally not designed for the stresses of classroom use.

3. ASL RECOGNITION SYSTEM

3.1 Data Set

With the assistance of educational technology specialists and linguists, we developed a list of appropriate phrases which are assigned to actions in the game. Phrases were selected with the goal of having three and four signs, using age-appropriate vocabulary (English translations listed in Table 1). This structure was chosen as part of the goal of encouraging the linguistic transition from single sign utterances to complete phrases. A vocabulary was chosen that was consistent with what they used in their classes and compatible with system constraints. The recognition engine is currently limited to a subset of ASL which includes single and double handed signs, but does not include more complex linguistic constructions such as classifier manipulation, facial gestures, and level of emphasis. Each phrase is a description of an encounter for the game character, Iris the cat. The students can warn of predators, such as "go chase snake" or identify the location of a hidden kitten, such as "white kitten behind wagon".

We collected the data in nine days with five children ages 9-11 at the Atlanta Area School for the Deaf in Clarkston, Georgia. Collecting data for use in statistical pattern recognition is both time consuming and tedious because a large number of samples must be collected and then labeled. We use a Wizard of Oz configuration and a "push-to-sign" mechanism to collect and segment relevant data samples [12, 7].

During game play, the main character Iris is asleep. The child must click to wake Iris, sign the phrase, and click to end the interaction. We use "push-to-sign", which segments the samples by the start and stop clicks of the mouse during

Action Button	Level 1	Level 2	Level 3
#1	Go chase snake	Go chase snake	Go chase snake
#2	Go chase snake	Go chase spider	Go chase spider
#3	Go chase snake	Go chase spider	Go chase alligator
#4	White kitten behind wagon	White kitten under chair	White kitten in flowers
#5	Black kitten under chair	Black kitten in flowers	Black kitten in bedroom
#6	Orange kitten in flowers	Orange kitten in bedroom	Orange kitten on wall
#7	Blue kitten in bedroom	Blue kitten on wall	Blue kitten behind wagon
#8	Green kitten on wall	Green kitten behind wagon	Green kitten under chair

Table 1: Phrases used for ASL game (English translation)

game play. This push-to-sign mechanism is similar to those found in many speech recognition systems (called push-to-talk for speech) and allows our ASL recognition system to perform recognition on only pertinent phrases of the children's sign. The data can be automatically labeled at a phrase level using information from the game. We use this method to remove both out-of-context and unscripted signing, as well as ignore the child's out–of–game comments. Thus the segmentation and labeling is done concurrently with the data collection. Post–processing and labeling of the data is still required, but the workload and boundary accuracy is greatly improved.

The Wizard of Oz setup and push–to–sign mechanism in the game allowed us to collect a large amount of signing during testing. This data set is unusual because it consists of samples of children signing and interacting naturally with the game. Most sign language data sets are collected in the lab under controlled conditions with well–enunciated signing. Our data set of signing contains a variety of signing inflections and emphases as well as sign accents common among the children.

3.2 Image Processing

Our data consists of video of the user's signing and accelerometer data from glove-mounted, wireless accelerometers. In our system, we require the children to wear small pink-colored gloves. This bright color is easily identified by a computer vision algorithm. Tracking skin tones can be particularly problematic for computer vision in unconstrained environments. Additionally, it is difficult to distinguish when the hands perform signs near the face. Many algorithms have been suggested to segment hand region robustly, even under illumination change [17, 28, 23, 24, 21]. However, some of them address only a narrow range of illumination change, and some results do not guarantee real-time processing (at least 10 fps with 720×480 sized images in our system) or robustness for long image sequences of gestures. Some methods extract similar color regions as well as hand color region, and the performance strongly depends on the result in the first image frame.

In our approach, the image pixel data is converted to HSV color space and used to create histograms for segmentation of the hand region and background, as shown in figure 4. HSV histograms are used to produce a binary mask using a Bayes classifier [21] and noise is removed by morphological filters including size filtering and hole filtering. The position of the desk and the colored gloves provide a significant marker for starting the gesture recognition; the light color of the desk provides a high contrast environment. The children click the mouse to start and end each phrase, which provides both location with color cues, as well as a start and end gesture. From these cues, we can extract the mouse hand region well for the first frame of the image sequence by simply applying a threshold.

We initially use the hand segmentation to create the starting histogram. Each frame is a segmentation cycle, which provides feedback to the system and helps enhance the discrimination of the color models. HSV histograms are updated with a weight value $\omega(0 < w < 1)$, based on the obtained mask and then the histograms are normalized:

$$H \leftarrow (1 - \omega)H + \omega H^{new}$$

where H denotes the histogram value for each bin [21].

Figure 4 shows the hand tracking process for later frames. The segmentation of the hand region and the update of HSV histograms are the same as the procedure in the first frame. To find both hands in the binary mask, we consider the size of hand shapes and the distance between the center position of the candidate blobs, as well as the hand positions in the previous image frame.

Figure 5 shows the results of the image processing for several image sequences processing occurs at $48.574 ms/frame$ (20.59 fps) in a laptop computer with $1GHz$ processor. We found that the tracking results were acceptable, even when the child wears a shirt with similar color patterns.

3.3 Accelerometer Processing

Our accelerometers (shown in 3) are a custom in–house design created for wearable, wireless sensing. These small wireless sensor platforms provide a Bluetooth serial port profile link to three axes of accelerometer data. The accelerometer is sampled by a PIC microcontroller at approximately 100Hz. The sensors run on a standard camera battery.

Each accelerometer data packet consists of four values – a 16 bit hexadecimal sequence number and three 10 bit hexadecimal values (one each for the X, Y, and Z axes). The axis values represent the gravitational affect of acceleration. Once the data packets are read from the accelerometer they are post-processed. First they are synchronized with our video feed so that the accelerometer data packets for each hand are associated with the correct video frames. Second, the data is smoothed to account for variable number of packets associated with each frame. Because of sampling issues, each video frame can vary in accelerometer packets by one or two packets.

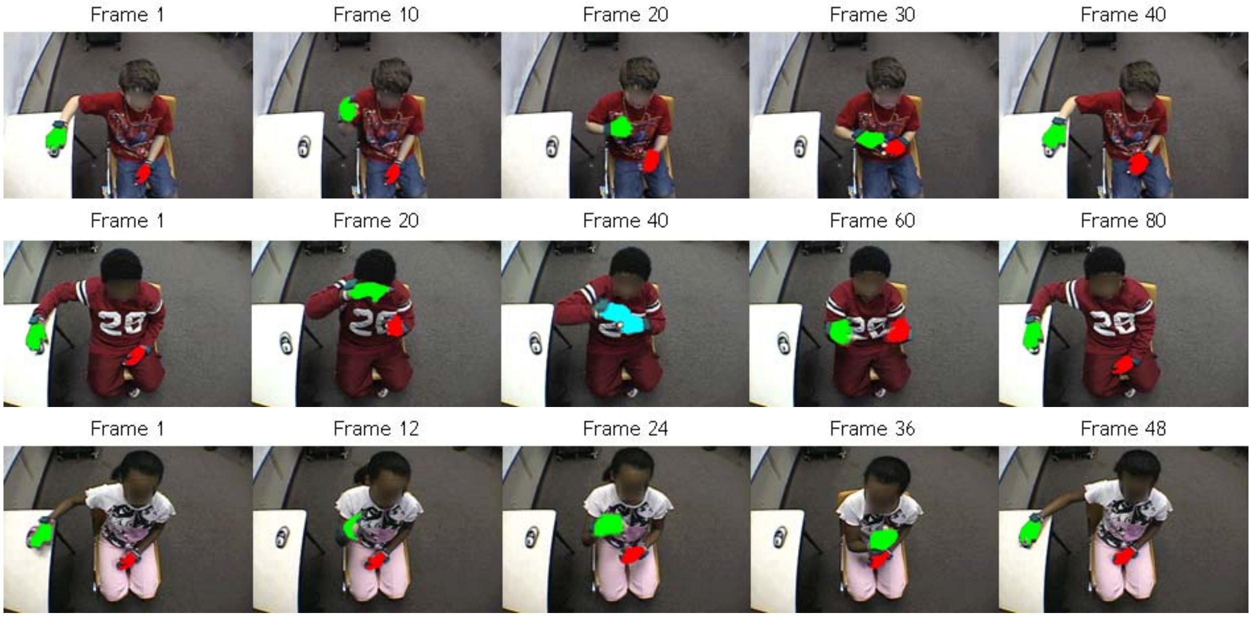

Figure 5: Segmented hand regions in the image sequences. green: right hand, red: left hand, cyan: occluded hands

3.4 Feature Vectors

Our feature vectors consist of the combination of both vision data and accelerometer data. The accelerometer data consists of (x, y, z) values for accelerometers on both hands. The vision data consists of the following hand shape characteristics: the change in x, y center positions between frames, mass, the length of the major and minor axes, eccentricity, orientation angle of the major axis, and direction of the major axis in x, y offset. The camera captures images at 10 frames a second and each frame is synchronized with the averaged accelerometer values. For recognition we adopt left to right HMMs with four states.

3.5 Experiment

The data collected represents each of the five children playing all three levels of the game at least five times each. Each sample consists of one signed phrase from the game. Samples were initially classified as a correctly or incorrectly signed phrase, based on feedback from our consultants from AASD. The "correct" samples were those that were signed correctly according to game play. These samples were further pruned to removed any samples which were evaluated as correct for content but had problems with the signing such as false starts, fidgeting or poorly formed signing. This set of good samples were then labeled to create a transcript of the signed phrase. The final data set represented 541 signed sentences and 1,959 individual signs.

We used the Georgia Tech Gesture Toolkit (GT^2K) [27] to train and test our system. GT^2K adapts HTK (Hidden Markov Model Toolkit) for gesture recognition. HTK provides HMMs in the context of a language infrastructure for use in speech recognition [9]. We used the language tools to train a single model for each sign and then ran additional training for context (equivalent to speech triphone modeling). This procedure allowed us to create stable individual models, and then to combine those models to represent the co–articulation effects of continuous signing. HTK also provides an infrastructure for rule–based and statistical grammars.

ASL is a structured language complete with a grammar, vocabulary, and other linguistic features. Thus, the application of a relevant grammar together with statistical word models can provide a practical solution to remove ambiguity due to disfluency of the deaf children in their signing. Table 4 shows the grammar adopted to our system given in HTK expression, where sil0, sil1, sil2 and sil3 denote silence models by which transitional motions and pauses between words can be segmented in the signing. Coarticulation is the effect that words or signs have on each other when they proceed or succeed one another. The ordering of the silences helps maintain consistency for modeling coarticulation effects.

3.6 Results

3.6.1 User Dependant Models

We evaluated our approach using several different methods. Table 2 shows results from testing user–dependent models – models which are trained and tested using a single child's data. The user–dependent models were generated by training on 90% of the samples randomly selected from a single child's dataset and testing on the other 10%. This was done 100 times for each child and averaged. These show how well the models perform for training and testing for individual users. We achieved an average word accuracy of 93.39% for the user–dependent models.

3.6.2 User Independent Models

Table 3 shows results from testing for user–independent models using leave–one–out validation. User–independent models can be used to recognize signs from multiple children. The user–independent models were generated by training on a dataset consisting of four children and testing on the

Data Set	Word Accuracy		Sentence Correctness	
	Mean	StdDev	Mean	StdDev
Participant 1 (90/10 split)	94.1105%	3.162212	69.3663%	11.95803
Participant 2 (90/10 split)	91.4754%	4.283267	69.3663%	11.95803
Participant 3 (90/10 split)	94.9271%	2.404333	70.9118%	11.48396
Participant 4 (90/10 split)	**95.6477%**	2.61991	**74.6389%**	12.1403
Participant 5 (90/10 split)	90.8016%	3.868346	55.779%	12.2334

Table 2: Comparison of recognition results for user–dependent models

Data Set	Word Accuracy	Sentence Correctness
Leave one out: Participant 1	87.33%	51.69%
Leave one out: Participant 2	76.90%	38.24%
Leave one out: Participant 3	**92.62%**	**61.61%**
Leave one out: Participant 4	90.94%	56.76%
Leave one out: Participant 5	83.60%	44.90%
Leave one out: Average	86.28% - stdev 6.29	50.64% - stdev 9.3

Table 3: Comparison of recognition results of user–independent models

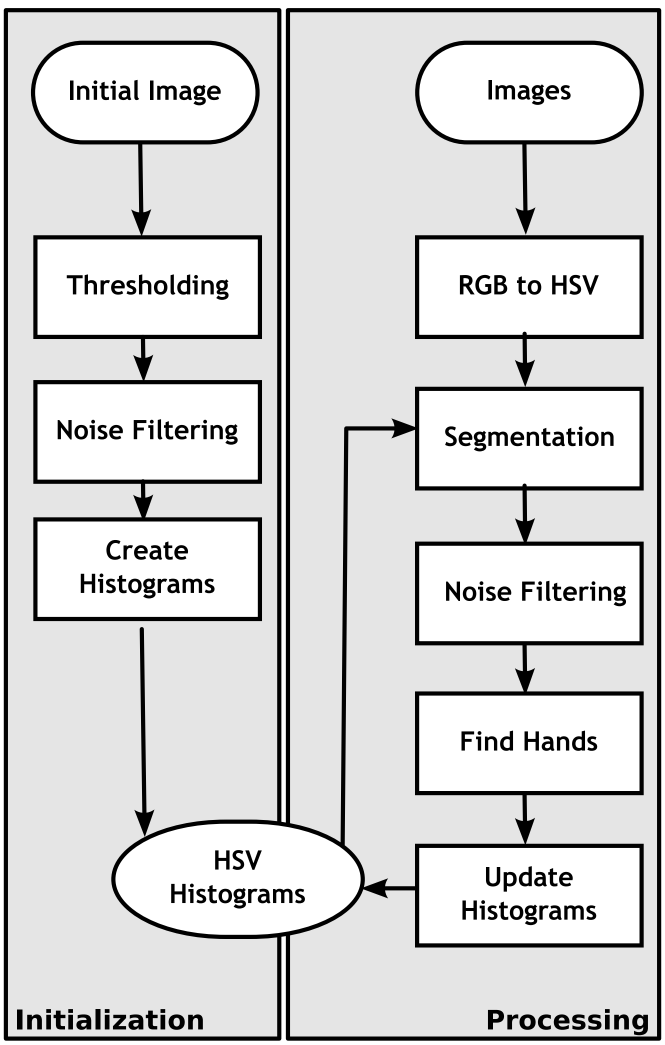

Figure 4: Hand segmentation for the first image frame

other child's dataset. This is done for each child. This helps us understand how the models generalize across users. We achieved an average word accuracy of 86.28% for the user–independent models.

3.6.3 All samples

We achieved on average 92.96% of accuracy in word-level with 1.62% of standard deviation when we chose samples across all samples and users (we trained and tested using data from all students). All 541 sentence examples were randomly divided into a set of 90% training sentences and a set of 10% independent test sentences. The test sentences were not used for any portion of the training. We repeated this training and test cycle 100 times using HTK and calculated the average of the recognition accuracy.

3.6.4 Summary

We ran three sets of experiments with our data sets: user–dependant, user–independent, and across all samples. The user–dependent models show how well recognizer performs for a single individual. All of the students signing can be modeled fairly well, with a greater than 90% word accuracy. However, when recognizing sentences, words can be inserted and deleted which creates errors that lower sentence accuracy. The range in word accuracies from 90.8% to 95.6% shows a strong variation in the user's signing samples – some are clearly modeled better than others.

Participant four has the highest word accuracy and the second lowest standard deviation on the tests. These results can be an indicator that participant four probably had clear, consistent signing. Participant two the second lowest word accuracy and the highest standard deviation. When used as the test set for leave–one–out validation, participant two scored the worst against the models. Participants three and four were the top two performers for word accuracies and standard deviation in both tests. Participants that signed clear and consistent data (evident by high word accuracies and low standard deviations over the user dependent models) were also well modeled by the more general user independent models.

The limited size of the participant pool restricts the generalizations that can be drawn from the experiment, but these experiments show interesting trends that should be investigated for larger population sizes. Increasing the number of users for both creating and testing models will give a broader idea of the generalization. There is a tension between the usefulness of generalization and added accuracy of user–specific training; this can be seen in speech recognition packages that ship with models trained from large populations and allow the user to help train the models by providing additional training samples.

The leave–one–out validation shows how the models generalize to users they have never seen. The wider range of word accuracies from 76.9 % to 92.6% show the impact that including or excluding certain students from the training set can affect the outcome. The consistency in performance by participants in the user–dependent and user–independent tests indicates that the user–independent models are doing a good job of generalizing. The independent models are generalizing well, even across models of varying qualities.

4. CONCLUSIONS

We present continued work on a computer game designed to help children practice their ASL skills and encourage their linguistic development. Linguistically, the system is designed to help children practice their vocabulary and encourage them to generate phrases which convey complete thoughts and ideas. CopyCat provides a new domain for the sign language recognition community and has provided a unique data set of native signers interacting naturally with a computer system. The recognition engine represents an important step towards the recognition of conversational sign in contrast to other systems which largely use scripted sign collected in controlled laboratory environments.

We have expanded the functionality of our previous systems [22, 2, 15] to handle non-scripted, live signing which includes pauses and variable coarticulation effects. We show progress with the gesture recognition component of the game. We use a hand segmentation algorithm which is robust against illumination change and guarantees a real-time processing. We also adopted a strong grammar to alleviate the effects of disfluencies and to achieve a 92.96% word accuracy. Our dataset is unique in the sign recognition community because it uses signers conversing in a spontaneous, unscripted manner. These results show the feasibility of the system and provide a platform for further developing the recognition capabilities.

4.1 Future Work

Though this work has extended the functionality of our recognition system, it is clear that there is much work to do. The dataset collected for our experiments is both exciting and challenging. This evaluation uses samples selected from the dataset for their correctness and clarity in signing. These samples are used for building models and recognizing signs. Our next challenge is enabling the recognition engine to deal with the disfluencies present in otherwise linguistically correct samples. Disfluencies of importance include long pauses, fidgets, hesitations and false starts in signing.

We have completed the next round of iterative development with a user study at AASD. The study included another round of interface development and data collection, as well as pre–study and post–study linguistic evaluations to begin to explore the learning effects of the system. Preliminary analysis of the data from this study has been extremely informative. We are currently focused on classifying disfluencies and out–of–band communication by the children during game play to provide more informed models and increase recognition accuracy. The full analysis of this study will provide us with further insight to educational value of the system, as well as increase our data bank for further recognition engine work.

5. 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 for their generous help with this project.

6. REFERENCES

[1] B. Bauer, H. Hienz, and K. Kraiss. Video-based continuous sign language recognition using statistical methods. In *Proceedings of the 15th International Conference on Pattern Recognition*, volume 2, pages 463–466, September 2000.

[2] H. Brashear, T. Starner, P. Lukowicz, and H. Junker. Using multiple sensors for mobile sign language recognition. In *Proceedings of the Seventh IEEE International Symposium on Wearable Computers*, pages 45–52, 2003.

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

[4] G. Fang, W. Gao, and D. Zhao. Large vocabulary sign language recognition based on hierarchical decision trees. In *International Conference on Multimodal Interfaces*, pages 125–131, 2003.

[5] Gallaudet. 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., 2001.

[6] W. Gao, G. Fang, D. Zhao, and Y. Chen. Transition movement models for large vocabulary continuous sign language recognition (csl). In *Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, pages 553–558, 2004.

[7] V. Henderson, S. Lee, H. Brashear, H. Hamilton, T. Starner, and S. Hamilton. Development of an American sign language game for deaf children. In *Proceedings of the 4th International Conference for Interaction Design and Children*, Boulder, CO, 2005.

[8] J. L. Hernandez-Rebollar, N. Kyriakopoulos, and R. W. Lindeman. A new instrumented approach for translating American sign language into sound and text. In *Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, pages 547–552, 2004.

[9] http://htk.eng.cam.ac.uk.

[10] IDRT. Con-sign-tration. Product Information on the World Wide Web, Institute for Disabilities Research and Training Inc.,

```
$predator = snake | spider | alligator ;
$color1 = white | green | blue ;
$color2 = black | white | green ;
$color3 = orange | black | white ;
$color4 = blue | orange | black ;
$color5 = green | blue | orange ;

$sentence = $color1 sil1 kitten sil2 behind wagon
          | $color2 sil1 kitten sil2 under chair
          | $color3 sil1 kitten sil2 in flowers
          | $color4 sil1 kitten sil2 in bedroom
          | $color5 sil1 kitten sil2 on wall ;

( sil0 go sil1 chase sil2 $predator sil3 | sil0 $sentence sil3 )
```

Table 4: Strong grammar adopted to the system

http://www.idrt.com/ProductInfo.php?ID=32&u=1, 1999.

[11] J. S. Kim, W. Jang, and Z. Bien. A dynamic gesture recognition system for the Korean sign language KSL. *IEEE Transactions on Systems, Man and Cybernetics*, 26(2):354–359, 1996.

[12] S. Lee, V. Henderson, H. Hamilton, T. Starner, H. Brashear, and S. Hamilton. A gesture-based American sign language game for deaf children. In *Proceedings of CHI*, pages 1589–1592, Portland, Oregon, 2005.

[13] R. Liang and M. Ouhyoung. A real-time continuous gesture recognition system for sign language. In *Third International Conference on Automatic Face and Gesture Recognition*, pages 558–565, 1998.

[14] R. I. Mayberry and E. B. Eichen. 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, 1991.

[15] R. M. McGuire, J. Hernandez-Rebollar, T. Starner, V. Henderson, H. Brashear, and D. S. Ross. Towards a one-way American sign language translator. In *Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, pages 620–625, 2004.

[16] E. L. Newport. Maturational constraints on language learning. *Cognitive Science*, 14:11–28, 1990.

[17] N. Oliver, A. Pentland, and F. Berard. Lafter: Lips and face real time tracker. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 123–129, 1997.

[18] G. Pollard and TSD. Texas School for the Deaf. Aesop: Four Fables, 1998.

[19] D. Roberts, U. Foehr, V. Rideout, and M. Brodie. Kids and Media @ the New Millennium, 1999.

[20] H. Sagawa and M. Takeuchi. A method for recognizing a sequence of sign language words represented in a japanese sign language sentence. In *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, pages 434–439, Grenoble, France, March 2000.

[21] L. Sigal, S. Sclaroff, and V. Athitsos. Skin color-based video segmentation under time-varying illumination. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(7):862–877, 2004.

[22] T. Starner and A. Pentland. Visual recognition of American sign language using hidden markov models. In *Proceedings of the International Workshop on Automatic Face and Gesture Recognition*, 1995.

[23] M. Storring, H. Andersen, and E. Granum. Skin colour detection under changing lighting conditions. In *Proceedings of the Seventh Symposium on Intelligent Robotics Systems*, pages 187–195, 1999.

[24] M. Storring, H. Andersen, and E. Granum. Estimation of the illuminant colour from human skin colour. In *Proceedings of International Conference on Automatic Face and Gesture Recognition*, pages 64–69, 2000.

[25] C. Vogler and D. Metaxas. ASL recognition based on a coupling between hmms and 3d motion analysis. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 363–369, 1998.

[26] C. Vogler and D. Metaxas. Handshapes and movements: Multiple-channel american sign language recognition. In *Springer Lecture notes in Artificial Intelligence*, volume 2915, pages 247–258, January 2004.

[27] T. Westeyn, H. Brashear, A. Atrash, and T. Starner. Georgia tech gesture toolkit: Supporting experiments in gesture recognition. In *ICMI '03: Proceedings of the 5th International Conference on Multimodal Interfaces*, pages 85–92, New York, NY, USA, 2003. ACM Press.

[28] J. Yang, L. Weier, and A. Waibel. Skin-color modeling and adaptation. In *Proceedings of Asian Conference on Computer Vision*, pages 687–694, 1998.