Improving the Efficacy of Automated Sign Language Practice Tools

Helene Brashear

Georgia Institute of Technology, GVU Center, College of Computing, Atlanta, Georgia, USA

brashear@cc.gatech.edu

Abstract:

CopyCat is an America Sign Language (ASL) game, which uses gesture recognition technology to help young Deaf children practice ASL skills. 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. We have created an automatic sign language recognition system for the game. We believe that we can improve the accuracy of this system by characterizing and modeling disfluencies found in the children's signing.

Introduction

Since early childhood is a critical period for language acquisition, early exposure to ASL is key for deaf children's linguistic development [14,15]. Ninety percent of deaf children are born to hearing parents. Most of these parents do not know or are not fluent in sign language [7]. Often a child's first exposure to signing is at school.

CopyCat is a research prototype combining an interactive computer game with sign language recognition technology. CopyCat aims to assist young deaf children's language acquisition by interactive tutoring and real-time evaluation [12]. The goal is to encourage the linguistic transition from single, isolated utterances to phrase level signing. CopyCat is designed to support ASL-based communication between the user and characters in the computer game. The child is asked to sit in front of the computer which is equipped with a video camera for computer vision recognition system. He or she wears colored gloves with wrist-mounted accelerometers to assist the recognition. While playing the game, the child communicates with an animated character through ASL. This game is both mentally and physically engaging and allows the child to practice ASL.

Related Work

Sign languages are used around the world by the deaf and speech impaired as a means of communication. These sign languages use hand, body and face gestures as well as spatial structures to communicate information. Automatic sign language recognition (ASR) is the process of using sensors to collect data from a user's signing and use computers to recognize the signs. Sign languages are rich, multi-faceted languages, and their full complexity is beyond current gesture recognition technologies. As a result of this, most researchers focus on a subset or aspect of a sign language to reduce the complexity of their task. Some of these subsets are finger spelling, hand-based gestures, facial gestures, body gestures, inflections, and spatial use.

Research on sign language recognition has been done around the world, using many sign languages, including American Sign Language, Korean Sign Language, Auslan, Taiwanese Sign Language, Chinese Sign Language, Japanese Sign Language, and German Sign

Language. Many sign language recognition systems use Hidden Markov Models (HMMs) for their abilities to train useful models from limited and potentially noisy sensor data. Sensor choices vary from data gloves and other tracker systems to computer vision techniques using a single camera, multiple cameras, and motion capture systems to hand crafted sensor networks. For a more detailed summary of work in the field of ASR see the work of Ten Holt *et al* [10] and Loeding *et al* [13].

Holt *et al* [10, 9] provide a comprehensive summary of t problems of ASR research. Holt describes several significant problems specific to ASR:

- Distinguishing gestures from signs
- Context dependency (directional verbs, inflections, etc)
- Basic unit of modeling (what are the phonemes? how do we describe them?)
- Transitions between signs (movement epenthesis vs. coarticulation effects)
- Repetition (cycles of movement may vary in length)

Transitional Movements

Without modeling of the basic structure of signs (which is debated), it is problematic to use traditional phonemic methods of modeling. Transitions between signs include both coarticulation effects and movement epentheses. Coarticulation effects refer to the changing of signs when they overlap[11]. Movement epenthesis is the actual movement between signs when there is a difference between a signs ending location and the starting location of another sign [17]. Distinguishing gestures from sign is a fundamental problem, since many of these non-sign gestures occur during conversation and act as disfluencies.

Speech recognition typically trains co-articulation models in context by creating biphones and triphones. Gao et al used this approach in modeling Chinese Sign Language with HMMs [8]. Vogler and Metaxas modeled each transition individually as a phoneme using parallel HMMs [18]. Rule based methods have also been used; these methods use heuristics such as acceleration and velocity to detect borders and transitions [16].

Data Sets

Most data sets for automatic sign language recognition are scripted data sets collected in the laboratory by the researchers [9]. Though scripted datasets provide a good testing bed for developing research systems, the field of speech recognition has found that they are limited in their representation of how language is used and lack examples of common conversational artifacts such as accents (since they are often over-enunciated), disfluencies, and inflections [11]. Additionally conversational signing may contain register variation which may result in more or less formal signing depending on the signer's environment. This register variation can affect how signs are performed, what vocabulary is chosen and what grammar is used [17]. Datasets collected in formal, scripted settings may lack many of the important language facets that are needed to fully model the language for use in live recognition systems.

The dataset from the CopyCat project [1] is particularly interesting because it was designed to capture more conversational signing. The dataset was collected from multiple children in a school environment. The children signed to characters in the game as conversational partners in order to complete tasks within the game story line.

Research

The CopyCat dataset was collected from 9 children ages 8-11 at the Atlanta Area School for the Deaf. Children played the game for 5 sessions each, over a period of two weeks. There were 30 game phrases per a session and a game vocabulary of 20 different hand-based signs. The grammar was limited to `Subject+Preposition+Object` format and includes sentences of four, five and six sign lengths. The system used cameras, colored cotton gloves, and accelerometers as sensors.

The dataset was collected using a Wizard of Oz approach. The Wizard of Oz 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, thus emulating the missing functionality of the system [3]. During game play, the children interact normally with the computer game while a human Wizard simulates the computer recognizer and evaluates the correctness of the player's signing. The child is not aware of the Wizard's presence and believes that he/she is using a fully functioning system.

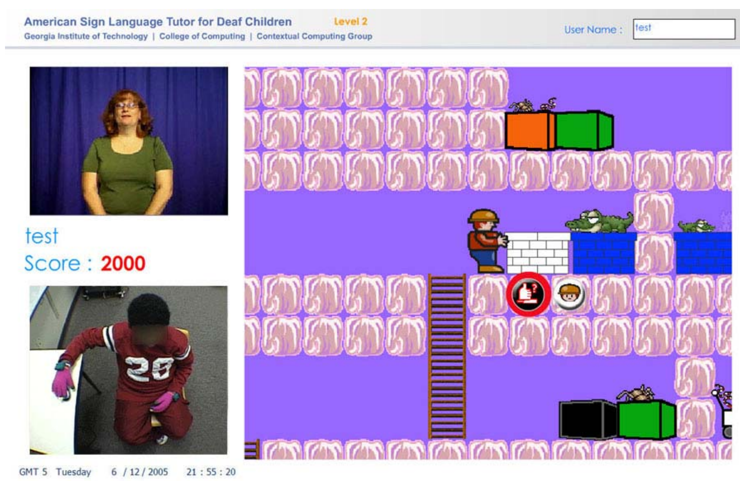

Figure 1: CopyCat screenshot: The child directs Abe (A), the hero, to defeat the villain (B). If the child is unclear on the game scenario, the he/she presses the help button (E) and a tutor video (C) of the properly signed phase plays. The child then presses the start button (F) and signs to the character Abe, making sure that Abe can ``see'' him or her in the camera view (D).

Thesis

My sign language recognition research is focused on creating user-independent models for recognition of hand-based American Sign Language gestures[1]. Our data set will be a collection of samples of the children interacting with the game via a Wizard of Oz setup at the Atlanta Area School for the Deaf (AASD). These models will be used to create our live, interactive game. Since most of the sign language sets (including our own past work) used for machine learning have been collected in a controlled, laboratory environment, these sets do not fully explore common disfluencies in sign. Our data set provides many samples of children signing naturally as they interact with the online characters. It has many examples of non-signing activities such as scratching and fidgeting and also includes false starts, hesitations, and pauses.

Thesis Statement: I will show that modeling the disfluencies that occur in a natural signing context can improve the accuracy of a sign recognition system for an ASL practice tool.

The transition from lab collected signing samples to real-world datasets for sign language recognition necessitates expanding models to include more diverse linguistic information. Just as the speech recognition community found that there is more to speech recognition than well-enunciated speech signals, there is more to sign language than perfectly performed signing. The linguistic scope of sign recognition is still largely limited by technology, with most groups focusing on hand gesture recognition.

Contributions: I will **identify significant gestures in our dataset**, including game vocabulary, communications directed towards game characters, and disfluencies in sign. We will enumerate and model the disfluencies that are relevant to our data set. I will **create an ontology for disfluency classification** for recognition purposes. This ontology will help group disfluencies by their structure and analyze their impact on the recognition system. I can **improve our system by modeling select classes of gestures from the ontology**. These models will be added to the recognition engine in order to improve the recognition rates.

Table 1: Categorization of sign examples on two axes: Clear vs. Unclear and Game Correct vs. Game Incorrect.

	Machine Learning	
	Clear	**Unclear**
Game Correct	Clear, accurate signing which follow the game grammar; these are the ``textbook'' correct samples	Signing which follows the game grammar, but may contain disfluencies; contains signing which is intelligible and grammatically correct, but may contain unusual linguistic artifacts
Game Incorrect	Clear, accurate signing which does not reflect game grammar; may contain good signs without any grammatical structure	Signing which does not follow game grammar and is incorrect; may include gibberish, wrong signs and disfluencies

Modeling Signs

The CopyCat dataset can be divided into four categories, as shown in Table 1. The traditional approach to pattern recognition datasets would be to choose the good examples (the intersection of ``Game Correct'' and ``Clear'' on Table 1). These samples would then be used to both build models and to evaluate the accuracy of these models. We have shown results from this approach for a dataset collected from an earlier phase of data collection for CopyCat [1].

We have previously used a single hidden Markov model for each sign along with a strong grammar designed to account for coarticulation effects, movement epentheses, and some pauses. These results showed user-dependent models with testing word accuracies from 90.80% to 95.65%. User- independent models were tested and showed word accuracies ranging from 76.90% to 92.62%. The signing skills of the participants varied, but the word accuracies show that their signing could be modeled and recognized with some success, even with such a small number of children to sample from.

The next step for expanding the functionality of the system is to move beyond these traditional techniques and recognize the signs using more detailed language models. The children have a variety of disfluencies in their signing both as a result of conversational mannerisms, as well as a result of their concentration and effort in the game. It is important to explore the kinds of disfluencies the children exhibit, as well as further analyze the structure of these disfluencies production in order to improve our modeling techniques. A live version of the game would need ASR that can differentiate between a badly formed sign and disfluencies such as head scratch or a long pause. Though there has been limited investigation by linguists, the work of Eklund [4] provides a short overview of some perspectives on disfluency research in sign languages.

Preliminary Results

A preliminary analysis of the data evaluated 16 of the 45 total sessions in order to further refine the groupings of sign. This evaluation included between one and three sessions from each child spread across various sessions. Of 514 total phrases, only 117 of them were from the ``Game Correct'' and ``Clear'' category from 1. The majority of the phrases (354) were in the ``Game Correct'' and ``Unclear'' category. These samples included a wide range of variations in signing that were conversationally correct, but not perfect. ``Game Incorrect'' phrases accounted for 43 of the samples, but only seven of those contained bad signs. The rest contained ``Clear'' or "Unclear" signing which was understandable, but wrong for the game (for example referring to the wrong color).

During game play of Castle Quest, the children interacted with Abe as a natural conversational partner. This interaction resulted in instances of ``out-of-band'' comments to the character. These instructions to Abe were not part of the game story line, but instead, were directly related to game play. Self-corrections accounted for a large part of the ``out-of-band'' comments and some of the disfluencies. Detecting self-corrections may be aided by some ``out-of-band'' comments such as `wrong` and `start again` or some non-sign gestures such as a hand waving ``erase'' gesture that several children exhibited. Studies have shown that adult signers appear to normally edit their sign during conversation (similarly to spoken language) and produce self corrections such as `wrong` or a head shake[2,6], however there have not been systematic studies of conversational repairs in ASL[5].

There is a wide variation in how pauses occur in the library of sign examples. Some pauses occur between signs at various stages: post-completion of the first sign, during the movement epenthesis, and immediately prior to the second sign. Hesitations and false starts are commonly interjected into parts of the phrase where children have to concentrate on the subject matter to successfully complete a game goal. These hesitations often occur mid-sign as the child attempts to think of the next sign in the phrase.

False starts are also sometimes symptomatic of extra concentration and self-corrections. In particular this seems to manifest itself as starting a sign with the wrong hand shape and then changing hand shapes, because the intended sign has changed. For example, the child starts to sign with the closed hand shape for the sign `spider`, then realized that they need to sign `alligator` and switches to the claw hand shape to begin the sign for `alligator`.

While there may be some consistency between non-sign gestures for the children, it is not clear whether this generalizes outside of this particular dataset. Fidgeting and wiggling account for some active variation in signing and that will probably generalize fairly well since it only causes slight permutations in the sensor data.

The children were inconsistent in their performance of signs which contain cycles. The actual repetition of cycles was sometimes more or less than required for the sign. Modeling this repetition will involve changing the structure of the models at a sign level.

Next Steps

It has become clear that it will be important to expand sign labeling schemes to include more linguistic information than the traditional ``gesture label'' by adding tags for variations in the sign that are a result of signer preference, accent, or regional variation. While some of these variations can be accounted for mathematically in the models, it is clear that a more detailed labeling scheme could provide more accurate models. The next step will be to begin to catalog these variations and develop approaches to modeling them.

Conclusion

In order to create live applications that use automatic sign language recognition, language models much be expanded from tradition gesture recognition based approaches to include more detailed linguistic information. Most automatic sign language recognition systems are built on scripted sessions collected in the laboratory. The CopyCat dataset provides an interesting alternative to these scripted systems because it was collected using Wizard of Oz methods. These methods allowed researchers to collected data of children signing conversationally to characters in the game. Further analysis reveals a variety of interesting variations and disfluencies in the children's sign which serve to inform new sign language modeling techniques and bring live automatic sign language recognition closer to a reality.

Acknowledgments

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.

References

1. H. Brashear, K.-H. Park, S. Lee, V. Henderson, H. Hamilton, and T. Starner. 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, New York, NY USA, 2006. ACM Press.
2. V. L. Dively. Conversational Repairs in ASL. In C. Lucas, editor, Pinky extension and eye gaze: Language use in Deaf communities, pages 137-169, Washington DC, 1998. Gallaudet University Press.
3. A. Dix, J. Finlay, G. Abowd, and R. Beale. Human-Computer Interaction, chapter 6.4 Iterative Design and Prototyping. Prentice Hall, 2004.
4. R. Eklund. Disfluency in Swedish Human-Human and Human-Machine Travel Booking Dialogues: Dissertation No. 882. PhD thesis, Linköping Studies in Science and Technology, Linköping University, Sweden, 2004.

5. K. Emmorey. Language, Cognition, and the Brain: Insights From Sign Language Research. Lawrence Erlbaum Associates, Mahwah, New Jersey, 2002.
6. K. Emmorey and B. Falgier. Talking about space with space: Describing environments in ASL. In E. A. Winston, editor, Story telling and conversations: Discourse in deaf communities, pages 3-26, Washington DC, 1999. Gallaudet University Press.
7. 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.
8. W. Gao, J. Ma, J. Wu, and C. Wang. Sign language recognition based on hmm / ann / dp. In International Journal of Pattern Recognition and Artificial Intelligence, volume 5, pages 587-602, 2000.
9. G. T. Holt. The Eye of the Beholder: Automatic Regonition of Dutch Sign Language. Master's thesis, University of Groningen, Netherlands, 2004.
10. G. T. Holt, P. Hendriks, and T. Andringa. Why Don't You See What I Mean? Prospects and Limitations of Current Automatic Sign Recognition Research. Sign Language Studies, 6(4), Summer 2006.
11. D. Jurafsky and J. H. Martin. Speech and Language Processing. Prentice Hall, Upper Saddle River, New Jersey, 2000.
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. B. L. Loeding, S. Sarkar, A. Parashar, and A. I. Karshmer. Progress in Automated Computer Recognition of Sign Language. In Proceedings of International Conference on Computers Helping People with Special Needs, pages 1079-1087. Springer-Verlag Berlin Heidelberg, 2004. Lecture Notes in Computer Science.
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. E. L. Newport. Maturational Constraints on Language Learning. Cognitive Science, 14:11-28, 1990.
16. 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.
17. C. Valli and C. Lucas. Linguistics of American Sign Language: An Introduction. Galludet University Press, Washington DC, 1998.
18. C. Vogler and D. Metaxas. Toward scalability in asl recognition: Breaking down signs into phonemes. In Proceedings of the Gesture Workshop (Gif-sur-Yvette, France), volume 1739, pages 211-224. March 1999. Springer Lecture Notes in Artificial Intelligence.

About the author

Helene Brashear is currently a PhD student at the College of Computing at the Georgia Institute of Technology. Her research focus is in using pattern recognition to create intelligent systems for assistive technology. She is working on her dissertation "Improving the Efficacy of Automated Sign Language Practice Tools" with her advisor, Dr. Thad Starner.