# Gesture Recognition System for Isolated Sign Language Signs

Kalin Stefanov and Jonas Beskow

KTH Royal Institute of Technology
TMH Department of Speech, Music and Hearing
Lindstedtsvägen 24, SE-100 44 Stockholm, Sweden
{kalins,beskow}@kth.se
http://www.speech.kth.se/

**Abstract.** This paper describes a system for recognition of isolated Swedish sign language signs for the purpose of an educational "signing game". The primary target group is children with communicative disabilities, and the goal is to offer a playful and interactive way of learning and practicing sign language signs for these children and their friends and families. Two datasets consisting of 51 signs have been recorded for a total of 7 (experienced) and 10 (inexperienced) adult signers. The signers performed all of the signs five times and were captured with a RGB-D (Kinect) sensor, via a purpose-built recording application. A recognizer based on manual features is presented and tested on the collected datasets. Signer dependent recognition rate is 95.3% for the most consistent signer. Signer independent recognition rate is on average 57.9% for the experienced signers and 68.9% for the inexperienced.

**Keywords:** sign language recognition, key word signing, depth sensor, real-time

## 1 Introduction

Sign language and different forms of sign based communication is important to large groups in society. In addition to members of the deaf community, that often have sign language as their first language, there is a large group of people who use verbal communication but rely on signing as a complement. A child born with hearing impairment or some form of communication disability such as developmental disorder, language disorder, cerebral palsy or autism, frequently have the need for this type of communication known as *key word signing*. Key word signing systems borrow individual signs from sign language to support and enforce the verbal communication. As such, these communication support schemes do away with the grammatical constructs in sign language and keep only parts of the vocabulary.

While many deaf children have sign language as their first language and are able to pick it up in a natural way from the environment, children that need signs for other reasons do not have the same rights and opportunities to be introduced to signs and signing. The Swedish Tivoli project, that forms the context of the system presented in this paper, aims at creating a learning environment where children can pick up signs in a game-like setting. An on-screen avatar presents the signs and gives the child certain tasks to accomplish, and in doing so the child gets to practice the signs. The system is thus required to interpret the signs produced by the child and distinguish them from other signs, and indicate whether or not it is the right one and if it was properly carried out.

## 2 Related Work

This paper presents a gesture recognition system that attempts to model and recognize manual features of sign language. Cooper, Holt and Bowden [1] provides comprehensive overview of the research on sign language recognition (SLR) and the main challenges. Manual features of sign language are in general, hand shape/orientation and movement trajectories which are similar to gestures. A comprehensive survey on gesture recognition (GR) was performed by Mitra and Acharya [2].

A time-domain process demonstrates a Markov property if the conditional probability density of the current event, given all present and past events, depends only on the $j$th most recent event. If the current event depends solely on the most recent past event, then the process is termed a first order Markov process. This assumption is reasonable to make, when considering the positions of the hands of a person through time. The generalized topology of a hidden Markov model (HMM) is a fully connected structure, known as an *ergodic* model, where any state can be reached from any other state. When employed in dynamic gesture recognition, the state index transits only from left to right with time, as depicted in Figure 1. Here the state transition probabilities $a_{ij} = 0$ if $j < i$, and $\sum_{j=1}^{N} a_{ij} = 1$.

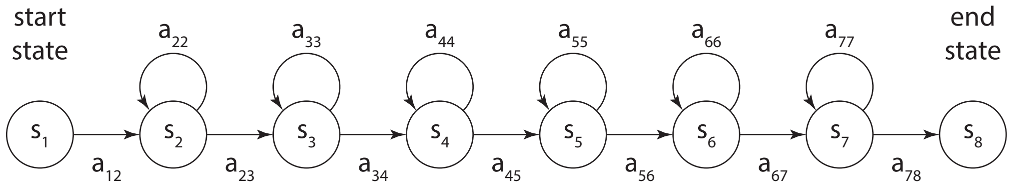


**Fig. 1.** 8-state left-to-right HMM used for gesture modeling.

## 3 Recognizer

The system is trained to perform signer independent recognition in real-time. The size of the vocabulary is 51 signs from the Swedish sign language (SSL). The vocabulary is composed of four subsets - objects, colors, animals, and attributes. The recognition rate (accuracy) in the signer dependent case is presented in Table 1. The models were trained on 4 instances of each sign for each participant and tested on the 5th instance of the sign performed by that participant. The recognition rate (accuracy) in the signer independent case is presented in Table 2. The models were trained on 30 (six different experienced signers) or 45 (nine different inexperienced signers) instances of each sign and tested on the 5 instances of the sign performed by the left out participant. All results are based on leave-one-out cross-validation procedure.

**Table 1.** Signer dependent results

| Experienced | $\mu$ | $\sigma$ | Inexperienced | $\mu$ | $\sigma$ |
|---|---|---|---|---|---|
| 1 | 92 | 5.1 | 1 | 84.3 | 4.2 |
| 2 | 85.5 | 11.1 | 2 | 92.2 | 6 |
| 3 | 84.4 | 7.5 | 3 | 75.7 | 12.1 |
| 4 | **95.3** | 3.3 | 4 | 94.9 | 1.7 |
| 5 | 88 | 14.6 | 5 | 94.5 | 4.2 |
| 6 | 87.8 | 7.8 | 6 | 93.3 | 9.4 |
| 7 | 80 | 10.8 | 7 | 89.4 | 5.8 |
| | | | 8 | **95.3** | 5.5 |
| | | | 9 | 94.1 | 6.8 |
| | | | 10 | 89.8 | 6.4 |

**Table 2.** Signer independent results

| Experienced | 1-BEST | Inexperienced | 1-BEST |
|---|---|---|---|
| 1 | 62 | 1 | 58.8 |
| 2 | 53.3 | 2 | 72.2 |
| 3 | 48 | 3 | 57.3 |
| 4 | **65.5** | 4 | 69.4 |
| 5 | 62.6 | 5 | 75.3 |
| 6 | 59.2 | 6 | **76.1** |
| 7 | 54.9 | 7 | 73.7 |
| | | 8 | 65.5 |
| | | 9 | **76.1** |
| | | 10 | 64.7 |
| $\mu$ | 57.9 | $\mu$ | 68.9 |
| $\sigma$ | 6.1 | $\sigma$ | 7 |

As expected, the performance in the signer independent case is significantly lower than in the signer dependent case - 57.9% and 68.9% compared to 87.6% and 90.3% when averaged over all signers. These accuracy rates are however for the full set of 51 signs. In our application, there is no situation where the recognizer needs to pick one sign from the full set, instead there is always the case of one out of a small number (e.g. choose one out of five objects). For this type of limited recognition tasks, accuracy will increase drastically. Furthermore, we can control the mix of signs in the game, meaning that we can make sure that signs which are confused by the recognizer never appear together, therefore, the recognition accuracy of the signer independent recognizer will not be a limiting factor in the game.

In this work each sign is modeled with the same simple HMM (see Figure 1). More complex and different models for each sign will be considered in future work. Further improvements are expected by introducing adaptation of the HMMs based on a small set of signs from the target signer that could be collected during an enrollment/training phase in the game.

## Acknowledgments

The project Tivoli (Sign learning through games and playful interaction) is supported by the Swedish Post and Telecom Authority (PTS). We would like to thank Krister Schönström for the support during recording the experienced signers dataset.

## References

1. H. Cooper, B. Holt, R. Bowden, Sign language recognition, in: T. B. Moeslund, A. Hilton, V. Kroger, L. Sigal (Eds.), Visual Analysis of Humans, Springer, 2011, pp. 539-562.
2. S. Mitra, T. Acharya, Gesture recognition: A survey, IEEE Transactions on Systems, Man, and Cybernetics 37 (3) (2007) 311-324.