

Sign Language Perception Research for Improving Automatic Sign Language Recognition

Gineke A. ten Holt^{a,b}, Jeroen Arendsen^a, Huib de Ridder^a, Andrea J. van Doorn^a, Marcel J.T. Reinders^b and Emile A. Hendriks^b

^aHuman Information Communication Design, Delft University of Technology, Landbergstraat 15, 2628 CE Delft, the Netherlands;

^bInformation and Communication Theory Group, Delft University of Technology, Mekelweg 4, 2628 CD Delft, the Netherlands

ABSTRACT

Current automatic sign language recognition (ASLR) seldom uses perceptual knowledge about the recognition of sign language. Using such knowledge can improve ASLR because it can give an indication which elements or phases of a sign are important for its meaning. Also, the current generation of data-driven ASLR methods has shortcomings which may not be solvable without the use of knowledge on human sign language processing. Handling variation in the precise execution of signs is an example of such shortcomings: data-driven methods (which include almost all current methods) have difficulty recognizing signs that deviate too much from the examples that were used to train the method. Insight into human sign processing is needed to solve these problems. Perceptual research on sign language can provide such insights. This paper discusses knowledge derived from a set of sign perception experiments, and the application of such knowledge in ASLR. Among the findings are the facts that not all phases and elements of a sign are equally informative, that defining the ‘correct’ form for a sign is not trivial, and that statistical ASLR methods do not necessarily arrive at sign representations that resemble those of human beings. Apparently, current ASLR methods are quite different from human observers: their method of learning gives them different sign definitions, they regard each moment and element of a sign as equally important and they employ a single definition of ‘correct’ for all circumstances. If the object is for an ASLR method to handle natural sign language, then the insights from sign perception research must be integrated into ASLR.

Keywords: sign language perception, automatic sign language recognition, information distribution in signs, variation in signs

1. INTRODUCTION

The current tendency in man-machine interaction is toward more and more natural methods of interfacing. It is not surprising, therefore, that automatic gesture processing is receiving a lot of attention. Most of this attention is focused on technical issues. However, to process natural gestures, attention must also be paid to the nature of human communication. The importance of this element depends on the application for which the gesture processing is to be used. In natural language processing or language teaching environments, the human criteria for communication must be adapted, because ultimately the machine’s goal is human communication. Current gesture processing largely ignores this issue, concentrating mostly on data-driven methods. This can lead to undesirable results, an example of which is the outlier problem: a gesture executed in a strange, but nevertheless correct way, will most often be rejected by a statistical gesture processing method. Such a method learns the definition of the gesture from example data. If the strange variant was not in its training data, then it will reject it henceforth. ‘Common’ and ‘correct’ are largely equivalent for such methods. For humans, the notion “uncommon but correct” is conceivable. Apparently, humans employ different criteria for judging correctness than mere frequency of occurrence. If a machine algorithm is to process natural gestures, it must adapt similar methods and cannot rely on data-driven methods alone. More knowledge on human gesture processing must be

Further author information: (Send correspondence to G.A. ten Holt)

G.A. ten Holt: E-mail: g.a.tenholt@tudelft.nl, Telephone: +31 (0)15 2788435

Human Vision and Electronic Imaging XIV, edited by Bernice E. Rogowitz,
Thrasyloulos N. Pappas, Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 7240, 72400C
© 2009 SPIE-IS&T · CCC code: 0277-786X/09/\$18 · doi: 10.1117/12.808750

gathered, and ways of using this knowledge in automatic gesture processing must be developed. This not only could solve problems such as the one described above, it may also benefit gesture processing in other ways, such as reducing computational load and removing the need for large training sets.

Many gesture processing techniques are first tried out in the field of sign language recognition, since sign language signs are gestures that are both complex and well-defined. Automatic sign language recognition (ASLR) is also an interesting field in its own right, since it can help create better communication aids for the deaf and interactive tools for sign language learners. The research described in this paper is all done on Sign Language of the Netherlands (SLN). However, many of the principles can be extended to gestures in general. Current ASLR, like other gesture processing areas, mostly focuses on problems of information acquisition (how to obtain information from image streams or instrumented gloves) and data-driven classification (which method to use, how to construct training sets, etc.). However, for successful processing of natural sign language, other issues must also be addressed, such as: which elements of a sign are informative? Are all parts of a sign equally important, and if not, where should we focus our attention and resources? And how much variation is allowed in the execution of a sign? This paper discusses the impact of such issues on ASLR. A set of previously performed sign language perception experiments is presented, their results are compared and several practical applications are proposed. The impact of the insights gained on sign language processing theory is also discussed.

1.1 Related Work

ASLR research is comparatively young: most research has been done since the 1990s. The majority of these studies either give no theoretical background to motivate their choice of features to represent a sign, or limit themselves to mentioning that handshape, hand orientation, hand motion and hand location determine the meaning of sign language signs and then extract (a subset of) features representing these qualities.^{1,2} Studies using instrumented gloves to collect data (e.g.^{3,4}) mostly use all sensors available. Studies using computer vision often limit the sign representation to location and motion plus some basic shape characteristics, because handshape and hand orientation are hard to extract (e.g.^{5,6}). Or they choose to use rotation- and scale invariant features.⁷ Lee et al.³ model handshape, hand orientation and motion separately, but do not describe how they arrive at values for these elements. The same is true for Bowden et al.,² who categorize motion, handshape, hand configuration and location into distinct categories, but do not describe how they arrive at the categories. Derpanis et al.^{8,9} and Vogler and Metaxas^{10,11} are among the few who give a motivation for their sign representations, which (partly) reflect existing sign phonological models. On the whole, little use is made of theoretical knowledge in the field of ASLR. An extensive overview of ASLR research is given in.¹²

2. INFORMATION DISTRIBUTION IN THE TIME DOMAIN

Three experiments were performed to gain insight into the distribution of information over a sign in time. A sign (or gesture) can be divided into certain phases in time. The three main phases are (in chronological order) preparation, stroke, and retraction.^{13,14} These terms originated in the study of co-speech gestures, but are used for sign language signs as well. In figure 1 the phases are shown schematically in white, gray and black. Preparation is the phase in which the hands are moved to the location and configuration necessary to perform the intended action (e.g. if the intended action is at eye-height and the hands are at the table, the hands must first be lifted). The stroke is the phase containing the desired action, and the retraction contains the relaxation back to the rest position or to the next movement. In fluent signing, the retraction of a sign can merge with the preparation of the next sign. In the experiments described here, however, all stimuli were isolated signs starting from and ending in a certain rest position, so all three phases were always present. In the experiments it was investigated which moments in a sign facilitate human sign detection and sign recognition. The first two experiments investigated how early in a sign recognition can take place, to learn how much of a sign is actually needed. The third experiment investigated the information content of sign phases in isolation.

2.1 Experiment 1: Detecting a Sign

This experiment was set up to investigate at which moment of a movement signers realize that it is a sign, and not a different movement. Participants saw a movie and were asked to press a button as soon as they saw a sign beginning. The movies contained either a sign, or a non-communicative movement (a fidget) followed by a sign,

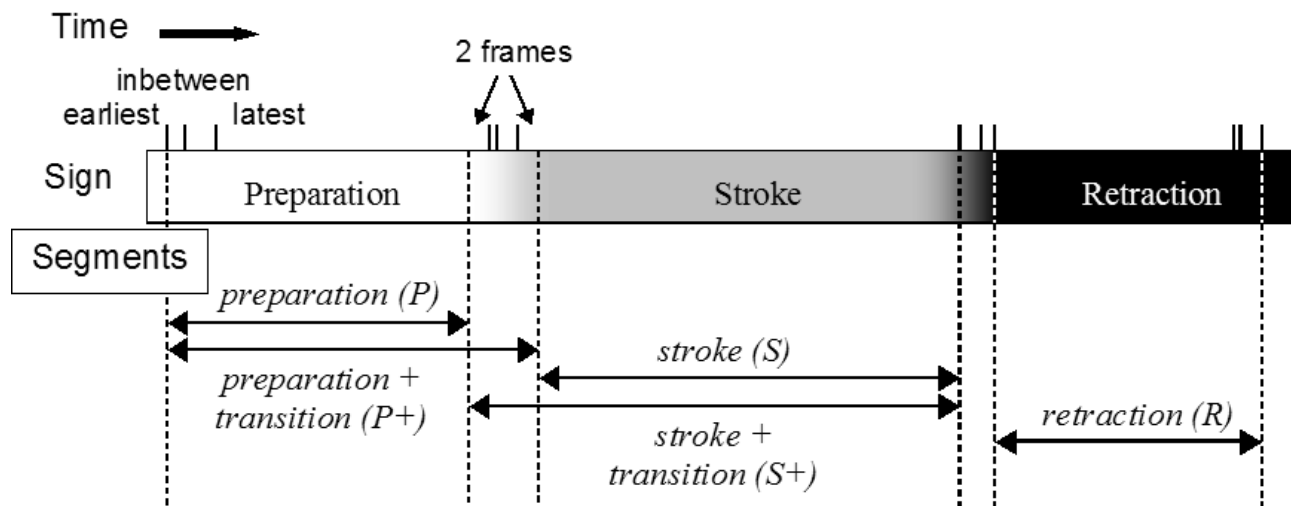


Figure 1: Structure of a sign in time. The bar represents the time course of a sign, its main phases are shown in white, gray and black. The vertical lines above the bar indicate the phase border determinations of three independent annotators. Beneath the bar is shown how segments were taken from a sign in experiment 3. Special attention was paid to the transition from preparation to stroke, since previous research suggests this is a salient region. The transition was defined as the interval between the earliest and the last border placement of the annotators, plus 2 frames on each side. Both preparation and stroke were tested with and without the transition. This gives a total of five segments.

or two fidgets. The fidgets were distractors that made sure participants could not simply react to any movement. Thirty-two signs and nine fidgets were used to make one hundred and twelve stimuli. Sixteen signers and eight non-signers participated in the experiment. A more extensive description can be found in.¹⁵

2.2 Experiment 2: Recognizing a Sign

In this experiment, the same setup and stimuli were used as in experiment 1, but this time the participants were asked to press the button as soon as they recognized a sign, and were then asked to identify the sign. The purpose was to determine at which moment in a sign signers can identify it. Thirty-two signers participated (no non-signers were included since they would not be capable of identifying a sign). See¹⁶ for more details.

2.3 Experiment 3: Recognition of Isolated Sign Phases

This experiment was set up to investigate the information content of the different phases of a sign. In the experiment signers were asked to view isolated fragments of a sign movie and try to identify the sign. The fragments corresponded to the phases preparation, stroke and retraction. Because the boundaries between these sign phases are not sharp, three independent annotators were asked to determine them. Using their annotations (shown as vertical bars in figure 1), the segments preparation, stroke, and retraction were extracted from sign movies as depicted in figure 1. Because previous research indicated that the transition between preparation and stroke may be especially salient,^{15,17} both preparation and stroke were extracted with and without this region included. The difference between the scores of these pairs would reveal the information content of the transition region. All segments were extracted from twenty different signs (a subset of the material used in experiments 1 and 2), and twenty-seven signers participated in the experiment. See¹⁸ for more details.

2.4 Results

Here, the combined information gathered in experiments 1–3 for the overlapping stimuli (fifty-two stimuli from twenty different signs) is discussed. Figure 2 shows the time lines of ten example stimuli as horizontal bars. Descriptions of the signs and details on phase border determination can be found in.¹⁸ The vertical black ticks indicate the phase borders (with transition areas between the end of the preparation and the start of the stroke,

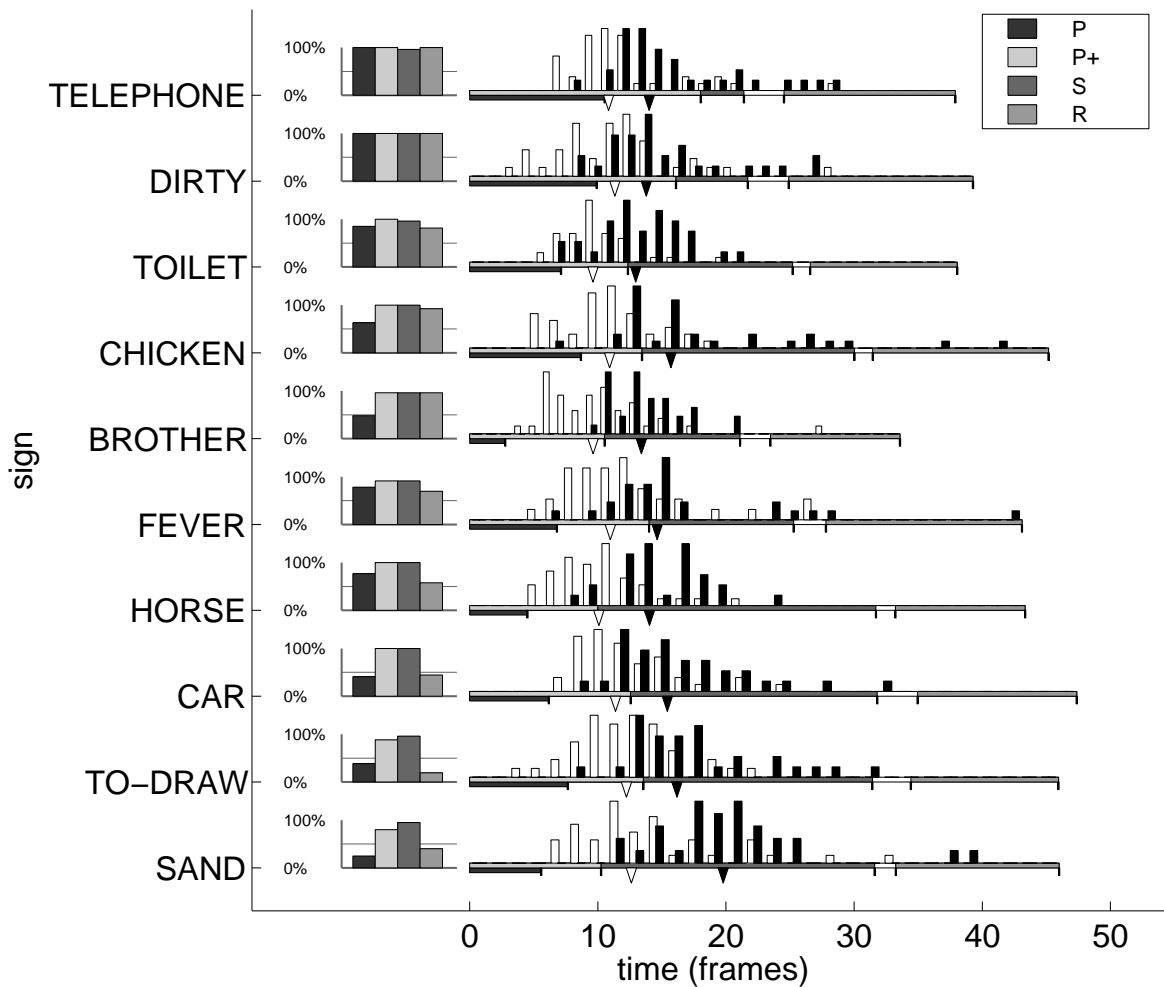


Figure 2: Overview of recognition scores and detection and recognition times for ten signs (experiments 1–3). The horizontal bars depict the time lines of the signs. The vertical black ticks indicate the phases borders (with transition areas between the end of the preparation and the start of the stroke, and between the end of the stroke and the start of the retraction). The four main phases are indicated by shades of gray. The segment P is put below the segment P+ because they partly overlap. On top of the time line, normalized histograms of the detection and recognition times from experiments 1 and 2 are shown in white and black respectively. Below the time line, white and black triangles indicate the median detection and recognition time respectively. The recognition percentages of the phases in experiment 3 are shown as bar graphs to the left of each time line. The segment S+ is not shown, since it gave results identical to those of S.

and between the end of the stroke and the start of the retraction). On top of the time line, histograms of the detection and recognition times from experiments 1 and 2 are shown in white and black respectively. The histograms were calculated by dividing each sign into a constant number of bins, regardless of sign length. Each histogram was normalized separately (divided by the maximum), so that all values lie between zero and one. The triangular marks below the time line indicate the median detection and recognition times, also in white and black respectively. The four bars to the left of the time line show the recognition percentages of the four fragments from experiment 3. A t-test confirmed that detection takes place earlier than recognition. Averaged

over all overlapping stimuli, detection takes place after 28% of the entire sign, recognition after 34%. Both detection and recognition were positively correlated to both the start of the transition and the start of the stroke (determined as shown in figure 1). Detection happened on average after 1% of the stroke and recognition after 18%. Apparently, detection and recognition are possible on the basis of just the first part of the sign — not even the entire stroke-phase is necessary. The results of experiment 3 were in accordance with these findings. The stroke in isolation gives recognition scores not significantly different from 100%, and adding the transition to the preparation increases recognition scores to the level of the stroke, confirming that the first part of the sign contains enough useful information to make recognition possible. There was no correlation between the length of the preparation (without transition) and the recognition times, nor between the recognition score of the preparation and the recognition times. Preparation length and preparation recognition score had no correlation with detection times either. The graphs in figure 2 show that there can be large differences between individual signs regarding the usefulness of the information contained in preparation, retraction and transition. The detection and recognition times of individual signs show less variation.

3. INFORMATIVE ELEMENTS OF SIGNS

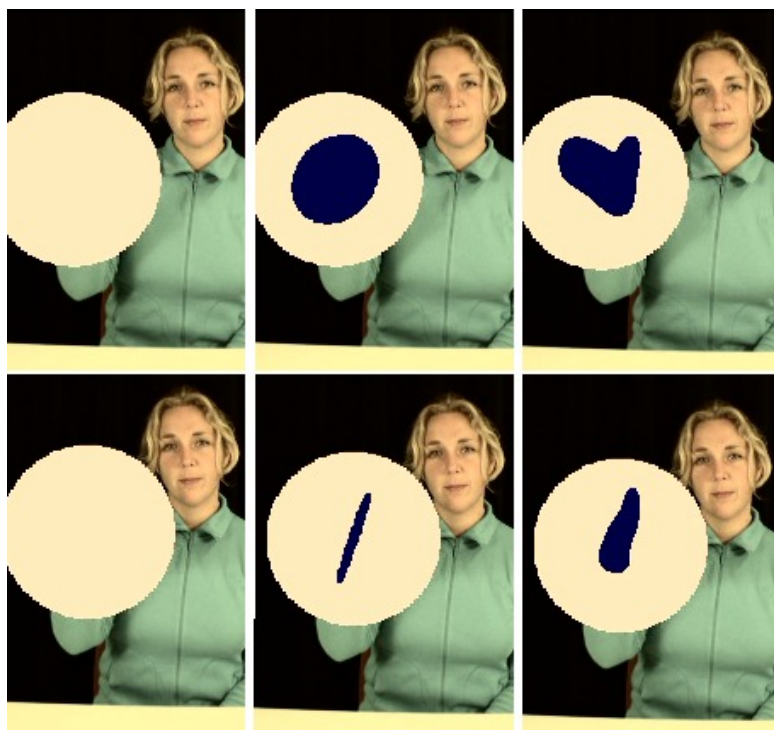
Signs can not only be subdivided in time, they can also be studied in terms of their informative elements. Signs are linguistically equivalent with words, and like words, they consist of smaller elements that determine their meaning. In words, these elements are called *phonemes*. In sign languages, the corresponding elements are not clearly described yet. There is considerable discussion on which elements would play the role of sign language phonemes, and what their organization should be.¹⁹ However, it is generally assumed that hand location, hand orientation, hand motion and handshape are important for the meaning of a sign, since pairs of signs (*minimal pairs*) can be found that differ only in one of these elements, and yet have different meaning. Other qualities of signs, such as speed of movement or scale, are generally not considered important for determining meaning. It is not clear, however, if each of these informative elements is equally salient. Nor is it clear how much variation is allowed within the elements. For example, if a position value is described as “in front of the eyes”, how large is this area exactly? Is the allowed variation the same for all position values, or are some values vaguer or more ‘lenient’? And how is this for the other elements?

The experiments described below were performed to increase knowledge on these matters. In the first, handshape and hand orientation were separated using image processing techniques, to investigate their salience for sign recognition. The second experiment investigated how much the informative elements of a sign can be changed before it becomes unacceptable as an example of that sign.

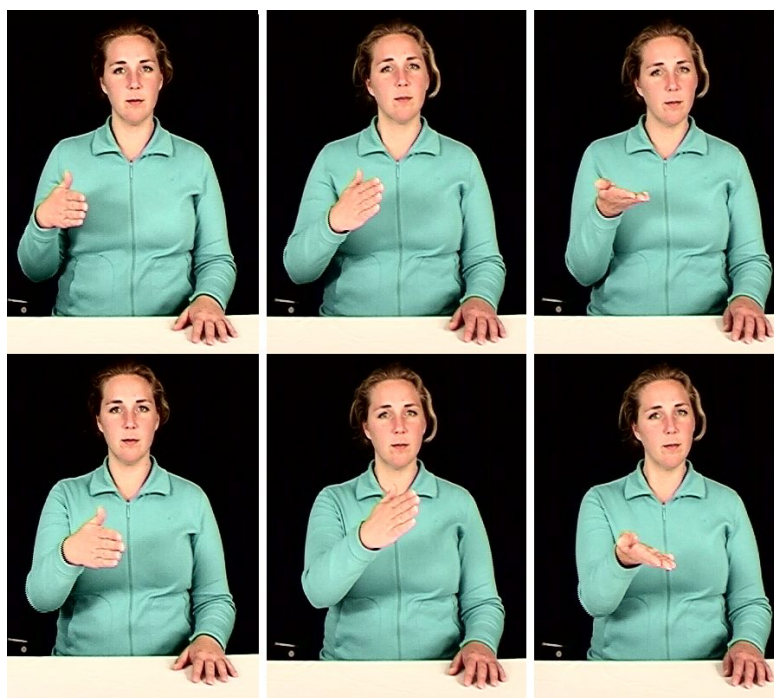
3.1 Experiment 4: Importance of Hand Configuration for Recognition

Signers were asked to view sign movies in which the hand configuration was obscured to varying degrees. The term “hand configuration” is used for the combination of handshape and hand orientation. They were then asked to try to identify the sign. In condition 1 (C1) the hand was entirely covered, so that only location and motion could be observed. In condition 2 (C2), a representation of palm orientation was shown on top of the covered hand. In condition 3 (C3), a crude version of the handshape silhouette was shown on top of the covered hand. The top row of figure 3 shows an example of a stimulus sign in conditions 1–3. Condition 4 (C4) was unmodified and used as a control condition. Twenty signs were used, in four conditions, giving a total of eighty stimuli. Participants saw each stimulus 5 times and scored between 0 and 5, depending on how many showings they identified correctly. Twenty-seven signers participated. The experiment was described more extensively in.²⁰

The results are shown in figure 4. In C1 most signs cannot be recognized. C2 hardly differs from C1. In C3, scores vary greatly, both between signs and within a sign (between participants). An analysis of variance confirmed that there was no difference in score between C1 and C2, and that C3 gave higher scores, though still not as high as in the unmodified condition (not shown). We conclude that location and motion alone (C1) are in general not enough for successful sign recognition. Adding hand orientation information (C2) does not improve recognition scores, but adding crude handshape information (C3) does, although the amount with which the scores improve varies between signs. A few signs can be recognized well even in C1 and C2.



(a)



(b)

Figure 3: (a) Example of stimuli in experiment 4. Two frames of the sign FLAG are shown in conditions 1 (left column), 2 (center column) and 3 (right column). (b) Example of stimuli in experiment 5. Two frames of the sign SAW are shown: in the 'normal' form (left column), with modified movement direction (center column) and with modified hand orientation (right column).

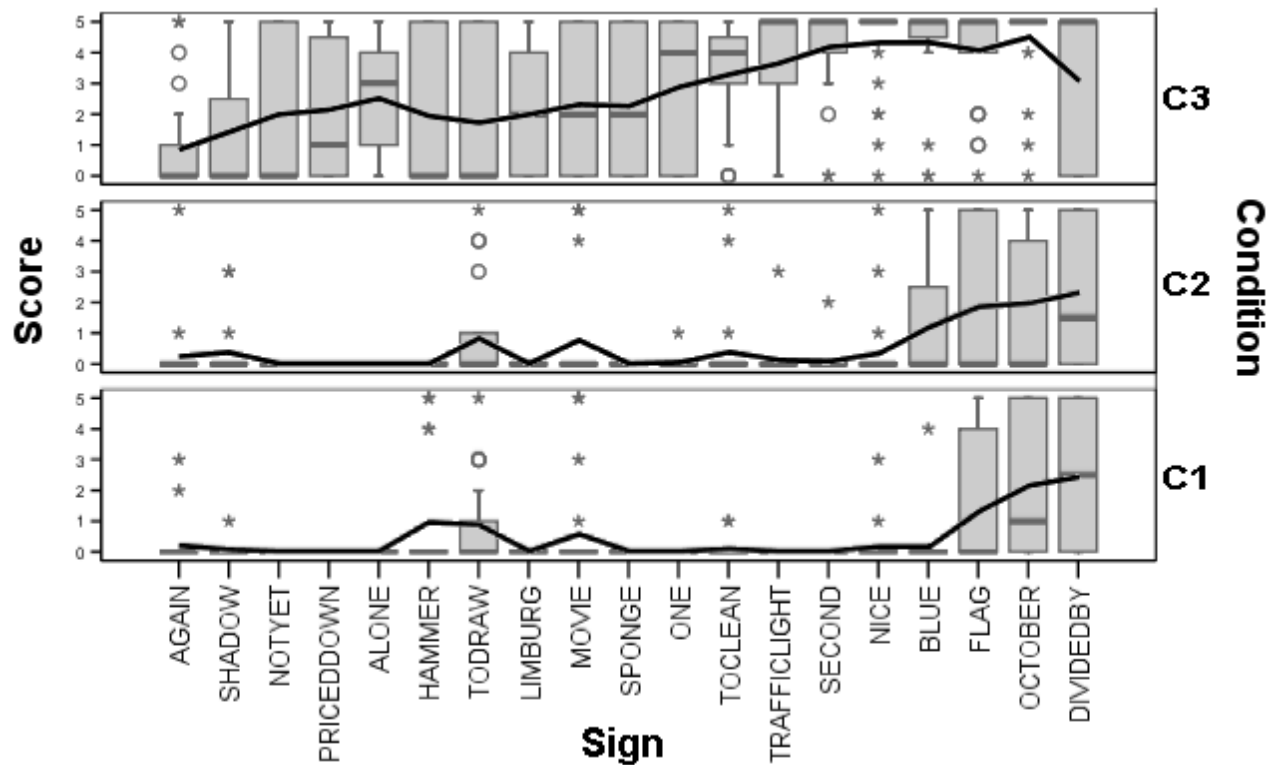


Figure 4: Results of experiment 4. Box plots of recognition scores (0–5) per sign in conditions 1 (no hand configuration), 2 (hand orientation visible) and 3 (crude handshape visible). The black line indicates the mean score, the thick gray bar within the box is the median. Circles and asterisks are outliers and extreme values respectively.

3.2 Experiment 5: Allowed Variation within Elements

This experiment was designed to learn more about which variations are acceptable in signs. To this end, four signs were recorded in a number of different variations. Variations included changing the direction of movement, the scale of movement, the speed of movement, the handshape, the hand orientation, etc. An extensive description can be found in.²¹ There were one hundred and thirty-one stimuli in total. Figure 3 shows an example. Signers were asked to view the movies and indicate on a five point scale how acceptable the sign was. Twenty-six signers participated.

The results show that participants vary widely in the strictness of their judgments, with some participants rejecting far more variants than others. However, the *order* of acceptability for the movies tends to be highly correlated between participants (using Kendall's tau). Closer inspection shows that acceptability ratings also vary according to the type of variation, as shown in figure 5. Temporal variations (changing the speed of the movement or inserting holds) appear to be acceptable. Spatial variations, on the other hand (e.g. changing the location) are generally unacceptable. Variations combining the two are only slightly more acceptable than spatial variations. Apparently, temporal variations do not affect acceptability much, whereas spatial variations quickly cause a sign to be considered unacceptable. However, iconicity is a complicating factor here. All four test signs contained iconicity, i.e. they mimicked the object or action they represented. Some spatial variations damaged the iconicity, and it may be for this reason that those variations were unacceptable. Temporal variations hardly ever damage iconicity. Also, because of iconicity some variations were acceptable for certain signs but not for others. For example, tilting the hand would be acceptable in the sign 'to saw', but not for the sign 'to carry a tray'. In the first case, it is a conceivable modification of the action the sign depicts: sawing at an angle. In the

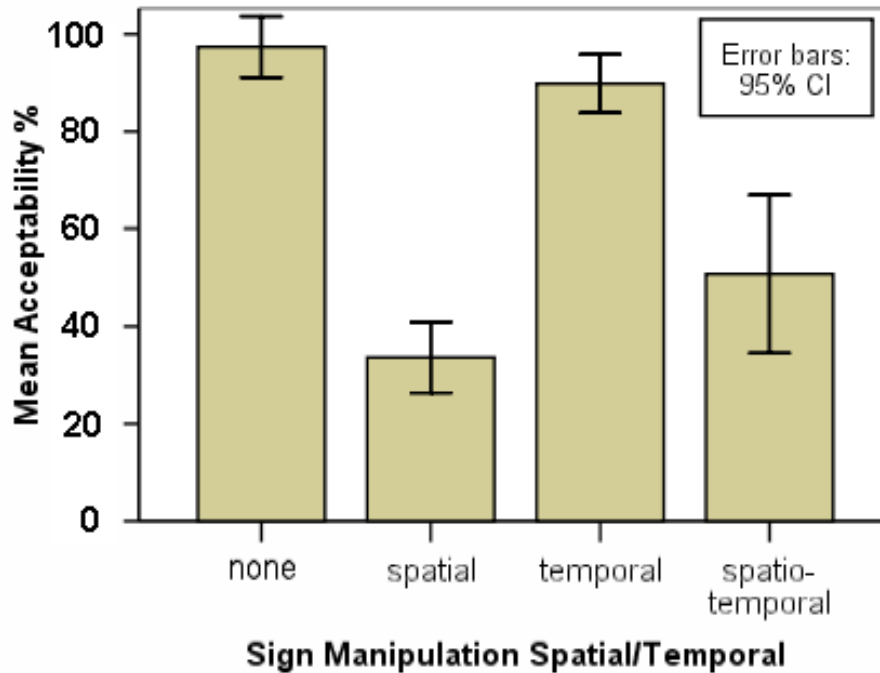


Figure 5: Results of experiment 5. The bars show acceptability of different types of variation in a sign. Temporal means that the timing of elements was changed, spatial means changes in location, direction, etc. Spatio-temporal means variations that combine the two types.

second case, it is not: tilting the tray would cause the contents to fall off. Through iconicity, the meaning of a sign has an influence on which variations are acceptable. This complicates the formational rules for signs.

3.2.1 Human and Machine Judgment Compared

The human judgments from experiment 5 were compared to the judgments of three different ASLR algorithms.²² These algorithms were trained to assess the correctness of a sign. They were trained on 120 signs, executed in regular form, from 75 different signers. The algorithms learned the definition of one sign by comparing the 75 examples of the sign to the 119 x 75 other examples. In this way, they learned which feature values represented 'normal' variation of the target sign, and which indicated a truly different sign. The algorithms were tested on a subset of the sign variants created for experiment 5: only three of the four signs were used, and all variants that were invisible to the machines were removed (such as handshape variations, which would 'look' identical to the machines because they did not extract detailed handshape information). The correctness values given to these sign variants by an algorithm were regarded as machine acceptability ratings. The variants were ranked according to these ratings, and the resulting rankings were compared to the rankings human signers had made in experiment 5 (see²³ for more details). Somewhat surprisingly, the rank correlation (Kendall's tau) between the machine and human rankings was low (average 0.30). In contrast, the correlation among human rankings was generally high (average 0.68), as was the correlation among the three machine rankings (average 0.61). Apparently, training on a set of 'normal' signs does not enable a machine algorithm to extract an accurate (human) model of what a 'correct' sign is. Also, the ASLR method that gave the highest average correlation was not the one with the best classification performance. This means that classification performance on regular signs is apparently not a reliable predictor for success in acceptability assessment of deviant sign forms. Table 1 shows the human-machine correlations per individual sign. The degree of correlation seems to be related to both the algorithm and the sign it is used on, suggesting that there is not one algorithm that always gives the most human-like ratings, but that this differs per sign.

Table 1: Correlations between acceptability judgments of three ASLR methods with the average human acceptability ratings, for the signs CURTAIN, OVEN and (HAND)SAW. The rightmost column gives the average correlation for the three signs. For comparison, the bottom row shows the correlation of human acceptability judgments with the average human judgment (average of the correlations of individual humans with the average human). It is clear that the human-human correlation is higher than the human-machine correlation for all three signs. The human-machine correlations vary both with the type of ASLR algorithm and with the specific sign involved. This may be because some algorithms are weaker in recognizing certain types of sign than others. But more complex factors, such as iconicity, may also play a part in explaining the lack of correlation for some algorithms and signs (e.g. CDFD on sign OVEN).

| Recognition Method | CURTAIN | OVEN | SAW | Average of three signs |
|--------------------|---------|------|------|------------------------|
| SDTW | 0.49 | 0.10 | 0.34 | 0.31 |
| CDFD | 0.40 | 0.02 | 0.41 | 0.28 |
| QDFFM | 0.33 | 0.14 | 0.43 | 0.30 |
| Humans | 0.68 | 0.67 | 0.68 | 0.68 |

4. DISCUSSION

4.1 Practical Application

The knowledge derived from the presented experiments can be given practical application in ASLR or gesture processing. For example, the time domain experiments reveal that it is not necessary to process the entire sign. Processing only the first half, or only the stroke phase, would give enough information to identify a sign. The discarded parts would still contain useful, if redundant information (segments P and R do enable some recognition), and using more information is often desirable. However, this should be weighed against the increase in computational costs and the extra variation that such parts can contain. For example, preparation and retraction of a sign can be very different for the same sign depending on where the hands are located before the sign and have to go afterward. Such variable phases are hard to model statistically, and in the end they would probably prove unhelpful for classifying the sign. In which case it would be better to disregard them from the start.

Focusing on certain sign phases does of course presuppose the ability to detect such phases automatically in the data stream. Unfortunately, automatic detection of movement phases is not a trivial matter.²⁴ A practical solution would be to annotate the phase borders of a single example per sign by hand, and to find the corresponding phase borders in other examples of the same sign through a synchronizing algorithm such as dynamic time warping.²⁵

In the domain of the informative elements both experiments have a fairly direct application. Experiment 4 showed that handshape is more salient for sign recognition than hand orientation. So if an algorithm using only the location and motion elements (most current algorithms) is to be extended, it would seem most beneficial to add the handshape element to the representation. Although in practice, detecting the hand orientation may be necessary in order to correctly classify a handshape. The participants in experiment 4 may also have derived some notion of orientation in condition 3, even though they only saw the handshape silhouette. Experiment 5 yielded that temporal variations in signs are relatively harmless, whereas spatial variations more quickly cause a sign to become unacceptable. This shows us that in ASLR, modeling temporal patterns is probably not useful. Such patterns can still be *typical*, of course. However, they are apparently not *necessary*: a sign example not following the pattern is still a valid version of the sign, as demonstrated in experiment 5, so an ASLR algorithm should not reject it. This is an example of how perceptual knowledge can be used for the correct handling of atypical examples (outliers), something which is hard to achieve for a data-driven method.

Experiment 5 and the human-machine comparison described in section 3.2.1 also yielded insights that are not immediately applicable, but nevertheless important for ASLR. One striking matter is the fact that determining the “ground truth” for ASLR (which variants should the algorithm recognize as a certain sign?) is not trivial: human observers vary in their determination of what can still be called a ‘correct’ version of a sign, with some being more strict, others more lenient. Therefore, choosing a strictness level seems necessary before variants can be judged as ‘correct’ or ‘incorrect’. Furthermore, the amount of variation that is allowed for a certain sign

element appears to depend on a number of factors. The type of variation is important (temporal vs. spatial), and possibly even the exact nature of the element. Then there is the presence of iconicity: if a variation damages the iconicity, it immediately seems to become unacceptable. Knowledge on these factors could enable us to construct rules for allowed variation in every situation. However, to do this, more research is necessary. Current ASLR methods try to learn such rules from training data, but do not arrive at the same criteria as human beings. Training the methods on more deviating sign variants may improve their performance. However, constructing such variants and determining a ground truth (which variants are ‘correct’) remains a problem.

4.2 Implications for Sign Recognition Theory

In experiments 2 through 4 the recognition of sign language signs under difficult circumstances was investigated — information was sparse or the response had to be as quick as possible. The results differ for individual signs, with some signs being recognized earlier (experiment 2), or yielding higher scores in the information-sparse conditions (i.e. segments P and R in experiment 3, and conditions 1 and 2 in experiment 4). No single sign characteristic explains these differences, i.e. there is no significant correlation between better score and rarity of handshape, better score and presence of a path movement, etc. However, in the experiments with sparse information, the well-recognized signs tend to contain more ‘restrictive’ elements than the poorly-recognized signs. We define restrictive elements as elements that only occur in a relatively small group of signs. For example: a rare handshape that only occurs in a small number of signs is a restrictive element. Presence of a path movement is also a restrictive element: the group of signs with no path movement is larger than any group of signs sharing a specific path movement. The presence of such elements is beneficial for recognition, because they limit the set of possible signs (assuming a cohort model of sign recognition²⁶). The following elements seem to function as restrictive elements: rare handshapes, path movements, a change in handshape or hand orientation, and repetition. Well-recognized signs appear to contain more of these elements than poorly-recognized signs. However, the large number of factors and small number of signs in the dataset makes it impossible to test this statistically. More research with a larger sign set, designed to separate out these factors, is necessary.

In the experiments that focus on a quick response, an analysis of variance showed that the presence of a rare handshape causes quicker recognition. The other restrictive elements listed above give no effect, except for the presence of a hand orientation change which actually causes recognition to be slower. This may be due to the difference in tasks: when timing is an issue, a stimulus that is high in information-density may cause a delay, because more information needs to be processed. When timing is not important but information is sparse or obscured, high density of information becomes a benefit, because a small part of the stimulus may still be very informative. Another possibility is that recognizing patterns for which several frames of a movie are necessary (such as path movements and handshape/hand orientation changes) takes time, whereas spatial patterns (such as handshapes and locations) can be observed at a glance. This would explain why the restrictive elements that are patterns in time do not cause quicker recognition in the experiments where a fast response is needed, while one of the spatial patterns does have an effect. Once again, more data are needed to test the influence of various elements on recognition performance.

Studying the data from several perceptual experiments has given us more knowledge on the nature of sign language recognition. We have gained a fairly good insight into the importance of different time phases within a sign, with the stroke being the most important phase and the beginning of the stroke a very salient moment. The matter of informative elements within a sign remains more difficult, mainly because of the large number of factors that appears to play a role. Some straightforward conclusions can be drawn about the salience of the handshape element and the relative insignificance of timing within a sign. However, to really gain insight into allowed variation, more research is necessary. Since this is an area in which ASLR can profit substantially, solving the problem of intra- and inter-signer variation, it would certainly be worthwhile.

ACKNOWLEDGMENTS

The research discussed in this paper was made possible by the Delft ICT Research Centre and the VSB-fund, grant number 2003451. The authors would also like to thank the Dutch Foundation for the Deaf and Hard-of-hearing Child and the Royal Auris Group, who were partners in this project.

REFERENCES

- [1] Bahrami, S. and Vannobel, J. M., "A parallel and hierarchic framework to visual gesture recognition," paper presented at GW2001 (April 18-20 2001).
- [2] Bowden, R., Windridge, D., Kadir, T., Zisserman, A., and Brady, M., "A linguistic feature vector for the visual interpretation of sign language," in [8th ECCV], *LNCS* **3021**, 390–401, Springer, Heidelberg (May 11-14 2004).
- [3] Lee, C.-S., Bien, Z., Park, G.-T., Jang, W., Kim, J.-S., and Kim, S.-K., "Real-time recognition system of korean sign language based on elementary components," *6th IEEE International Conference on Fuzzy Systems* **3**, 1463–1468 (Jul 1997).
- [4] Wang, C., Gao, W., and Ma, J., "A real-time large vocabulary recognition system for chinese sign language," in [GW2001], *LNCS* **2298**, Springer, Heidelberg (April 18-20 2002).
- [5] Starner, T., Weaver, J., and Pentland, A., "Real-time american sign language recognition using desk and wearable computer based video," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **20**(12), 1271–1375 (1998).
- [6] Zieren, J. and Kraiss, K.-F., "Robust person-independent visual sign language recognition," in [IbPRIA], *LNCS* **3523**, 520–528, Springer, Heidelberg (June 5-7 2005).
- [7] Holden, E.-J., Lee, G., and Owens, R., "Australian sign language recognition," *Machine Vision and Applications* **16**, 312–320 (December 2005).
- [8] Derpanis, K. G., Wildes, R. P., and Tsotsos, J. K., "Hand gesture recognition within a linguistics-based framework," in [8th ECCV], *LNCS* **3021**, 282–296, Springer, Heidelberg (2004).
- [9] Derpanis, K. G., Wildes, R. P., and Tsotsos, J. K., "Definition and recovery of kinematic features for recognition of american sign language movements," *Image and Vision Computing* **26**, 1650–1662 (2008).
- [10] Vogler, C. and Metaxas, D., "Parallel hidden markov models for americal sign language recognition," in [7th ICCV], 116–122 (September 20-27 1999).
- [11] Vogler, C. and Metaxas, D., "Handshapes and movements: Multiple-channel american sign language recognition," in [GW2003], *LNCS* **2915**, 247–258 (April 15-17 2004).
- [12] Ong, S. C. and Ranganath, S., "Automatic sign language analysis: A survey and the future beyond lexical meaning," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **27**(6), 873–891 (2005).
- [13] Kita, S., van Gijn, I., and Van der Hulst, H., "Movement phases in signs and co-speech gestures, and their transcription by human coders," in [GW1997], Wachsmuth, I. and Frolich, M., eds., *LNCS* **1371**, 23–35, Springer, Heidelberg (September 17-19 1998).
- [14] McNeill, D., [Hand and Mind], University of Chicago Press, Chicago (1992).
- [15] Arendsen, J., van Doorn, A. J., and de Ridder, H., "When and how well do people see the onset of gestures?," *Gesture* **7**(3), 305–342 (2007).
- [16] Arendsen, J., Koenderink-van Doorn, A. J., and de Ridder, H., "When do people start to recognize signs?," Submitted to *Gesture*.
- [17] Emmorey, K. and Corina, D., "Lexical recognition in sign language - effects of phonetic structure and morphology," *Perceptual and Motor Skills* **71**(3), 1227–1252 (1990).
- [18] ten Holt, G. A., van Doorn, A. J., de Ridder, H., Reinders, M. J., and Hendriks, E. A., "Which fragments of a sign enable its recognition?," *Sign Language Studies* **9**(2) (2009).
- [19] Emmorey, K., [Language, Cognition and the Brain: Insights from Sign Language Research], Lawrence Erlbaum Associates, Mahwah, NJ (2002).
- [20] ten Holt, G. A., van Doorn, A. J., de Ridder, H., Reinders, M. J. T., and Hendriks, E. A., "Signs in which handshape and hand orientation are not or only partially available: what is the consequence for lexical recognition?," *Sign Language Studies* (Accepted).
- [21] Arendsen, J., van Doorn, A. J., and de Ridder, H., "Accpetability in signs," Unpublished.
- [22] Lichtenauer, J. F., Hendriks, E. A., and Reinders, M. J., "Sign language recognition by combining statistical dtw and independent classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **30**(11), 2040–2046 (2008).

- [23] Arendsen, J., Lichtenauer, J. F., ten Holt, G. A., van Doorn, A. J., Hendriks, E. A., Spaai, G. W., Reinders, M. J., and de Ridder, H., "Acceptability ratings by humans and automatic gesture recognition for variations in sign productions," in *[FG'08]*, IEEE (2008).
- [24] Wilson, A. D., Bobick, A. F., and Cassell, J., "Recovering the temporal structure of natural gesture," in *[FG'96]*, 66–71, IEEE (1996).
- [25] Kruskal, J. and Liberman, M., "The symmetric time warping problem: From continuous to discrete," in *[Time Warps, String Edits and Macromolecules: The Theory and Practice of Sequence Comparison]*, Sankof, D. and Kruskal, J. B., eds., 125–161, Addison-Wesley (1983).
- [26] Marslen-Wilson, W. D., "Functional parallelism in spoken word recognition," *Cognition* **25**, 71–102 (March 1987).