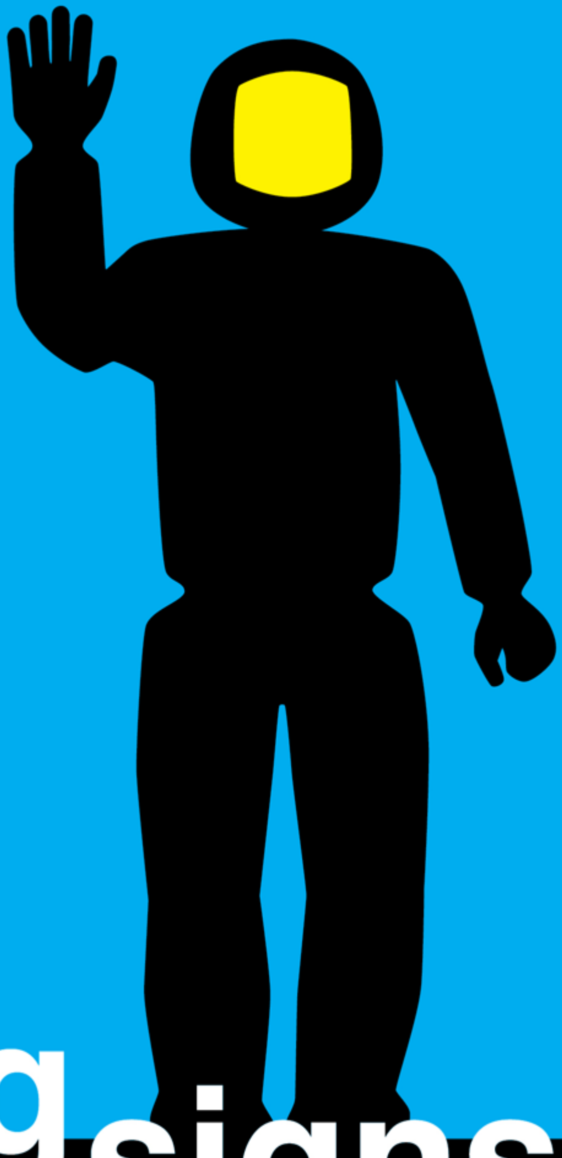


seeing signs

on the appearance of manual movements in gestures

jeroen arendsen

Seeing Signs

On the appearance of manual movements in gestures

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. dr. ir. J.T. Fokkema,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op maandag 19 oktober 2009 om 15.00 uur

door

Jeroen ARENDSEN

Ingenieur Industrieel Ontwerpen
geboren te Uden

Dit proefschrift is goedgekeurd door de promotor:

Prof. dr. H. de Ridder

Co-promotor:

Dr. A.J. van Doorn

Samenstelling promotiecommissie:

Rector Magnificus	voorzitter
Prof. dr. H. de Ridder	Technische Universiteit Delft, NL, promotor
Dr. A.J. van Doorn	Technische Universiteit Delft, NL, copromotor
Prof. dr. K. Emmorey	San Diego State University, USA
Prof. dr. ir. B.H. Eggen	Technische Universiteit Eindhoven, NL
Prof. dr. I. Heynderickx	Technische Universiteit Delft, NL
Dr. A. Kendon	University of Pennsylvania, USA
Dr. M. Güllberg	Max Planck Institute for Psycholinguistics, NL
Reservelid	
Prof. dr. ir. M.J.T. Reinders	Technische Universiteit Delft, NL,

Dit proefschrift is mede mogelijk gemaakt door financiering vanuit het ICT Delft Research Centre en door een bijdrage vanuit het VSB-fonds, beursnummer 2003451: project "Een elektronische leeromgeving voor jonge kinderen met auditieve en communicatieve beperkingen voor het leren van Nederlandse Gebarentaal (ELo)".

ISBN 978-90-9024630-7
Cover design: Elif Özcan Vieira, 2009
http://jeroenarendsen.nl/seeing-signs

Contents

Introduction

Computers and robots are starting to become 'aware' of our gestures. They have already been 'listening' to our speech for some time. Of course, so far, with the exception of some wild science fiction machines, they heed the commands we type and they go where we point them. But the dawn of a more social human-computer interaction (HCI) may well be at hand. Perhaps someday we may witness robots that walk amongst us, at our beck and call. Or is this only a dream? Are robotic waiters, snapping to attention at our merest glance and wave, only a distant, unlikely future? Can we, as Ray Kurzweil (1999) predicts, truly expect machines to be socially adept enough that we will want them to share our lives, to occupy our spaces, to watch us, and to talk to us?

On the one hand, the technological developments in gesture recognition are promising (see Mitra et al. 2007), and the introduction of two 'killer applications', the Wii and the iPhone, has done much to bring gesture recognition to the center of attention of the HCI community. On the other hand, gesture recognition technology still faces many problems and it is not very 'humane' yet. The Wii requires a Wii-mote (a remote with motion sensors), while the iPhone accepts only '2D gestures' on its small touch screen. And while both interfaces can be considered gestural, they are hardly big steps towards the above mentioned robot waiter. However, recognition of more natural gestures with cameras is developing rapidly. Already in different labs, unobtrusive, vision based gesture recognition is being combined with advanced speech recognition in multimodal interfaces, such as described by Sharma et al. (2003), and also in robots, as for example described by Stiefelhagen et al. (2007). How 'natural' or 'humane' such machines communicate will be a central question in future developments.

Humans communicate with other humans using verbal and nonverbal behaviour including speech and various kinds of gestures. In the course of our lives we learn to produce these communicative acts and to perceive them in others. Some people even become 'eloquent' communicators. In any case, becoming an active player in the grand symphony of communication that surrounds us is not a trivial achievement and takes little children years of their lives. Unfortunately, for people who are trying to program computers to recognize gestures, little of the knowledge required for eloquent communication is documented. Children are usually expected to just pick it up from the examples provided by others like parents and peers, and to some extent gesture perception may be facilitated by our biological nature (cf. Rizzolatti & Arbib, 1998). It is striking that certain aspects of the perception of gestures are almost universally treated as trivial matters while they are in fact big challenges for automatic gesture recognition. As an example, imagine you are in another country: you may sometimes have trouble understanding the meaning of a certain gesture, but is it not amazing that you are at least able to see when these strangers are gesturing to you and that you are often able to guess what they mean? Knowledge about the precise perceptual strategies that enable us to see gestures in the continuous stream of human behaviour, to segment them from each other and from the rest of the action, is still sparse. However, it is clear from our example that, since humans are carriers of the secrets of gesture perception, a substantial amount of research on gesturing should be devoted to human gesture perception. This will be the main focus of this dissertation.

1.1. Topic and aim of this dissertation

This dissertation presents the results of a series of studies on the appearance of manual movements in gestures. The main goal of this research is to increase our understanding of how humans perceive signs and other gestures. Since generated insights from human perception may aid the development of technology for recognizing gestures and sign language automatically with cameras and computers, a small part of the dissertation will be devoted to possible implications of our findings for automatic gesture recognition.

Studying human perception with the goal of improving machine 'perception' is not new, and can be placed in a long tradition of studying (human) nature to inspire new engineering solutions. An example that has many parallels with our work is the study of human speech perception to aid the development of automatic speech recognition. Speech recognizers have already benefited from knowledge of how humans perceive speech and further progress (there is still quite a performance gap between human speech perception and automatic speech recognition (Lippmann, 1997)) will undoubtedly also be inspired by insights into human speech perception (Dusan & Rabiner, 2005). That does not mean automatic speech or gesture recognizers must always mimic human perceptual strategies. Other strategies can also be applied successfully. However, we believe that automatic gesture recognition currently faces many performance problems that might be solved, or where some progress might be made, by considering human perceptual strategies. One example is the handling of unexpected 'meaningless' movements, which is a topic we shall return to in the studies. Humans seem to have little difficulty in ignoring the meaningless movements, whilst paying attention to meaningful gestures. Machines typically pay attention to all of the movements and have great difficulty in gracefully ignoring those actions that were not intended for them to react upon. Understanding how humans accomplish this 'detection of gestures' might help in engineering a robust automatic gesture detection method.

1.2. Gesture technology and user experience in HCI

The work in this dissertation has been done at the faculty of Industrial Design Engineering, where a user centered approach to human-computer interaction (HCI) is advocated. This includes the development of natural, multisensory interactive systems yielding pleasant, satisfying user experiences (see Schifferstein & Hekkert, 2008). Our contribution consists of building up knowledge about human gesture perception. This knowledge may be applied in the development of automatic gesture recognition systems that fit typical or natural human behaviour and capabilities.

Of course, humans will adapt to and use any technology in the field of gesture recognition if they feel they need it, even if it does not fit their capabilities or preferences well, but if that is the case they probably will be less satisfied. To return to the example given previously, if an application with automatic gesture recognition can gracefully ignore someone's fidgeting movements (e.g. rubbing his nose or wiping his lips) and attend his meaningful gestures then a user of that application can behave freely. If, on the other hand, said application also attends the fidgeting and, for example, responds with some error message, then people can be expected to try to suppress their fidgeting movements. Some people may be used to such a high level of control over their body motions, but others may be annoyed by the need to suppress part of their natural behaviour and experience it as a restraint on their physical freedom.

One example of an application of automatic gesture recognition that has played a role in shaping the research in this dissertation is ELo. ELo is an Electronic Learning environment for deaf and hearing impaired children to practice Sign Language of the Netherlands (SLN) signs (Spaai et al, 2008). The ELo application was developed in a joint project by the ICT Delft Research Centre (ICTDRC), a multidisciplinary group at the Delft University of Technology, the Dutch Foundation for the Deaf and Hard of Hearing Child (NSDSK), and the Royal AURIS group. The work done within the ICTDRC consisted primarily of developing the gesture (or sign) recognition algorithms (mainly by Jeroen Lichtenauer, developed further by Gineke ten Holt, both also PhD students involved in the ELo project) that were required for specific ELo functions (i.e. checking the sign productions of a practicing child). To aid this development, given that it entails new progress in the field of machine vision, the ICTDRC also defined a work package to study the human perception of signs and other gestures (this dissertation and work by Gineke ten Holt).

1.3. Research questions and methods

The research questions were to some extent inspired by issues raised during the ELo development, but predominantly they were chosen because they were considered of general importance to learn more about the perception of gestures. The questions should also be of interest to the wider community involved in developing gesture and sign recognition. The questions were:

1. How do we perform the temporal segmentation of manual movement? Which boundaries exist between (phases of) movements that are useful for gesture or sign recognition?
2. How do we discriminate (meaningful) gestures (or signs) from other human behaviour?
3. How much time does it take humans to detect the beginning of a sign?
4. How much time does it take humans to recognize the lexical meaning of a sign?
5. Given that there appears to be a high degree of variability in producing signs how do humans handle this variability, for example how acceptable are different types of variation?

Several experiments were performed addressing these questions. In each experiment, the same method was followed: movies of signs and other manual movements were shown to human observers who were given a task related to their perception of the movies. Thus, insight was gathered into human gesture perception and into perceptually important characteristics of signs. In most cases, the experimental designs included a large number of variables and/or variations because the work is exploratory in nature. The approach in analyzing or interpreting the data was to stay as close as possible to the data and observations.

1.4. Outline of this dissertation

Figure 1 shows the outline of this dissertation and suggests several reading paths. Chapters 2 through 5 contain reports of the experiments, including their own introductions and discussions. It is advised to read chapter 2 before reading either chapter 3 or 4, since these chapters contain experiments that followed logically from the experiment in chapter 2. Chapter 5 can be read separately. Chapter 6 contains a general discussion of the work and some considerations for the future. At the end of the dissertation there is also a summary (in English and Dutch).

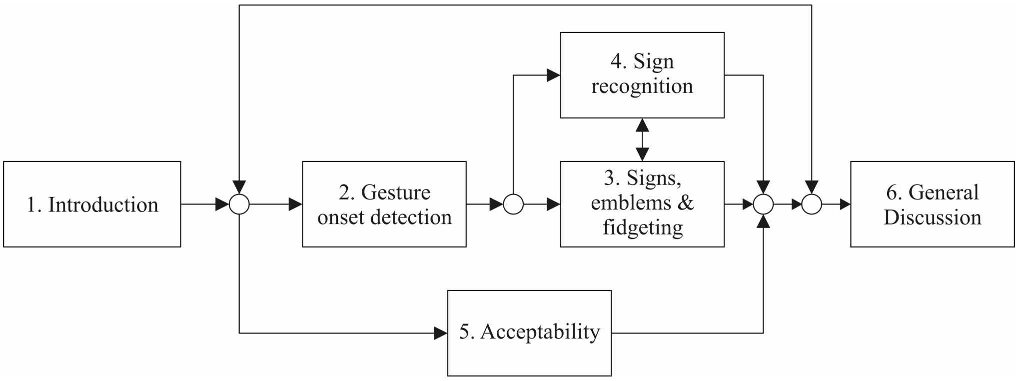


Figure 1. Outline of this dissertation. Each block represents a chapter and the arrows between them suggest different reading paths.

Chapter 2

♣

When and how well do people see the onset of gestures?

We studied if and when people detect the beginning of a gesture, in our case a sign in Sign Language of the Netherlands (SLN), by presenting movie fragments consisting of sequences of rest positions, fidgets, and signs to deaf signers, hearing signers and non-signers. Participants were instructed to respond as soon as they saw that a SLN sign had begun. All participants showed themselves highly capable of responding to sign beginnings. Signs that are two-handed, performed in signing space, have a highly marked hand shape, and contain path movement were discriminated best. Considering a sign as having a preparation, a stroke, and a recovery, response times showed strong clusters around 500 milliseconds after the beginning of sign preparation, or 200 ms after the onset of the stroke. The non-signers needed more time before responding; deaf signers took more time than hearing signers. Response time was influenced by three factors (shorter for signs that have a highly marked hand shape, are one-handed, and are preceded by fidgets). The results show that it is possible for people to discriminate fidgeting and signs based on appearance, even if one does not know sign language. No single feature of the movement appears necessary to detect the beginning of a sign. In most cases visual information available up to an early stage of the stroke is sufficient but in some cases the information in the preparation is enough.

This chapter appeared as:
Arendsen, Jeroen, Andrea J. van Doorn, & Huib de Ridder. (2007). When and how well do people see the onset of gestures? *Gesture, 7*(3), 305-342.

2.1. Introduction

This work is part of a project to develop video-based gesture recognition technology. We chose to focus on recognition of single signs from the Sign Language of the Netherlands (SLN) because they are gestures with more or less established criteria of well-formedness. Recognizing signs with a camera and a computer is not without challenges. However, some progress has been made (Starner et al., 1998; Zhang et al., 2004; Bowden et al., 2004; Zieren & Kraiss, 2005). Some of the many challenges are tackling movements of the hands not intended as a sign, allowing for variable rest positions, and providing real-time feedback. In this study we address these issues by studying how humans perform on these challenges, which may serve as a general benchmark for computer vision.

In most social contexts, the hands of a signer are also engaged in habitual touching of nose, chin, ear, or other body parts. The last category of habitual movements, called manipulators by Ekman (1999), is further referred to as "fidgeting" after Sacks and Schegloff (2002). People also use many different rest positions for their hands (Kita et al., 1998; Sacks & Schegloff, 2002). Under these noisy conditions we wish to automatically detect the beginning of a sign, to know quickly if a movement is a sign or a fidget, and to extract the relevant boundaries of the movement for further analysis. In our overall project, only the appearance of the manual movement (captured by camera) will be used. In accordance with this restriction, here people's perceptual abilities are investigated with material that excludes the non-manual component, discourse, or any other contextual factors (as far as possible).

In this research, we regard gestures as a broad category of movements that are all intentionally communicative. This includes signs in sign languages, emblems, pantomime, co-speech gestures of various types, and pointing (Kendon, 2004). Gestures are intended to communicate (Melinger & Levelt, 2004), and perceived as such (Kendon, 1994). Kendon (2004) emphasizes the importance of the appearance of a movement itself, rather than context, in discriminating gestures from other kinds of movements, such as practical actions, shifts in body position and fidgeting. He suggests that a movement perceived as a gesture has certain visible features and the more strongly it has these features the more likely it will be regarded by observers as gestural.

How people are able to see that a movement is intended to communicate (as a gesture) is a poorly studied problem. Nevertheless, humans appear to be able to do this very well. In his appendix on procedures McNeill (1992) remarks that the first step in video transcription for gesture studies is to identify the movements that are gestures, defined as all body movements except fidgeting. The inter-coder reliability of these identifications, determined in several studies, falls between 77% and 96% (McNeill, 1992). Clearly, coders are well able to distinguish between fidgeting and gestures, yet no explanation is given of how this is possible. Note that such coding is done with software that controls playing the video, rewinding it and even playing it in slow-motion. Whether people are able to perform as well in real time remains unclear, and is one of the questions that the experiment reported here aims to answer.

It is unclear which objective properties of (a movie of) a movement allow its categorization. Kendon (2004) suggests that certain forms and movement patterns lead to an immediate appearance of gesturalness. To back up this suggested "direct recognition" Kendon seeks support in two lines of work: one that outlines people's ability to attribute intentionality to movements (Michotte, 1962; Kiraly et al. 2003) and the other line showing people's ability to perceive biological motion when it is reduced to kinematics using point light displays (Johansson, 1973; Poizner et al., 1981). The only objective clues for gesture discrimination that Kendon offers himself are: a sharp boundary of onset and offset, and movement that is an excursion, going from and returning to a rest position (as opposed to a sustained change of position). In summary, there is some circumstantial evidence that gestural features exist, but we do not know exactly what they are, how many there are, which combination of them is sufficient, or if any of them is necessary.

The beginning of a sign or a gesture

Assume people can distinguish signs from fidgeting. We then want to know by what moment in the unfolding of the action they can do this. A movement has a beginning, an end, and perhaps several phases. As the movement progresses certain positions, shapes, kinematics, boundaries, etc. may appear that we perceive as gestural features. If we want to create automatic gesture recognition technology which must detect quickly that an observed movement is potentially the beginning of a sign, so that processing resources may be focused on it, we need to know what to look for. We need the cues that suffice to mark movement as potentially a sign in progress so that all available resources can be recruited to analyse it further (e.g. extract the meaning).

Kita et al. (1998) have analysed movement phases in signs and co-speech gestures. Their main contribution is a coding scheme for movement phases, now widely used in studies on gestures and signs, which builds on earlier work by Kendon (1980) and McNeill (1992). A "gesture unit" is defined to begin when the hands depart from a rest position (which can be highly variable, see also Sacks and Schegloff (2002)) and to end when the hands return to rest. Within this gesture unit one or more "gesture phrases" may occur (Kendon, 2004). A gesture phrase is comprised of a phase of movement called the "stroke", where there is a well-defined movement pattern in which the "effort" and "shape" of the movement[1] (see Dell (1977) for an explanation of these terms) are most clearly manifested and in which, in the case of manual expressions, the hand or hands assume, relatively, the most well-defined postures or shapes. Commonly a phase can be distinguished in which the hands are lifted or otherwise got ready for the performance of the stroke, a phase which is referred to as the "preparation". The gesture phrase also includes any phases when there is a pause in movement in which the articulators are held still in position, either after the stroke has been completed (a so-called "post-stroke hold") or before the stroke (a "pre-stroke hold"). Following the stroke, and hold that may ensue, the hand then relaxes and may return to a rest position. This phase of relaxation has been termed the "recovery". For Kendon (2004*)* the gesture phrase is comprised of the stroke and any preparation phase that precedes it, and any holds that may occur before or after the stroke, but it does not include the recovery phase, although this phase is part of the gesture unit within which the gesture phrase is contained.

In this paper we shall use the term "gesture" to mean what Kendon calls a gesture phrase and for us the phase of movement so nominated will also include any period of relaxation or recovery that follows. Kita et al. (1998) have provided criteria, which they claim are purely based on appearance, for identifying the boundaries of the entire movement unit, segmenting it into phases and identifying each phase type. For several studies they report a good reliability between coders on these tasks. Thus, not only can we see that a movement is intended to communicate (Kendon, 2004), we can also see where it begins and ends, and where the boundaries are between different phases (Kita et al., 1998), none of which requires knowing the meaning of the gesture or sign.

Kita et al. (1998) instructed coders to mark boundaries between movement phases when there is an "abrupt change of direction" and "discontinuity in the velocity profile". However, it remains unclear what exactly constitutes an abrupt change of direction: is it opposed to a gradual or to a regular change? It is also unclear what is meant with velocity profile. It appears that despite these weak definitions different coders using the scheme do agree on boundaries rather well (Kita et al., 1998). This suggests that the perception of boundaries is not a matter of defining what to look for, but simply of how humans perceive movement.

Marr and Vaina (1982) have provided a theory for the representation and recognition of the movement of shapes. Their state-motion-state (SMS) moving shape representation is designed to fit the requirements laid down by Marr and Nishihara (1978) for efficient visual recognition of 3-D objects. States are moments when parts of a shape are either absolutely or relatively at rest. The SMS representation, together with rotational movement as a primitive, and segmenting at discontinuities in velocity allow segmentation of a considerable range of movements (Marr & Vaina, 1982). Rubin and Richards (1985) define the boundaries of visual motion as starts, stops and force discontinuities. They have provided the mathematics that shows that their representation satisfies criteria of stability (robust for inconsequential variations) and invariance. They have also tied their work to psychophysical evidence and they state that there is a subjective motion boundary if and only if there is a theoretical motion boundary (Rubin & Richards, 1985).

Parish et al. (1990) used the ideas of Marr and Vaina (1982) and Rubin and Richards (1985) to compute "event boundaries" in American Sign Language (ASL) movies using the local minima of a motion index (moments of absolute or relative rest). For the purpose of video telephony they examined how well a sequence of frames can be represented by a subset of the frames, namely those at the boundaries. Newtson and Engquist (1976) had shown earlier that the boundaries are perceptually most salient. Parish et al. (1990) report that event boundaries can, to a certain extent, be defined computationally. Their findings suggest that their algorithm finds the boundaries between preparation, stroke, and recovery (or in their words raising the hands, moving them and reassuming the rest position). They found subsampling from such computed boundaries to be better than choosing at regular intervals to keep the ASL movie interpretable. Parish et al. (1990) thus showed that motion perception theory can productively be applied in the analysis of sign language. We have no reason to assume that SLN will be different than ASL in this respect, although the experiment should be replicated with SLN to be sure.

Emmorey and Corina (1990) studied the necessary time for lexical isolation and recognition of movies of isolated ASL signs which they found to be less than for spoken words. This difference is attributed to the phonetic structure of signs, where much information is available quite early. On average, location and orientation are identified about 150 ms after the beginning of the sign, followed shortly by handshape at 170 ms and finally the movement at 240 ms (Emmorey & Corina, 1990). But note how these times were calculated: A sign was defined to begin when the hand(s) entered "signing space". Their choice suggests that entering signing space is a visible event, yet it is unclear which boundary the hands must cross after departing from their rest position. Nevertheless, their findings indicate that several pieces of information become available in parallel in the early stages of a sign. Emmorey and Corina (1990) further found that native signers were able to recognize the lexical meaning of signs faster than late signers[2]. Such differences in experience may play a role in discriminating signs from fidgeting as well.

For the present study we formulated a sequence of goals: First, to verify that humans can discriminate the beginning of a sign from that of a fidget in real time (no slow-motion and no replay) without contextual clues (just the movement). If Kendon (2004) is right then this should be possible, even if subjects do not know what the signs mean. The second goal is to study how much time they need to make their decision and to relate that to the beginning of the sign. This will narrow the search for the movement features that people use. Thirdly, effects of experience will be checked by recruiting native signers, late signers and non-signers as participants. If there is little difference then we might assume signers and non-signers are using the same features. That, in turn, suggests that signs share these features with other gestures. Fourth, we will study to what extent the presence, absence, and nature of several phonological attributes (handshape, location, movement, etc) in the signs influence capability or response time. Some of these features may also serve as a cue for discriminating signs from fidgets. The fifth goal is to check whether conditions prior to the sign (preceding fidget, variable rest positions) hamper detecting its beginning.

The movement phases (preparation, stroke, etc) of the signs will be coded by three people. This will show us whether coders agree on boundaries, which would suggest that more or less objective criteria are commonly used. As such it may replicate (and add to) the findings of Kita et al. (1998). Furthermore, we can relate response times on the task of detecting the beginning of a sign to the coded boundaries (e.g. the onset of the stroke).

2.2. Experimental method

Participants

A total of 23 people participated in the experiment. They provided the following information: Age, gender, deaf or hearing, deaf or hearing parents, age of onset deafness, age of SLN acquisition, fluency in SLN, usage of SLN (primarily, daily, regularly or exceptionally). The recruiting goal was to find eight native signers, eight late signers, and eight non-signers. Native signers would be those that were congenitally deaf, had started learning SLN at age zero to two years, and had become fluent signers.

The goal was not reached. During the recruiting period only three people were enlisted that qualified as native signers. Five other participants who were deaf signers with varying characteristics were added to form a group of eight deaf signers. Another group was formed with eight hearing signers, all late learners, and another one with seven non-signers (one of the eight did not provide usable data):

- 8 Deaf signers (15-40 yr, 3 male, fluent to good in SLN, mostly early acquisition);
- 8 Hearing signers (21-46 yr, 1 male, fluent to reasonable in SLN, all late acquisition);
- 7 Non-signers (26-58 yr, 3 male).

The signers were connected to the Dutch Foundation for the Deaf and Hard of Hearing Child (NSDSK), or were students or teachers of the SLN programme at the Hogeschool Utrecht or friends and relatives of them. The non-signers included several students or teachers from the Delft University of Technology and several friends and relatives of the authors.

Material

The stimulus material consisted of 112 movie fragments (further abbreviated as 'movies'). These movies were constructed using sequences of 32 SLN signs, 9 fidgets and 4 rest positions. Before explaining the way these sequences were put together, the process of selecting the signs, fidgets and rest positions will be explained first.

Signs - The set of 32 signs, see Table 1, had to be representative in the sense that many or most of the possible variations in surface form should be present. Especially those variations should be present that previous research had reported to be of importance in the perception of signs. The signs were the citation form of signs (clear, isolated pronunciation). The form characteristics were examined using the "Standaard lexikon Nederlandse Gebarentaal, deel 1" (Nederlands Gebarencentrum, 2002).

Cross-sections of these signs could be made based on handedness, handshape-markedness, location, movement path, orientation change, handshape change and repetition:

- 16 were one-handed and 16 were two-handed (of which 3 were alternating, 8 were symmetrical and 5 involved a strong and a weak hand)
- 6 had a highly marked handshape, 12 a highly unmarked handshape, 14 were in-between.
- 13 were located in neutral space, 10 were on/near the face, 4 on the body, and 5 on the hand/arm.
- 24 included path movement: (5 Arc, 4 Bounce, 3 Circular, 9 Straight, 1 Tracing and 2 Zigzags)

- 11 included a change in orientation
- 6 included a change in handshape
- 17 included repetition

Such phonological variations are important in distinguishing signs from each other. This does not mean that they are automatically important to distinguish a sign from fidgets. It should be noted that the descriptions of surface form used here are quite general. Current phonological descriptions of SLN (NSDSK, 1988; Crasborn, 2001) that aim at a complete description of a sign use much more detailed annotation (i.e. "on the nose" instead of "on the face"). This experiment did not focus on exploring the influence of every detailed variation. Rather it was an exploration, using the more general variations, whether such variations have any effect in people's performance in detecting the beginning of a sign or do not have consequences at all.

One of the variations, called handshape-markedness, requires explanation. The term "markedness" was borrowed from sign language phonology to indicate that a more exceptional handshape requires more marking within the phonological system to describe it (Van der Kooij, 2002). Instead of looking at the 71 different handshapes of SLN the handshapes were grouped into highly marked, unmarked and those that are in-between[3]. Highly marked are those handshapes that are infrequently used in the SLN Lexicon (Nederlands Gebarencentrum, 2002), and are acquired last (Conlin et al., 2000). Unmarked handshapes are used frequently, and learned first.

Table 1. Glosses of the selected 32 SLN signs. These are given in small caps as customary.

One-handed Signs: Dutch gloss (English gloss)	**Two-handed Signs: Dutch gloss (English gloss)**
ZAND (SAND)	AUTO (CAR)
SCHEP (SHOVEL)	MELK (MILK)
EUROPA (EUROPE)	FIETS (BIKE)
TEKENEN (DRAW)	SOEP (SOUP)
WC (RESTROOM)	JARIG (BIRTHDAY)
BROER (BROTHER)	PAARD (HORSE)
BAD (BATH)	BOTERHAM (SANDWICH)
AFDROGEN lichaam (TOWEL-OFF body)	EGEL (HEDGEHOG)
VIES (DIRTY)	OPRUIMEN (TIDY UP)
KOORTS (FEVER)	RAAM (WINDOW)
MAMA (MOM)	TELEVISIE (TELEVISION)
PAPA (DAD)	BOOM (TREE)
KIJKEN (LOOK)	FEEST (PARTY)
KIP (CHICKEN)	AANKLEDEN (GET DRESSED)
MIS (MISSED)	POES (CAT)
TELEFOON (PHONE)	KOE (COW)

Table 2. List of fidgets used in experiment. [Brackets] are used to indicate fidgets.

Fidget	Description	Location	2Hands
[Lip Touch]	Touch lips with side of closed hands/indexfinger	Face	No
[Nose Rub]	Rub hand/indexfingers underneath nose	Face	No
[Chin Rub]	Rub hand/fingers along chinline	Face	No
[Ear Grab]	Grasp Earlobe	Face	No
[Hair Brush]	Brush hair with fingers	Face	No
[Arm Fold]	Fold both arms over eachother	Arm	Yes
[Hand Squeeze]	Squeeze Hands together	Arm	Yes
[Chest Scratch]	Scratch chest through clothing	Body	No
[Table Drum]	Wrap fingers/knuckles on tabletop	Tabletop	No

Table 3. List of rest positions used in experiment.

Rest position	Description
2H-Table	hands resting on table, fingers brought together
1H-Space	hand held floating, elbow on table
1H-Face	head/chin resting on hand
1H-Body	hand on chest, elbow on table

Table 4. Overview of the combinations of rest position, fidget and sign. All 32 signs, of which 16 are two-handed, are recorded in isolation with a neutral rest position. The one-handed signs are recorded in four additional ways: two non-neutral rest positions (RSL: rest in the same location. RDL: rest in a different location) and two preceding fidgets (FSL: fidget in the same location. FDL: fidget in another location). In addition there are 16 dummy sequences of two fidgets.

Stimulus-type	Rests	Fidget	1-Handed Signs			2-Handed Signs			
			Space	Body	Face	2H-Space	2H-Body	2H-Face	2H-Arm
32 Isolated	2H-Table	-	1-H	1-H	1-H	2-H	2-H	2-H	2-H
32 Isolated non-neutral rests	1H-Space	-	RSL		RDL				
	1H-Body	-		RSL					
	1H-Face	-	RDL	RDL	RSL				
32 Combos	2H-Table	1H-Body	FDL	FSL	FDL				
	2H-Table	1H-Face	FDL	FDL	FSL				
	2H-Table	2H-Arm	FDL	FDL	FDL				
	2H-Table	1H-Table	FDL	FDL	FDL				
16 Dummies	2H-Table	Combo of 2 fidgets							

Some restrictions in the selection process were also important. No signs were selected that specifically required a gaze shift (e.g. GOD) or a facial expression (e.g. CRY). This was done because the aim of this experiment was to isolate the manual features that contributed to the detection of signs.

Fidgets and rest positions - The first step in selecting fidgets was to examine videotapes of signed interactions. Only fidgets that people were actually seen to make were selected. Next, fidgets were selected such that they would not only be on the face (which was where most fidgeting was found to take place). This led to a set of nine fidgets, see Table 2. Four rest positions were used in the experiment, see Table 3.

Sequences - The overview in Table 4 shows the design of the contrasts between sequences. The sequences were composed in such a way that it was not predictable if a movement, whether first or second, was a sign (target) or fidget (distracter), and should allow us to get insight into:

- The influence of a non-neutral rest position in which case the hands were already in signing space or even in the same location before the sign began. This can be analysed by comparing the one-handed signs with a neutral rest position (two hands on the table or 2H-Table) to those with a non-neutral rest position in the same location as the sign (RSL) or in a different location (RDL).
- The influence of a preceding fidget in which case the hands were already in signing space and moving before the sign began. This can be analysed by comparing the one-handed signs without preceding movement to those with a preceding fidget, either in the same location (FSL) or in a different location (FDL). However, fidgets are not made in neutral space (they always involve contact with the body or an object), or at least none that came to our attention. So, for signs in neutral space (that is, signs not contacting

body parts) a fidget was selected that came closest to the initial location. For example, SAND starts low in space and was therefore preceded by a [table drum].

The two-handed signs were excluded from these comparisons for two reasons. First, it was deemed necessary to keep the number of sequences as small as possible for practical purposes. Second, the combination of different locations is more complex when two hands are under consideration.

Recording - The material was recorded in the facilities of the NSDSK. A hearing, late signer working for the NSDSK as a teacher of SLN performed the fidgeting and signing. A high-quality digital camera was used at 720*576 pixels PAL. The clothing (red) and background (blue) were chosen to provide good contrasts with the skin and the table (white), behind which the signer was seated. Diffuse lighting was created to avoid drop-shadows. During recording the signer was instructed not to use mouthing, to keep looking into the camera and to keep a straight face. This was done to isolate the influence of manual movement.

Procedure

The procedure consisted of the following steps

- Participants were seated at a table with a laptop and a written instruction containing the following explanation: "You will see a series of movies, in each of which a person makes a SLN sign and/or other hand movements. Please press the spacebar for each movie in the series as soon as you see the beginning of a SLN sign." The experimental software on the laptop was started which first repeated the instruction in SLN (a digital video clip) on the screen. Subjects then provided their personal data.
- Next, subjects could get used to the experimental procedure with five movies that did not belong to the stimulus material but were additional representative recordings. When the five samples were accomplished, the experimenter asked the subject whether the procedure was clear. If not then the instruction was repeated, after which the test started. If a subject had misunderstood the instruction (apparent from his behaviour and the debriefing) the data were not used in the analysis.
- The subject's reaction time to seeing a red flash (further abbreviated as "rf-reaction time") was measured (the first of three times, see below). Subjects had to "press the spacebar as soon as you see a red flash on the screen". The average rf-reaction time will be used as a measurement of the latency between visual information presentation and recording of the motor response (response time) during the actual task. Our research focuses on the information that is carried by the visual signal. Subtracting the rf-reaction time from the response time gives us a measurement of the time at which the information in the visual signal was sufficient for subjects to see the beginning of a sign (without latencies). In addition, differences between subjects in response time due to for example age or computer experience are controlled for.
- The first series of 112 movies was presented in a random order. Before each movie participants were instructed to "press the spacebar when you are ready to start. Then watch the crosshairs in the centre of the screen and press the spacebar again to start the movie. Press the spacebar again as soon as you see the beginning of a SLN sign." The movie was played full screen. Subjects could respond to a movement or choose not to press the spacebar in which case the movie ran to its end. If they pressed the spacebar the movie stopped and disappeared and the time when the spacebar was pressed was recorded. Then a pop-up window appeared asking them "OK, or try again later?". Their choice between "OK" and "again" was recorded. In case of "again" the movie was presented later on again in a random position amongst the remaining movies.
- The second measurement of rf-reaction time was performed.
- A short break followed. Subjects were invited to share their experience which was

written down, but neither feedback nor further instruction was given other than "please continue as you did".

- The second series was presented, consisting of the same 112 movies in a new random order. The reasons for this repetition were to gather more data from a participant, and to be able to analyse whether subjects' responsive behaviour changed with increased exposure to the stimulus material. To aid this analysis five participants performed a third and fourth series a week later.
- The third measurement of rf-reaction time was performed.
- The software was closed and subjects were debriefed (unless they were invited to perform a third and fourth series) and their questions were answered. If their observed behaviour was unclear the experimenter probed their understanding of the task. Before going home the subjects were financially rewarded for their time and co-operation.

Coding

Coding of fidgets, signs and movement phases - Three coders marked the beginning and end of each fidget and sign. They also coded the movement phases. Custom software was written for this coding (and for a visual inspection of the responses in relation to the coded events). The coding scheme for the movement phases in signs, see Table 5, was taken from Kita et. al. (1998) with small adaptations. The coders received the following explanations:

- **Liberation**. Freeing the hands from rest position when necessary.
- **Preparation**. Strict preparation for the stroke (or independent hold). Movement towards initial location and formation of initial handshape and orientation. Defined as first phase of the sign or fidget.
- **Stroke**. The expressive phase of the movement, together with the post-stroke hold. Either stroke or hold is obligatory.
- **Hold**. Maintaining a posture after the stroke or take on a posture instead of a stroke.
- **Relax**. Relaxing the handshape and location. Including going back to the home position.
- **Settle**. Movement within the rest position. Hands, arms and body settle back. Not part of the sign anymore.
- All movement phases are optional.

Pre-stroke holds (Kita et al., 1998) were not coded, because they did not appear to exist in this stimulus material. Settling was added to get a clearer boundary where a sign ends. We often found body movement going on for a while after the sign had obviously ended.

Reliability of Coding - For the three coders the agreement between them was calculated, see Figure 2. This was done for the beginning of the preparation and the beginning of the stroke (or independent hold instead). The preparation and stroke were used in the analysis of the results. One coder had some experience with SLN, one coder had considerable experience and one coder was a good signer.

The procedure was as follows: For each occurrence of a sign in a movie (each "sign production") the differences between the three coders were calculated. Then the largest time difference between two coders was taken. If this difference exceeded a certain tolerance in time difference then the three coders did not agree. If it was within this tolerance the three coders were said to agree on the time of beginning. The percentage of sign productions (from the total of 96) on which three coders agreed is used as the measure of agreement. In Figure 2, this agreement is plotted on the vertical axis, with the corresponding tolerance on the horizontal axis. To claim, for example, an inter-coder agreement of more than 80% we had to allow for 120 ms difference between them (which is 7.5% of the average duration of signs).

The average setting, of the three coders, was calculated for every beginning of preparation and stroke for all signs and all further analyses were performed with those time-averages.

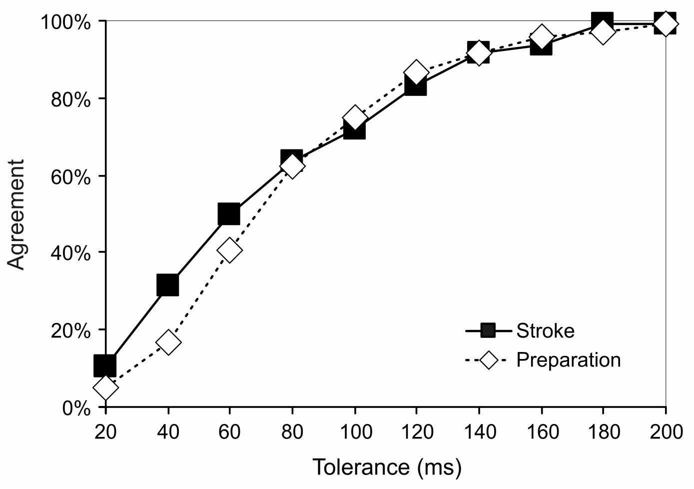


Figure 2. Reliability of the coding of preparation and stroke. Agreement is the percentage of sign productions (N = 96) on which three coders agreed on the time of beginning, given a certain tolerance in time difference.

Table 5. Coded movement phases. Each movement is first regarded as an excursion from and returning to a home position. An excursion can contain one or two movements (fidgets or signs). Each movement is divided into phases. A sign begins, by definition, at the start of the preparation and ends with the end of the recovery. Both liberating the hands and settling back into a home position are not considered to be part of the sign (or fidget), they are regarded as transitions from home to excursion and back. All movement phases are optional.

Home-Excursion	Movement phases		Definition of Sign
Home position	-		Not Sign
	Liberation		
Excursion	Preparation		Sign
	Expressive phase	Stroke	
		Hold	
	Recovery		
	Settling		Not Sign
Home position	-		

Table 6. Average duration (ms) of signs in isolation (sign1), signs preceded by a fidget (sign2), and for all of the signs together (total). Also given are the duration of the excursion and of the movement phases of the sign (in milliseonds and as a percentage of the duration of the sign).

Movement	Excursion	Sign	Preparation	Stroke+Hold	Recovery
Sign1	1835	1680	345 (21%)	835 (50%)	500 (30%)
Sign2	-	1385	280 (20%)	660 (47%)	455 (33%)
Total	-	1580	320 (20%)	775 (49%)	485 (31%)

Summary of coded duration of movement phases - An overview of the durations of signs and movement phases, as they were coded, is shown in Table 6. The average duration of isolated signs was 1680 milliseconds, counting from the start of the preparation to the end of the recovery. The average duration of the entire excursion (including part of the liberation and settling) was 1835 milliseconds, 155 milliseconds longer than the sign itself. Thus, about 10% of the excursion is not directly related to the sign itself. When a sign was made following a fidget it had an average duration of 1385 milliseconds, about 300 milliseconds shorter than the average duration of isolated signs. The nucleus, the stroke plus an optional post-stroke hold, accounted for about half of the duration of a sign, the preparation for twenty percent, and the recovery for thirty percent.

2.3. Results

This section begins with a structured overview of the data. Then the results are presented of two studies: one study looks into subjects' *capability* of seeing the beginning of a sign. The other study concerns subjects' *response times* for seeing the beginning of a sign.

Structuring the Data

Reaction time for seeing a red flash - For each subject the reaction time for seeing a red flash (rf-reaction time) was measured three times. The average for each subject was then calculated. The group means were 232, 239 and 252 milliseconds for the deaf signers, hearing signers and non-signers respectively. With standard deviations from 29 to 39 ms these outcomes were not significantly different (ANOVA, $F_{(2, 20)} = 0.613$, not significant (n.s.)). On an individual level, there was no correlation between subjects' rf-reaction times and their response times.

Linking the Responses to Coded Fidgets and Signs - If a subject responded on a certain movie (pressed the spacebar) then the response time was defined as the measured time when the spacebar was pressed (counting from the start of the movie) minus the subject's rf-reaction time. Using this response time, the responses were then linked to the actual movements based on the following rules and definitions:

- A response was linked to a sign or fidget if it occurred after its beginning.
- When a response was to a sign it was called a "hit" and a non-response (not pressing the spacebar) was called a "miss".
- A response to a fidget was defined as a "false alarm" and a non-response as a "correct rejection".
- When a combo was presented (i.e., fidget1 followed by sign2) a response to sign2 counted as a hit and also as a correct rejection of fidget1. If there was no response then this was counted as a miss on sign2 and a correct rejection of fidget1.
- If a dummy was presented (i.e., fidget1 followed by fidget2) a non-response was counted for both fidget1 and fidget2 as a correct rejection. If there was a response to fidget2 then that counted as a false alarm, but at the same time as a correct rejection of fidget1.
- In case of a combo respectively dummy, if a response was given after the beginning of sign2 respectively fidget2 in the sequence, but became an event before its beginning when the rf-reaction time was subtracted it was classified as ambiguous and removed from the data.
- If a response occurred prior to any movement (whether a fidget or a sign) in the movie it was classified as an accidental start and removed from the data.

Application of these rules to the measurements from the experiment yielded the results in Table 7. For isolated signs there were 3459 hits on the sign1 (with response times ranging from as early as 55 ms to very late responses of 2500 ms) and 121 misses. For combos there were 130 false alarms linked to the fidget1, and 1700 hits and 69 misses on sign2 (which were also 1769 correct rejections of fidget1). For dummies there were 58 false alarms on fidget1 and 68 false alarms on fidget 2, and 857 correct rejections of it. In total, there were 19 accidental starts and ambiguous responses which were excluded from further analysis.

Table 7. Number of responses that could or could not be linked to movements (that is: sign1, sign2, figdet1, fidget2) in the movies.
a. Movement to which the response was linked
b. The range (min-max) of the response times (corrected with rf-reaction time)
c. NA = Not Applicable

Type	Movement[a]	Response	Frequency	Response time (ms) [b]
Sign	None	Accidental Start	7	< -350
	Sign1	Hit	3459	55-2496
		Miss	121	NA[c]
Combo	None	Accidental Start	6	< -550
	Fidget1	False Alarm	130	24-1232
		Correct Rejection	(1700 + 69 =) 1769	NA
	Neither	Ambiguous	4	NA
	Sign2	Hit	1700	8-1910
		Miss	69	NA
Dummy	None	Accidental Start	1	-560
	Fidget1	False Alarm	58	67-762
		Correct Rejection	(68 + 857 =) 925	NA
	Neither	Ambiguous	1	NA
	Fidget2	False Alarm	68	99-2054
		Correct Rejection	857	NA
Total: 9175				

Table 8. Amount and percentage of selections of "again" per response type. Values are given for all subjects in total and per group.

Responses on signs			**OK**	**Again**	**Again%**
Hit	Group	Deaf signer	1391	10	1%
		Hearing signer	1492	16	1%
		Non-signer	1285	4	0%
	Total		4168	30	1%
Miss	Group	Deaf signer	34	10	23%
		Hearing signer	44	18	29%
		Non-signer	48	30	39%
	Total		126	58	32%
Responses on fidgets			**OK**	**Again**	**Again%**
False Alarm (FA)	Group	Deaf signer	21	38	64%
		Hearing signer	2	93	98%
		Non-signer	21	35	63%
	Total		44	166	79%
Correct Rejection (CR)	Group	Deaf signer	944	5	1%
		Hearing signer	1045	5	1%
		Non-signer	878	14	2%
	Total		2867	24	1%

Correction, repeated presentation of movies, repeated series - After each response subjects had to choose whether it was "OK" or they would like to try "again" later, in which case the movie was presented again at a random position in the remainder of the series. We found only three percent of the responses to a first presentation of a movie to be followed by "again". If a (movement in a) movie was presented for the second time (or more), then the chance that the response to it was followed by "again" increased to about 16%. The patterns of choosing "OK" and "again" did not appear to differ markedly between different series. For details and a complete overview, see Appendix 2-A.

In Table 8 the results are summarized by adding up the first and second series and all presentations, separated for each group of subjects. The third and fourth series are not used in this table because only five participants from two of the groups performed these series. False alarms are very often (79%) followed by "again" and this accounts for 60% of all the selections of "again". This was indicated in the debriefing by subjects as correcting mistakes.

When signs were missed this led to less corrections (32%) than with false alarms though still this is much higher than the selection of "again" in case of a hit or correct rejection (both 1%). In the pattern for each group one difference stands out: the hearing signers make more false alarms but correct almost every false alarm they make (98%). This is the only significant difference in usage of the "again" option found between the groups (Pearson Chi-Square 37.3, $p < 0.001$).

Study 1. Capability

In this first study the subjects' capability to perform the task was investigated. Good capability was defined as a high hit rate on signs and a low rate of false alarms on fidgets. Comparisons were made between (groups of) subjects, between (groups of) signs, between conditions with or without a preceding fidget, between conditions with different rest positions.

Capability comparison of the different subjects - The results, see Table 9, showed a remarkable performance by the non-signers. They had only a slightly lower group hit rate than the signer groups and about an equal false alarm rate. The differences between the groups in the distribution of hits and misses are significant when tested with all the measurements (Pearson chi-square 12.7, $p < 0.01$). The differences are not significant when tested with only the data that was confirmed with "OK" (Pearson chi-square 3.6, n.s.). Direct comparisons between any two groups all showed no significant differences. In other words, deaf signers did not clearly have higher hit rate than hearing signers. Furthermore, signers did not clearly have higher hit rates than non-signers. All three groups had almost the same, very high hit rate.

All groups had extremely low rates of false alarms. Intriguingly hearing signers performed worse than the other groups when the difference was tested with all the data (Pearson chi-square 6.7, $p < 0.05$) but they were better performers if it was tested with only the data that was confirmed with "OK" (Pearson chi-square 19.2, $p < 0.001$).

The individual results in capacity were further studied to check the variance within the groups and to check for effects of age, gender, being a native signer, age of SLN acquisition, SLN fluency and SLN usage. No effects were found. Within each group, the individual subject with the median hit rate had a hit rate of 98% (series 1 and 2 added together). Each group contained one or more people with a perfect 100% hit rate. These findings also demonstrate the similarity of the groups. For an overview of the individual results see Appendix 2-B.

Capability changes during participation - To check whether the performance improved during participation in the experiment a comparison was made between the results for each of the four series performed by five of the subjects (using only the responses confirmed with "OK"). The rate of false alarms for these five subjects was 0% in each series, but their hit rate was somewhat lower in the first series (95%) than in the subsequent series (100%), which was a significant improvement (Pearson Chi-square 57.4, $p < 0.001$).

Improvement was limited to the first series. Therefore, hit rates and rates of false alarms were calculated for groups of 20 presented movies of the first series with the "OK" responses of all subjects. This means that the results for the first 20 movies offered during the series were separated, then for the next 20 movies offered and so on. For the hit rate on signs, the improvement was limited to the first 80 movies (going from 90% to 98%). For the rate of false alarms on fidgets the improvement was limited to the first 60 movies (going from 5% to 1%). The patterns of improvement appeared similar for the three subject groups.

An additional check of the results for the first 20 movies presented, with all the data including those when "again" was selected, showed a hit rate of 87% and a rate of false alarms of 13%. A binomial test showed these rates to be well above chance level, which was also true for all subsequent groups of 20 movies.

Comparison of hit rates between signs - Table 10 shows that hit rates were not equal for all one-handed signs (Pearson Chi-square 148.9, $p < 0.001$). The two-handed signs were less often presented and the low number of misses they caused did not allow for a statistical comparison between them. A comparison between the grand totals for one- and two-handed signs showed a slightly better hit rate for two-handed signs (97.5%) than for one-handed signs (95.5%) (Fisher's exact, $p < 0.01$).

Table 9. Hit rate and false alarm (FA) rate per group. Hit rate is calculated from: hits/(hits+misses). False alarm rate is calculated from: FA/(FA+CR).

Data included	Group	Hit rate	FA-rate
Using only responses that are confirmed with "OK"	Deaf signer	98%	2%
	Hearing signer	97%	0%
	Non-signer	96%	2%
	All groups	97%	2%
Using all data (including responses corrected with "again")	Deaf signer	97%	6%
	Hearing signer	96%	8%
	Non-signer	94%	6%
	All groups	96%	7%

Table 10. Hits and misses for all productions of a sign.

One-handed signs	Hit	Miss	Total	Two-handed signs	Hit	Miss	Total
DRAW	223	5	228	CAR	46	1	47
TOWEL-OFF body	212	16	228	BIKE	46	0	46
LOOK	223	2	225	MILK	45	0	45
CHICKEN	219	7	226	HEDGEHOG	43	3	46
FEVER	195	39	234	BIRTHDAY	44	2	46
MOM	203	29	232	SOUP	45	1	46
MISSED	222	6	228	HORSE	46	1	47
DAD	220	15	235	SANDWICH	45	0	45
SHOVEL	218	7	225	CAT	45	1	46
PHONE	223	3	226	PARTY	45	0	45
DIRTY	216	7	223	COW	45	1	46
BROTHER	220	6	226	WINDOW	42	3	45
BATH	223	5	228	GET DRESSED	45	0	45
SAND	221	8	229	TREE	43	3	46
EUROPE	223	7	230	TELEVISION	45	1	46
RESTROOM	222	4	226	TIDY UP	45	1	46
Total	3483	166	3649	Total	715	18	733

Each of the one-handed signs was part of five movies: one with a neutral rest position, two with non-neutral rest positions, and two with preceding fidgets. The misses were evenly distributed across these actual productions of the signs. Four of the one-handed signs were noticeably high causers of misses. These were FEVER, MOM, TOWEL-OFF body, and DAD, with hit rates of 83%, 88%, 93% and 94% respectively. The misses on TOWEL-OFF body were almost all by hearing signers. DAD only produced misses with non-signers.

The results for the one-handed signs were compared on presence of a preceding fidget and on rest location. It did not make any difference in hit rate whether a sign was made directly from a rest position (95.5%) or if it was preceded by a fidget (95.4%). Chi-square tests showed neither a difference for the totals of 1-handed signs nor for each sign separately. The same was true for variations in rest position: there was neither a difference in hit rates between the totals nor for any individual sign. Finally, comparisons were made between (a) signs containing a location shift during the preparation and (b) signs where a preceding fidget was made in the same location or where the rest position was maintained in the same location: No effect was found. In summary, subjects' capability in detecting the beginning of a sign appeared neither to be influenced by the presence of preceding fidgets, nor by starting from non-neutral rest positions (with the hands already in signing space), nor by a lack of location shift during preparation.

Using all data from the first two series a comparison was made between hit rates of signs made in the four locations: space (97%), face (94%), body (96%) or arm (97%). The frequency distributions of hits and misses were not equal (Pearson Chi-square 21.2, $p < 0.001$). Using only the data that was confirmed with "OK" did not alter these findings, nor did excluding two-handed signs or all but neutral rest positions.

The markedness of the handshape had an effect within the group of 1-handed signs (taking all data of the first two series) with hit rates for unmarked handshapes at 93%, marked handshapes at 98% and in-between handshapes at 95% (Pearson Chi-square 18.1, $p < 0.001$). If the two-handed signs were included in the analysis and it was restricted to the isolated production from a neutral rest position, then there was no effect. Restricting the comparisons to responses that were confirmed with "OK" did not alter these findings.

Four different movement characteristics were studied for effects on hit rate: path movement, change of orientation, change of handshape, and repetition. The hit rates for the different types of path movement (none 95%, straight 98%, arc 97%, zigzag 97%, bounce 93%, circular 95%, tracing 93%) were significantly different (Pearson Chi-square 33.3, $p < 0.001$) when tested with all data of all signs or with just the 1-handed signs and/or the isolated signs (directly from a rest position). The presence of changes in orientation or changes in handshape did not have a clear effect on capability. Repetition had no effect.

Table 11. False alarms per type of fidget. Some fidgets occurred more frequently than others. All data is used from series 1 and 2.

Fidget	False Alarm	Correct Rejection	False alarm rate
[Arm Fold]	5	264	2%
[Chest Scratch]	21	418	5%
[Chin Rub]	22	362	6%
[Ear Grab]	36	301	11%
[Hair Brush]	21	317	6%
[Hand Squeeze]	22	470	4%
[Lip Touch]	15	315	5%
[Nose Rub]	67	305	18%
[Table Drum]	1	139	1%
Total	210	2891	7%

False alarm rates between fidgets - As shown in Table 11 the false alarms were not evenly distributed over the fidget movements (Pearson Chi-square 111.3, $p < 0.001$). [Nose rub] fidgets lead to false alarms in 18% of the cases and accounted for about a third of all false alarms. [Ear grab] fidgets caused an 11% false alarm rate. All others were at 6% or lower.

The actual "fidget productions", the occurrences of a fidget in a movie, were not evenly responsible for false alarms (Pearson Chi-square 29.1, $p < 0.001$). Two particular [nose rub] movements (out of seven) had very high false alarm rates of 33% and 30%. Those two movements (out of 64 fidget productions) together caused 40 false alarms, almost a fifth of the total of 210. A visual inspection of these two movements did not clarify why they caused so many false alarms.

It was checked whether the groups of subjects differed in response pattern to certain fidgets (interaction). Although the numbers of false alarms was too low for a quantitative statistical comparison the pattern appeared to be similar for all groups.

Study 2: Response time for hits

In this second study the response times are examined for hits on signs. First, the results are compared between subjects and series, then between signs. In addition, the response times are investigated in relation to the progress of movement phases (e.g. the onset of the stroke).

Comparison between subjects and between series - For each series of a subject the median was determined of all hits on signs that were confirmed with "OK". The median is taken because the distributions typically have a long, thin tail to the right with late responses (avg. skewness 2.2) and a strong central cluster (avg. kurtosis 8.1). On average the median response time (of all signs in a series from a subject) is about 500 milliseconds after the beginning of the (preparation of the) sign. For the three subject groups, the group mean of the median response times for each of the series are given in Figure 3. The group of hearing signers contained four subjects who performed a third and fourth series, which are shown in the graph but are left out of further statistical tests. Both factors, group and series, have a significant effect on response times (ANOVA: Group $F_{(2,\ 39)} = 5.88$, $p < 0.01$; Series $F_{(1,\ 39)} = 8.82$, $p < 0.01$). There is no significant interaction between them and if we test on the mean of the mean of each series of each subject, instead of the mean of the median, these results do not alter. Of the three groups the non-signers need most time to respond (first series 680 ms, second 540 ms, average 610 ms), and hearing signers (first series 510 ms, second 400 ms, average 450 ms) take less time than deaf signers (first series 550 ms, second 460 ms, average 510 ms).

The response times drop about 100 milliseconds from the first series (570 ms average) to the second series (470 ms average). Over the third and fourth series together the response times appear to drop another 90 milliseconds (ending up at 310 ms for the hearing signers). Age and gender do not have an effect on response time. For detailed individual response time results see Appendix 2-C.

Comparison between signs: Response times and onsets of stroke - For each sign the median was determined of all hits on that sign that were confirmed with "OK" (all subjects, all series). This was repeated per group of subjects and per series to monitor previously found effects. The median is taken because the distributions typically have a tail to the right of late responses and a strong central cluster. Each median represents the response time for that sign movement.

When making comparisons between different sign characteristics it is not irrelevant to consider the influence of a possibly important variable: the onset of the stroke within the sign movement (Kita et al., 1998). A note of caution about the following analysis is that the concept of stroke does not refer to a single objective property of the movements, nor was it part of the subjects' task. Three coders marked the onset of the stroke. The relation that is examined is between the average of this subjective coding of the stroke and the median response time of subjects in the task of detecting the beginning of a sign.

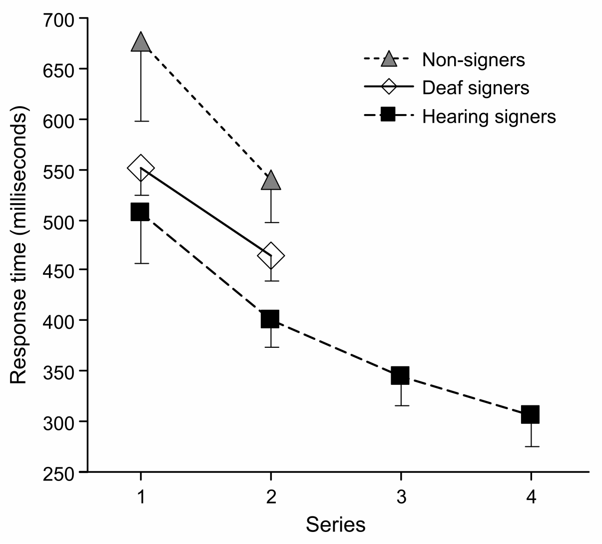


Figure 3. Mean response times per group and series. One standard error of the mean is plotted in a single branch downward (to avoid cluttering the image)

In Figure 4 the relation between the median response time and the onset of the stroke is given for each sign production. The time of the onset of the stroke is measured from the beginning of the preparation (that is the beginning of the sign), and thus equals the duration of the preparation. There is a significant correlation ($R = 0.65$, $R^2 = 0.423$, $F_{(1, 94)} = 68.8$, $p < 0.001$) between response time and onset of the stroke. If we estimate a linear relation then the slope is less than one (0.705). The correlation found is strongest for signs that begin directly from a rest position (sign1, $R^2 = 0.503$, $p < 0.001$). For signs that are preceded by a fidget (sign2) there is no significant correlation, yet the estimated fit for all sign productions does appear to approach the data fairly well. In further comparisons between groups of signs, the onset of the stroke is used as a covariate in explaining the variance of the response time.

The median response times often occur somewhat after the onset of the stroke, which is shown in Figure 5. On the Y-axis, the onset of the stroke is subtracted from the response time. Linear curve estimations are now done separately for the groups of subjects and series. The correlations between median response time and onset of stroke remain significant even with these smaller data sets (ranging from $R^2 = 0.16$ to $R^2 = 0.65$). In both graphs of Figure 5 three lines are drawn: the top line is the predicted median response time behaviour for the non-signers, the middle line is for deaf signers, the lowest for hearing signers.

In the first series, shown in the left graph, the non-signers are predicted to respond about 320 milliseconds after the onset of the stroke, given an average duration of the preparation. In the same situation the deaf signers are predicted to respond after 255 milliseconds and the hearing signers after 190 milliseconds. In the second series the non-signers estimate, for an average duration of the preparation, is at 215, the deaf signers at 165, and the hearing signers at 90 milliseconds. The early stages of the stroke (say, less than 400 ms) usually provide sufficient information to detect the beginning of a sign. The slopes fall slighty (-0.17 to -0.31) indicating that the information in the preparation does play a role in detecting the beginning of a sign. Interestingly, several median response times in the second series occur before the onset of the stroke. Henceforth, for signs with a long preparation the estimated response time minus the onset of stroke even becomes zero for the hearing signers. The information in the preparation is sometimes sufficient to detect the beginning of a sign, especially if the preparation is long, and if it is the second time a subject sees the movie.

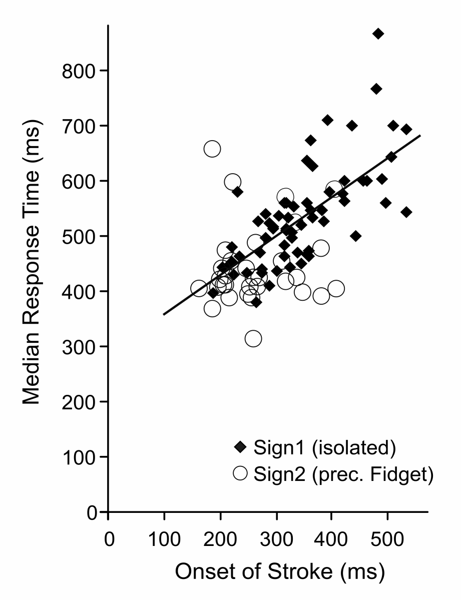


Figure 4. Median response times and onset of stroke for each production of a sign (sign1 = directly from a rest position; sign2 = preceded by a fidget). The line represents a linear fit resulting from a regression analysis.

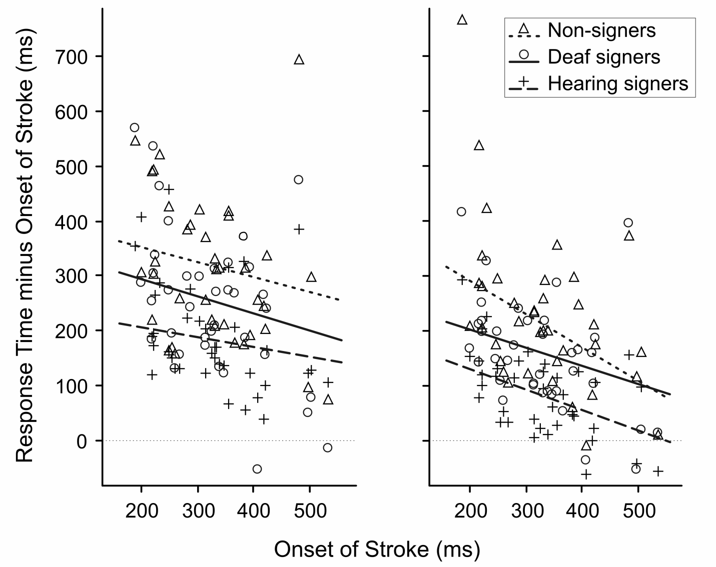


Figure 5. Median response times minus onset of stroke (per group) versus onset of stroke for each production of a sign. A random selection of 40% of the signs is plotted, to reduce cluttering. The lines represent a linear fit resulting from a regression analysis with all of the sign productions. The left graph is of the data from series 1, the right of series 2. In both series the top line is the response time prediction for non-signers (slowest), then the middle line for deaf signers, and the bottom line for hearing signers.

Comparison between signs: Response times and characteristic features - For the comparisons between (groups of) signs on characteristics the overall median response times are used. Differences between groups of subjects or series are disregarded. Only relative effects are reported. Comparisons are made using separate one-way ANOVAs, with the onset of the stroke as a covariate.

There is an effect of a preceding fidget: a preceding fidget causes a drop in response times of 75 ms ($F_{(1, 45)} = 10.78$, $p < 0.01$, tested with one-handed signs in neutral production and with preceding fidgets). Differences in rest position are non-significant with the onset of the stroke as a covariate ($F_{(3, 43)} = 2.0$, n.s.).

In a comparison between the 16 two-handed signs (in a neutral production) and the 16 neutral productions of one-handed signs, the two-handed signs showed somewhat higher response times (50 ms) than the one-handed ($F_{(1, 29)} = 4.2$, $p < 0.05$). It should be noted that the one-handed signs were presented in five conditions, whereas the two-handed signs appeared only once in a series.

Using the same 32 isolated neutral productions of signs, the signs with highly marked handshapes had significantly shorter response times than either signs with non-marked handshapes (95 ms shorter) or signs with in-between handshapes (85 ms shorter) ($F_{(2, 28)} = 4.64$, $p<0.05$).

The following sign characteristics did not have a significant effect on response times: location, symmetry (in two-handed signs), path movement, change of orientation, change of handshape, and repetition.

2.4. Discussion

Study 1 shows that all participants are highly capable of responding to signs (hit rates at least 95%) and not to fidgets (rate of false alarms around 5%). There was no clear capability difference between the groups of subjects. A few signs caused some confusion, most notably FEVER (hit rate 83%) and MOM (88%). Several form variations had small effects on performance level: it was higher for signs that are two-handed (+2%), in space (+3% compared to on or near the face), have a highly marked handshape (+5% compared to unmarked), and/or contain path movement (+3% for straight path versus none). There was no effect from a preceding fidget or non-neutral rest position.

Study 2 shows that response times had strong clusters around 500 milliseconds after the beginning of the sign. The non-signers had the highest response times (~600 ms). Hearing signers (~450 ms) had shorter response times than deaf signers (~510 ms). If you think of a sign as consisting of the sequence of preparation, stroke and recovery then responses occurred on average around 200 ms after the onset of the stroke. The onset of the stroke in a sign is correlated with the median of response times of hits on that sign. However, information in the preparation is used and can even be sufficient to detect the beginning of a sign, especially if the movie was seen before and the preparation is long. Response time was further influenced by three factors between signs: it was shorter for signs that have a highly marked handshape (95 ms shorter compared to unmarked), are preceded by fidgets (75 ms shorter), and/or are one-handed (50 ms shorter).

Signs and fidgets can be discriminated on their appearance - Our findings support the idea (Kendon, 2004) that signs, as cases of gesture, can be discriminated from other movements, in our study fidgeting, by their appearances. Experience with sign language is not necessary for this ability. These results should not be generalised without further research to signs in other settings, for example a piece of a signed conversation (compare Gerwing and Bavelas (2004)).

Early stages of a sign are sufficient - The information provided in the early stages of the sign, the preparation (about 300 ms) and the first half of the stroke (about 400 ms) in most cases appears to be sufficient to detect the beginning of a sign. The overall median response time is 500 milliseconds. The preparation is not void of information, sometimes it is even sufficient. This limits the search for the movement features that people use to discriminate signs.

Deaf signers respond later than hearing signers - The finding in our experiment that deaf signers showed higher response times than hearing signers might seem to go against what one might have expected. Some believe that native, deaf signers develop abilities that should allow them to perform faster at tasks involving visual perception. This would be especially true for tasks involving looking at sign language.

A possible explanation for the somewhat higher response times of deaf signers is that they responded more conservatively. Our results show that the deaf signers have the same high hit rate (the observed mean hit rate was slightly higher, yet not signicantly so). If we look at the rate of false alarms we notice that the hearing signers have a higher rate of false alarms But this is only true when we count all of the responses and not just the confirmed measurements. The hearing signers almost always used the option "again" when they had a false alarm. They appeared to more heavily rely on this option to correct possible mistakes, while taking the opportunity to respond quickly. The deaf signers seem much less likely to profit from this opportunity. They used the option "again" much less often yet still managed to have a very low rate of false alarms. This is a conservative pattern.

Rothpletz et al. (2003) studied response times to the onset of target events in the visual periphery. They also found a longer response times for deaf participants as compared to hearing participants. Rothpletz et al. (2003) further provide a well balanced overview of other research on differences between deaf and normal-hearing individuals on specific visual tasks. Studies are quoted showing that deaf individuals were often equal, sometimes worse and sometimes better performers. There appears to be no evidence to accept either general visual deficits or enhancements in deaf individuals, yet some differences may arise from difference in experience, Rothpletz et al. (2003) conclude.

Non-signers respond later than signers - The finding that signers have shorter response times than non-signers appears to be what one might expect, at first glance, yet it raises questions. First, do signers use different features from non-signers? Given the relatively small difference (100-150 ms) it does not appear likely that entirely different features are used. Second: are signers able to more quickly pick up certain information from a movement's appearance that helps them see that it is a sign? This reminds us of a research paradigm developed by Rosch (1978) to study categorization: the time it takes to identify a case as a member of a category is a function of its prototypicality; the more prototypical the case, the faster its identification. We then need to assume two things to explain the finding. First, for experienced signers, signs form specific (sub)categories of visible movement. Such categories are structured using prototypical features (e.g. handshapes, locations, motions, etc). Second, non-signers also categorize gestures using prototypical features, but these are slightly different. If we make these assumptions we may think of sign language acquisition as a process where (among other things) one develops specialized "categorical" perception (Emmorey et al., 2003); not wholly different from a hearing person's perception of gestures, only *fine-tuned* to the language. But then it is logical to expect signers to be faster because the movements are more prototypical for signs (as they know them) than they are prototypical for gestures (as a non-signer knows them). Note that there is a difference between prototypicality (how well a case represents a category) and membership of a category (Rosch, 1978; Lakoff, 1987): it is still entirely possible that a movement is as much a member of the category "gesture" as it is of "sign", hence the lack of differences in capability.

Some phonological properties have a small effect - As already remarked, only a highly marked handshape has the combined effect of better performance and shorter response time. This is in line with the findings of Emmorey & Corina (1990) that the handshape information becomes available early (they also found marked handshapes to decrease the required time to recognize signs). It has also been suggested (Kita et al., 1998) that handshape and orientation (hand-internal information) appear gradually toward the end of the preparation. This may be important information that allows people to detect the beginning of a sign, even before the stroke.

If path movement was present in a sign, subjects had higher hit rates, but not shorter response times. Perhaps this is because movement information simply becomes available later in the signal (Emmorey & Corina, 1990). It may become completely apparent early in the stroke. But by that time other features may also have appeared that suffice to detect the beginning of a sign.

No single feature was necessary to detect the beginning of a sign. Neither the absence of path movement, nor the absence of a highly marked handshape (or any other feature) reduced people's ability to chance level. The sign with the lowest hit rate (83%) was FEVER, which does not have movement at all. It is an "independent hold" in the words of Kita et al. (1998). The location is on the face, which in our study caused misses more often. The other independent hold, TELEPHONE, has a normal hit rate and a highly marked "Y" handshape.

Conditions prior to the sign have no negative effect - The only effect we found of conditions prior to the sign was that a preceding fidget caused shorter response times. Although signs in space had a higher hit rate it appears unimportant whether the hands enter this space from a rest position outside of it or not. There appears to be no important information in shifts in locations prior to the sign that can be used to discriminate signs. It is not unreasonable to expect this finding; after all, if the hands are shifted to a location, this might as well be done to fidget there.

Why does a preceding fidget in the same location not hamper sign detection? We had specifically constructed movies where we thought the fidget might camouflage the sign. For example, there was a sequence with a [nose rub] and then CHICKEN, made by the side of the nose. In another case a [chest scratch] preceded BATH, in exactly the same location. None of our camouflage attempts had any effect. We suspect that the boundaries (Rubin & Richards, 1985) between movements remained clearly visible. Segmentation of the motion into units and phases was not hampered, and sign detection proceeded in a regular fashion. How these boundaries can be extracted automatically is a topic for further research (see also Parish et al., 1990).

Implications for automatic gesture recognition - This study may serve as a benchmark for computer vision. That is, if the responses of a machine must match human performance it should quickly be able to see that an unfolding movement is a gesture (to which it should attend) or that it is just fidgeting. Furthermore, the machine should not require humans to use specific rest positions before gesturing, nor should it be troubled by sequences of movements. If it cannot do this then humans are forced to adapt their behaviour to the machine.

Our results tend to support the idea that detecting gestures in a continuous stream of human behaviour (for example in security camera (CCTV) footage (Troscianko et al., 2004)) is possible on the basis of the appearance of gestures alone. But here we must be cautious with expectations, for we did not find any specific gestural features that were always present in the movement of gestures and not in fidgeting motions. Alternatively, discrimination between fidgets and gestures may well require knowledge of the specific appearances of both fidgets and gestures. Further research is needed here.

Notes

1. Laban (and others following his lead) developed the Effort/Shape method to describe movement (Laban & Lawrence, 1947; Laban, 1975; Dell, 1977). An important effort factor in the description is the 'flow' of tension, which can be bound or free. With respect to gestures Dell (1977) notes that many conversational gestures are prepared and concluded with such flow changes. Kendon (1980), McNeill (1992), and Kita et al. (1998) refer to Laban's notion of effort (Dell, 1977) to explain the perception and segmentation of gestures. They state that the stroke is visible to the human eye because the efforts become more clearly defined and are focused on the form of the movement itself. In the preparation, if there is discernible effort it is focused in the movement by

which the hand reaches the position from which the stroke is performed.

2. Language users are often divided into those that are ‘native’ (typically, starting acquisition of the language before the age of two) and those that are ‘late’ (later start of acquisition) or otherwise ‘non-native’ (in the case of sign languages, it is usually considered important whether the parents are also deaf and use the same sign language).
3. An example of a highly marked handshape is the so-called ‘Y-hand’ (in SLN), which is formed by a closed hand with an extended thumb and pink. An example of an unmarked handshape is a closed fist or a flat, extended hand.

Appendix 2-A. Responses per series and movement type

After each response, or non-response, subjects had to choose either “OK” or “again”. When “again” was selected the movie was presented again later in the series at a random spot. This cycle could happen again and again in exactly the same way. The number of presentations is counted separately for each movement in a movie. In case of a response to a fidget1 in a movie, followed by “again”, subjects had not yet seen the sign2 or fidget2 of that movie. So their next response is to a second presentation for that fidget1 but a first presentation of the second movement in this movie. Table 12 shows how many responses of each type occurred. A distinction is made between the first, second and further series in which the subject participated.

Some cases may serve best to explain Table 12. Let us look at a simple case first: A subject, in his first series, is presented a sign and responds with a hit and confirms it with “OK” then this shows up in top-left cell (of the cells containing numbers) which contains the 1372 confirmed hits on a first presentation of a sign1 movement in one of the movies of series 1.

Another case which is slightly more complex: A subject is shown a combo of fidget1 and sign2 and does not press the spacebar and selects “again”. His non-response is recorded as a correct rejection of the fidget1 and a miss of the sign2 (both with “again”). The movie is then presented a second time (presentation = 2) and the subject presses the spacebar on the sign2 and confirms the measurement. This creates a hit on sign2 and a correct rejection of the fidget1 (both with “OK”).

Finally, a very complicated but existing case: A subject is presented with a dummy movie. The first time she presses the spacebar during the second fidget but selects “again”. This creates a false alarm on the fidget2 with “again” and a correct rejection on the fidget1 (with “OK” since the selecion of “again” is assumed to be related to the response on fidget2 and not to letting the fidget1 pass without pressing). To the second presentation she responds with a false alarm on the first fidget and again presses “again”. This creates a false alarm on the fidget1 with “again”. The movie is presented a third time upon which she responds with another false alarm on fidget2 and “again” (this movement was however presented only twice so far, since the previous presentation was cut short at the fidget1). This creates a false alarm for fidget2 (with “again”) and a correct rejection of fidget1 (with “OK”). On the fourth and fifth presentation the same thing happens until finally, the sixth time the movie is presented, the subject correctly rejects both fidgets of the dummy and confirms with “OK”. So, this has led to 6 responses to presentations of fidget1 and 5 to fidget2. This case is the most elaborate of the results. It occurred in the second series.

Table 12. Responses on the (repeated) presentation of movements

a. Note that for fidget1 it can happen that there are more second presentations than responses to a first presentation with "again". This is because when a hit to a sign2 or false alarm on a fidget2 is corrected with "again", then this is interpreted as a correction of that hit or false alarm and not as a correction of the non-response to the fidget1. The non-response to the fidget1 is in that case stored with "OK". However, the movie will be presented again with the fidget1 in it, leading to more second presentations.

b. The totals from this table and the overall frequencies of responsetypes diverge because the 19 measurements that were ambiguous or accidental starts are not in this table. If an ambiguous result is corrected with "again" it counts as a first presentation of both movements.

Series	Movement	Response	Presentation nr. X							
			#1		#2		#3		#4 (+ 5 + 6)	
			OK	Again	OK	Again	OK	Again	OK	Again
#1	Sign1	Hit	1372	15	32	2	1			
		Miss	61	22	3		1			
	Sign2	Hit	666	7	20					
		Miss	36	13	1					
	Fidget1[a]	FA	23	51	1	17		2		1
		CR	1018	13	66	1	18		4 (+1)	
	Fidget2	FA	6	14		4		2		
		CR	335	1	12		2		2	
#2	Sign1	Hit	1377	4	11		3			
		Miss	15	11	1	3				
	Sign2	Hit	678	2	7				1	
		Miss	7	7	1	1		1		
	Fidget1	FA	10	43		11		3		1
		CR	994	6	51	1	9	2	6 (+2+1)	
	Fidget2	FA	3	12	1	3		1		1
		CR	334		9		2		(+1)	
#3	Sign1	Hit	318	1	2					
		Miss		1						
	Sign2	Hit	156	2	2					
		Miss								
	Fidget1	FA		13		1				
		CR	226		17		5		2 (+1)	
	Fidget2	FA	1	5		3		2		1
		CR	74		2		1		1 (+1)	
#4	Sign1	Hit	315	2	3		1			
		Miss		2		1				
	Sign2	Hit	157		2					
		Miss		2						
	Fidget1	FA		11						
		CR	226	2	20		2			
	Fidget2	FA		7		2				
		CR	73		5		2			
	9156[b]	Total	8481	269 3%	269	50 16%	47	13 22%	23	4 15%

Appendix 2-B. Individual Capability Results

The individual capability results are shown in Figure 6. This graph (an ROC curve) is a common way to represent the sensitivity of people in a signal detection task (Haber & Hershenson, 1980). In this case the task is to detect the beginning of a SLN sign and the noise is the beginning of a fidget. If subjects are not sensitive to the difference between signal and noise then they will respond at chance level. This is visualized by the diagonal line in the graph. It runs from the lower left corner (a person does not respond at all, neither to signal nor to noise) to the upper right corner (a person responds to everything, be it signal or noise). The more a subject's result goes to the upper left corner of the graph the higher is the sensitivity shown by this subject in the task: It means that he/she responds to signal but not to noise. In this case, all respondents showed excellent sensitivity.

If a person has a relatively low hit rate in combination with a relatively low rate of false alarms then this person can be called "conservative" in his response. If a person combines a high hit rate with a high rate of false alarms then he can be called "liberal" in his response. In this case there are some small differences.

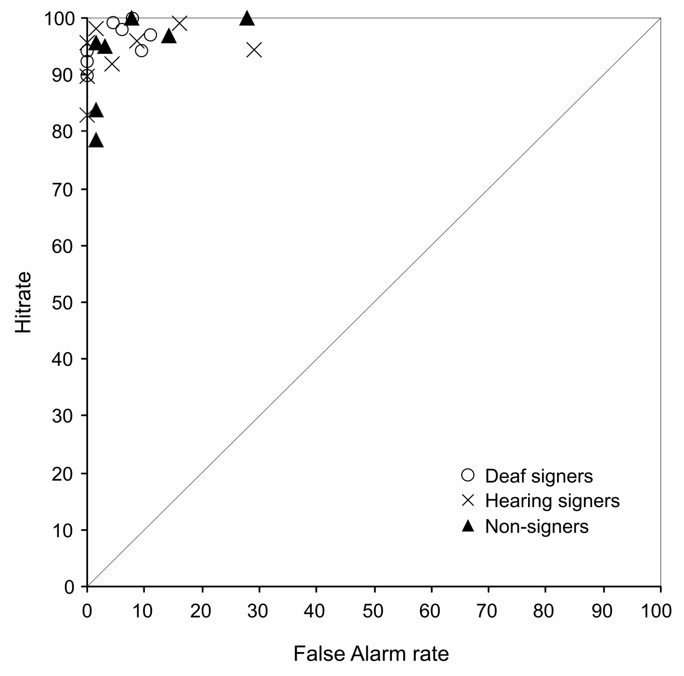


Figure 6. For each subject the hit rate and rate of false alarms is given. These are combined in a two-dimensional plot. On the x-axis the rate of false alarms is shown and on the y-axis the hit rate. All data is used from the first series, since the hit rates are lowest in this series and may show the most contrast between (groups of) subjects. For that same reason the responses that were followed by selecting "again" are also included.

Appendix 2-C. Individual Response Time Results

For each subject and each series performed by that subject the response times are given in Table 13 in the form of some descriptive statistics: median, mean, standard deviation (SD), standard error of the mean (SE) and number of measurements (N). Only the response times for hits on a sign production that were confirmed with “OK”are included (misses do not have a response times). N can vary because a subject can have a confirmed miss on a sign production and will therefore not have a hit on that same event. If all sign productions in a series are hit then N equals 96. The descriptive statistics are also given for the total of gathered data from that subject.

Table 13. Individual response time results.

Group	Subj.	Series	Median	Mean	SD	SE	N
Deaf Signer	1	1	613	685	357	36.8	94
	2	1	406	439	249	25.7	94
		2	361	371	138	14.1	96
		Total	379	404	203	14.7	190
	3	1	627	696	333	34.0	96
		2	522	573	257	26.2	96
		Total	575	635	303	21.9	192
	4	1	489	523	191	19.7	94
		2	508	539	166	17.3	92
		Total	496	531	178	13.1	186
	5	1	603	649	236	24.3	94
		2	501	534	174	17.8	96
		3	506	541	184	18.7	96
		4	439	475	162	16.6	96
		Total	508	549	200	10.2	382
	6	1	606	626	217	22.5	93
		2	467	501	205	21.3	93
		Total	545	564	220	16.1	186
	7	1	513	525	179	18.8	90
		2	378	380	116	12.0	93
		Total	437	451	166	12.3	183
	8	1	555	632	228	24.3	88
		2	517	558	235	24.7	91
		Total	528	594	234	17.5	179

Table 13. Continued

Group	Subj.	Series	Median	Mean	SD	SE	N
Hearing Signer	9	1	462	506	251	26.6	89
		2	354	363	137	13.9	96
		3	335	336	125	12.8	95
		4	319	313	113	11.7	94
		Total	367	378	179	9.3	374
	10	1	474	533	250	25.5	96
		2	430	427	100	10.2	96
		Total	447	480	197	14.2	192
	11	1	649	733	263	26.9	95
		2	523	538	144	14.7	96
		Total	574	635	232	16.8	191
	12	1	769	875	427	44.5	92
		2	433	492	206	21.0	96
		Total	565	679	384	28.0	188
	13	1	427	435	192	19.6	96
		2	354	365	127	13.0	96
		Total	394	400	166	12.0	192
	14	1	451	485	220	22.9	93
		2	396	403	118	12.0	96
		3	363	366	113	11.6	95
		4	319	321	104	10.6	96
		Total	371	393	157	8.1	380
	15	1	306	325	120	12.2	96
		2	263	283	109	11.1	96
		3	271	277	102	10.4	96
		4	220	230	95	9.7	96
		Total	269	279	112	5.7	384
	16	1	515	590	290	30.6	90
		2	456	494	194	20.2	93
		3	408	429	133	13.6	96
		4	369	390	107	10.9	96
		Total	442	474	206	10.6	375
Group	**Subj.**	**Series**	**Median**	**Mean**	**SD**	**SE**	**N**
Non-Signer	17	1	418	432	131	13.6	93
		2	402	410	138	14.1	96
		Total	410	421	135	9.8	189
	18	1	944	1032	464	50.3	85
		2	634	745	362	38.6	88
		Total	766	886	438	33.3	173
	19	1	571	584	170	17.6	93
		2	488	500	132	13.4	96
		Total	533	541	157	11.4	189
	20	1	561	641	270	27.9	94
		2	456	517	237	24.3	95
		Total	526	578	261	19.0	189
	21	1	821	881	287	29.4	95
		2	636	670	220	22.5	96
		Total	720	775	276	20.0	191
	22	1	907	1016	360	39.2	84
		2	696	787	296	30.2	96
		Total	791	894	346	25.8	180
	23	1	511	514	161	16.6	94
		2	462	478	151	15.4	96
		Total	481	496	156	11.3	190

Chapter 3

◆

Can non-signers differentiate between SLN signs, emblems and fidgeting?

The experiment reported here aims to determine whether there are visible differences between lexical signs, emblems (i.e. highly conventionalized gestures) and fidgeting. To focus on the appearance of the movements instead of their meaning we selected non-signers as participants. They were shown movies with a single lexical sign, an emblem, or a fidgeting movement and were instructed to press the spacebar as soon as they judged the movement to be a sign. Participants were found to be well able to let the fidgeting movements pass without pressing, but to press almost equally often in response to lexical signs as to emblems. Emblems that were commonly known in the Netherlands elicited pressing less often than emblems not commonly known. However, this difference was entirely due to four emblems with an offensive meaning which many participants did not judge to be SLN signs. These results show that, based solely on appearances, non-signers are typically not able to discriminate signs from emblems, but they are typically able to discriminate between fidgeting and movements that are intended to communicate (emblems and SLN signs).

This chapter is based on a paper presented at the Gesture Workshop 2009: Gesture in Embodied Communication and Human-Computer Interaction, February 25 - 27, in Bielefeld, Germany:
Authors: Arendsen, Jeroen, Andrea J. van Doorn, & Huib de Ridder.

3.1. Introduction

This paper presents a study on the perception of Sign Language of the Netherlands (SLN) signs and other types of gestures and hand movements. The research has two main goals. First, to get a better understanding of how humans observe other humans' gestures and signing, as part of their extensive repertoire of body movements. Second, to gather clues that may be useful for programming computers to detect and recognize gestures and signs, thus contributing to a related project on automatic sign language recognition (Lichtenauer et al. 2007; Spaai et al. 2008). For both goals it is interesting to know to what extent the appearances of sign language signs and other gestures are the same or not. Each similarity between them may result in speeding up technological advance, when algorithms aimed at solving certain problems for sign language recognition can also be applied productively to gesture recognition and vice versa. For example, Xiong and Quek (2006) proposed ways to deal with 'oscillatory gestures' which could also be useful for correctly interpreting signs that have a repetitive movement with a variable number of repetitions. The other way round, Vogler and Metaxas (2004) proposed a method to recognize in parallel different phonological aspects of signs, an approach that might also be interesting for distinguishing gestures from each other. For overviews of sign language recognition, see Von Agris et al. (2008), Ten Holt et al. (2006) and Ong and Ranganath (2005). For overviews of gesture recognition see Mitra and Acharya (2007) and Pavlovic et al. (1997).

In line with the definition of Kendon (2004), we consider gestures first and foremost to form a broad category of bodily motions produced with an intention to communicate (Melinger & Levelt 2004) and perceived as such (Kendon, 1994). This implies that gestures are contrasted on the level of intentions with other body movements such as practical actions aimed at achieving some noticeable result, moving or repositioning to direct attention elsewhere, and motion that is 'just fidgeting' without a clear intention; see Baldwin and Baird (2001) for an overview of what is known about how people (learn to) discern intentions in dynamic human action. Obviously, sign language falls under this definition of gesture. According to Kendon (2004) gestures stand out visibly from other bodily motion through their appearance of deliberate expressiveness. Typically, one does not need to be familiar with the meaning of a gesture to see that it is a gesture (see also Bucci et al. (2008) for a discussion of cases where people show a severely decreased ability to discriminate between gestures and 'incidental movements' and are said to suffer 'delusions of communication'). Arendsen et al. (2007) found support for this statement in an experiment where people, signers as well as non-signers, were asked to watch movies containing signs (as a particular class of gestures) and/or fidgeting movements and to press the spacebar as soon as they saw the beginning of a sign. Surprisingly, non-signers who were not familiar with SLN were equally capable as signers at discriminating signs from fidgeting. This finding raises the question whether the non-signers were really able to identify signs or that they pressed because they could see that the signs are intended to communicate and fidgeting movements not. If the latter is true, then non-signers may not see differences between signs and other gestures but they should be able to discriminate between different types of gestures and fidgeting. In the present study, this is investigated by including not only signs but also 'emblems', being highly conventionalized gestures (Ekman & Friesen 1969).

In the current experiment the focus is on isolated lexical SLN signs, in their citation form. As stated above these lexical signs will be compared with a set of highly conventionalized gestures. In any given hearing culture there are gestures that have well-understood meanings and can substitute words. Kendon (1995) provided examples of how these emblems or 'quotable gestures' (Kendon 1984, 1992) function in a conversation. Compared with the large lexicon of a signed language, the set of emblems is rather small, even in a region such as Southern Italy with a lively gesture tradition (Morris et al. 1979; Kendon 1995). Yet it is interesting to include emblems in the present study. The main reason is that, of all types of gestures, emblems are arguably most similar to lexical signs because of their conventionality and their ability to function as words (see Kendon (1995) for an analysis of the graded differences between emblems and other, less conventionalized gestures such as illustrators). This suggests that, if visible differences between lexical signs and emblems can be demonstrated for non-signers, it is reasonable to assume that there are also visible differences between lexical signs and all other types of gestures. In other words, signs will stand out as a class of gestures with a unique appearance.

The aim of the current experiment is to determine whether non-signers are able to see that movements are SLN signs, in the sense of being able to distinguish lexical signs from other highly conventionalized gestures (i.e. emblems) and fidgeting. We show movies of signs, emblems and fidgeting movements to non-signers and instruct them to press the spacebar when they discern a SLN sign. If non-signers can distinguish the lexical signs from the emblems they will press the spacebar only with the signs. If they 'simply' classify all movements that appear to have an intention to communicate into SLN signs they will press the spacebar with the emblems as well as with the signs, but not with the fidgeting movements.

3.2. Method

Participants

Participants were nine Dutch speaking people without significant hearing limitations and no knowledge of any sign language. They volunteered to attend a single session of about 20 minutes. One participant was a woman. Participants' age ranged from 18 to 48 years with an average of 28 years.

Material

Test material consisted of short movies of 20 emblems, 20 SLN signs, and 20 fidgeting movements. Their description plus label as well as snapshots from the movies can be found in Figure 7 to Figure 9 for emblems, signs and fidgets respectively.

Emblems - The emblems were taken from Morris et al. (1979) who studied the origin and distribution of their forms and meanings in Europe. There is one exception: the head toss was replaced by a temple tap (emblem 13) because it was without manual action. Morris et al. (1979) found that the emblems were all well known to have certain meanings, as testified by informants, in some regions of Europe (such as Southern Italy), but not in other parts of Europe. In the Netherlands, where the present study was carried out, some of the gestures are commonly known and others only rarely or not at all, according to the findings of Morris et al. (1979). A second, more recent source (Andrea & De Boer 1993) was used to check the meaning of the gestures in the Netherlands. This resulted in twelve emblems that are commonly known in the Netherlands. They are indicated with the symbol "NL" in Figure 7. In addition, in a recent SLN dictionary (Nederlands Gebarencentrum, 2002) eleven emblems were found to be in use as lexical signs. Their lexical SLN meaning turned out to be similar or equal to the meaning of the emblem as used in the hearing host culture of the Netherlands (with the exception of the flat-hand flick). They are indicated by the symbol "SLN" in Figure 7.

Lexical SLN Signs - The 20 lexical SLN signs were selected randomly from 32 signs used by Arendsen et al. (2007) in a previous experiment. The following three lexical SLN signs from Figure 8 were also described by Andrea and De Boer (1993) as common gestures in the Netherlands with a corresponding meaning: TO LOOK, BIKE, and CAR.

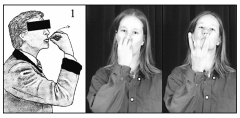
Fidgeting movements - The 20 fidgeting movements, shown in Figure 9, were an expansion of the set previously used by Arendsen et al. (2007). Two fidgeting movements are also mentioned by Andrea and De Boer (1993) as common gestures in the Netherlands: The [Arm Fold] as a gesture of defiance and the [Hand Squeeze] as an expression of glee. However, both are said to require accompanying exaggerated facial expressions and postures (Andrea & De Boer 1993) or else people will not interpret them as intentionally communicating movements but as fidgeting movements.

Recording - The movies were recorded with a hearing woman and a hearing man. Neither actor was a fluent signer, but both actors knew how to produce the signs. Each actor produced 10 SLN signs, 10 emblems, and 10 fidgeting movements (see Figure 7 to Figure 9). Both actors practiced the movements before recording. A third actor was recorded producing an additional set of movies of three signs, three emblems and three fidgeting movements chosen randomly from the used set. These were used to let participants practice.

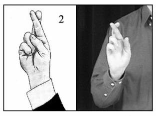
We aimed at isolating the manual features of the signs, emblems and fidgeting movements. Therefore, during the recording of the movies, the actors were instructed not to use mouthing, to continuously look into the direction of the camera lens, while keeping a straight face.

A digital camera (PAL) was used with a resolution of 720*576 pixels. The colors of the clothing (blue), the background (black), and the table behind which the signer was seated (white) provided good color contrasts including the color of the actors' skin. Diffuse lighting was created to avoid sharp, strong shadows.

The movies had an average duration of 3100 ms (range 2300 to 4460, with one outlier of 6100). The beginning and ending of the movements in the movies were coded by two coders. The beginning was defined as the moment the hands left their rest position and the ending as the moment they returned there. The movements had an average duration of 2500 ms (SD = 600 ms). The durations of fidgeting movements (M = 2880 ms, SD = 677 ms) were significantly longer than those of emblems (M = 2160 ms, SD = 406 ms) and signs (M = 2330 ms, SD = 263 ms).

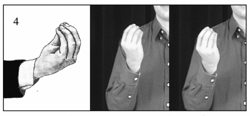


1. Finger tips kiss. The tips of the fingers and thumb are pressed to the lips, lightly kissed, and flicked out. (NL: Praise [1] [2]) (SLN: Magnificent)



2. Fingers cross. The middle finger is twisted over the forefinger, both pointing upwards. (SLN: To Hope; when made with 2 hands)

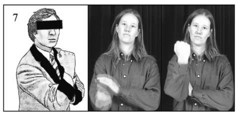
3. Nose thumb. One thumb touches the tip of the nose, with the fingers spread out in a fan and pointing upwards. The fingers are waggled. (NL: Mockery [1] [2])



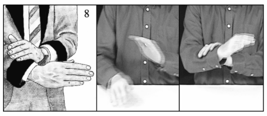
4. Hand purse. The fingers and the thumb of one hand are brought together in a point facing upwards. The hand is moved slightly. (NL: emphasis [1])

5. Cheek screw. A forefinger is rotated against the cheek, as if screwing something into the face. (SLN: Shy)

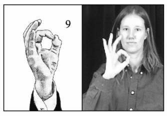
6. Eyelid pull. The forefinger tugs at the lower eyelid while looking straight at the companion. (NL: keeping an eye on it/you [1] [2])



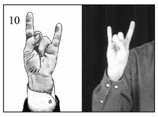
7. Forearm jerk. A fist is jerked forcibly upwards. The other hand is slapped down on the upper-arm, checking the upward movement. (NL: Fuck you [1])



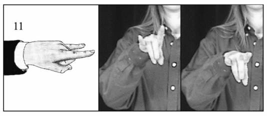
8. Flat-hand flick. A hand is flicked up into the air. This movement is checked by the other hand, copping down on to the right wrist. (NL: 'He is gay' [2]) (SLN: Leave)



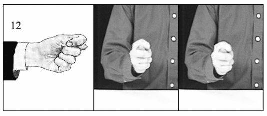
9. Ring. The hand is held up, palm out, with thumb and forefinger forming a circle. (NL: Very nice [1] [2]) (SLN: (Very) Nice)



10. Vertical horn sign. The hand is held up, forefinger and little finger extended vertically. The other fingers are restrained.



11. Horizontal horn sign. The hand is pointed forward, forefinger and little finger extended horizontally. The other two fingers are restrained

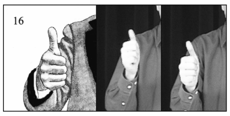


12. Fig. The tip of the thumb protrudes from between the first and second fingers of a fist. (NL: To Fuck [1] [2]) (SLN: To Fuck).

13. Temple tap. The forefinger is tapped a number of times to the temple. (NL: Smart [2]) (SLN: location temple refers to the mind; Wise)

14. Chin flick. The fingers are flicked out, their backs brushing against the underside of the chin. (SLN: Sick of it; with facial expression)

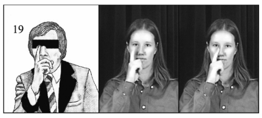
15. Cheek stroke. The thumb and forefinger are placed one on each cheek-bone and then gently stroked down each cheek. (NL: Skinny & sick [1]) (SLN: Pale)



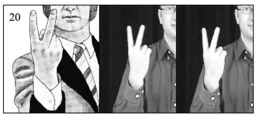
16. Thumb up. The clenched hand is extended, with the thumb vertically erect. (NL: OK, good [1] [2]) (SLN: Good)

17. Teeth flick. The thumb-nail is placed behind the lower edge of the upper incisors and then forcibly jerked forwards, making a clicking sound.

18. Ear touch. The ear is deliberately touched with the fingers of one hand. The ear-lobe is taken between the thumb and forefinger and tugged several times.



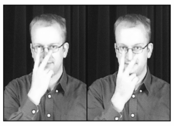
19. Nose tap. The forefinger is placed vertically alongside the nose and then gently taps it several times. The side of the nose is tapped with the tip of the forefinger.



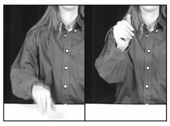
20. Palm-back v-sign. The hand, palm inwards with forefinger and middle finger forming a V shape, is jerked upwards. (NL: Victory [2]) (SLN: Two; no upward jerk)

Figure 7. Emblems used in the experiment. Each emblem is provided with a label, description and depiction given by Morris et al. (1979). If an emblem is commonly used in the Netherlands then this is indicated with the symbol "NL" and the meaning is given along with the source of this meaning, with [1] indicating Morris et al. (1979) and [2] indicating Andrea and De Boer (1993). If an emblem is also used as a lexical SLN sign then this is indicated with the symbol "SLN" and the lexical SLN meaning is given.

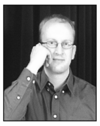
21. **DADDY** (PAPA)



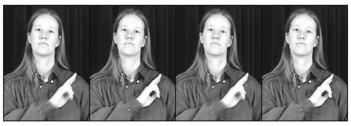
22. **LOOK (KIJKEN)**
(NL: (You should) look out [2])



23. **SAND** (ZAND)

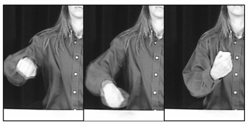


24. **FEVER** (KOORTS)

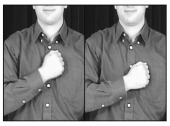


25. **BROTHER** (BROER)

26. **DIRTY** (VIES**)**



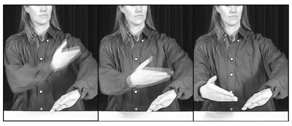
27. **SHOVEL** (SCHEP)



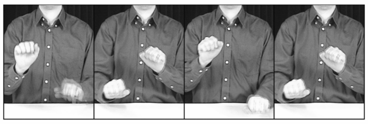
28. **BATH** (BAD)

29. **BATHROOM** (WC)

30. **DRAW** (TEKENEN)

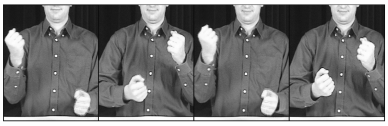


31. **SOUP** (SOEP)

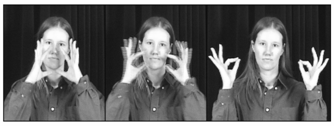


32. **BIKE** (FIETS)
(NL: To ride a bike [2])

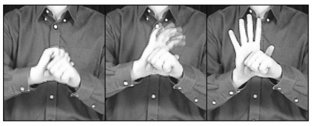
33. **COW** (KOE)



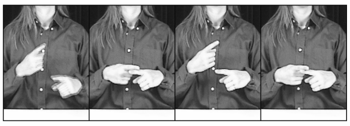
34. **CAR** (AUTO)
(NL: To ride a car [2])



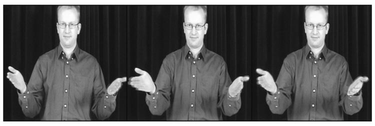
35. **CAT** (POES)



36. **HEDGEHOG** (EGEL)

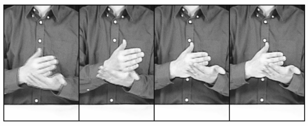


37. **HORSE** (PAARD)



38. **TIDY UP** (OPRUIMEN)

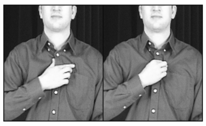
39. **TREE** (BOOM)



40. **SANDWICH** (BOTERHAM)

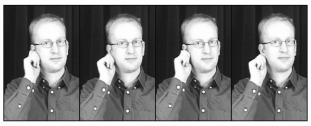
Figure 8. Lexical SLN signs used in the experiment. Glosses of the signs are given in English (and Dutch) as well as key frames (selected by hand) from the movies. Three of these signs were also mentioned by Andrea and De Boer (1993) as commonly used gestures. These signs are indicated with the symbol "NL" and their meaning in the Netherlands is given.

41. [Arm Fold]. Fold arms. (NL: Defiance (defiant facial expression and posture required))



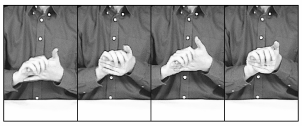
42. [Chest Scratch]. Scratch torso through clothing.

43. [Chin Rub]. Rub chin with index finger and thumb twice.



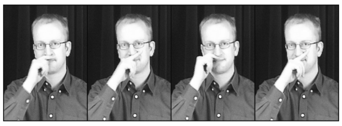
44. [Ear Grab]. Grabbing earlobe between indexfinger and thumb.

45. [Hair Brush]. Brush hair with fingers.



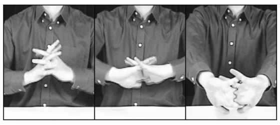
46. [Hand Squeeze]. Clasp and rub hands twice. (NL: Glee (gleeful facial expression required))

47. [Lip Touch]. Touch the corner of the mouth briefly as if wiping something off.

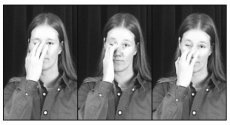


48. [Nose Rub]. Rub twice under the nose with index finger or closed hand.

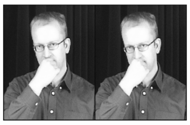
49. [Table Drum]. Drum twice on the tabletop with knuckles or fingers.



50. [Knuckle Crack]. Lock fingers of hands and crack knuckles.

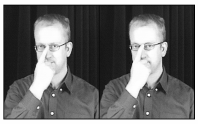


51. [Eye Rub]. Rub eye with index finger or closed hand.

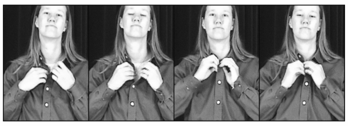


52. [Teeth Clean]. Touch teeth with fingernails, as if trying to get something out from in between.

53. [Nail Clean]. Use one thumb to remove dirt from under the other thumb's fingernail.



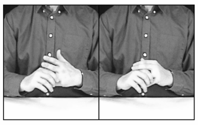
54. [Spectacle Push]. Use the index finger to push the glasses higher up on the nose.



55. [Button Up]. Fasten the top button of the shirt.

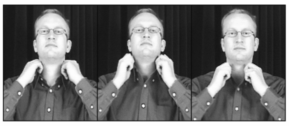
56. [Time Check]. Check wristwatch.

57. [Button Down]. Unfasten the top button of the shirt.



58. [Ring Twist]. Twisting a ring, trying to get it in a better position.

59. [Necklace Adjust]. Adjust necklace position.



60. [Clothes Adjust]. Straighten collar.

Figure 9. Fidgeting movements used in the experiment with a [label] between square brackets which is used to indicate a fidgeting movement in the remainder of the paper. The description that was used to instruct the actors during recording is given. Two of the fidgeting movements, the [Arm Fold] and [Hand Squeeze] are mentioned by Andrea and De Boer (1993) as commonly used gestures, but in both cases a specific facial expression is required to mark the movements as that gesture.

Procedure

Participants were seated at a table with a laptop. At the start participants provided some personal data, such as their name, gender and age. Then participants were shown the following on-screen message: "You are going to watch a series of movies in which an actor is producing some hand movements and you are requested to press the spacebar as soon as you think you see a SLN sign. A movie does not always contain a SLN sign. It may also contain other hand movements, in which case you should not press the spacebar." Next, participants practiced the experimental procedure, as described below, with the practice material. Then, participants performed the actual experiment with the 60 movies presented for each participant in a unique random order. Their response to each movie was recorded in the following way:

1. Before each movie participants were instructed to "Press the spacebar to start the movie. Press the spacebar again as soon as you see a SLN sign."
2. When started, the movie was played full screen. Participants could choose to respond by pressing the spacebar after which the movie immediately stopped or by not pressing the spacebar in which case the movie ran to its end.
3. Next, participants had to either confirm their response or choose to see the same movie again in which case it would not be played directly afterwards but be placed randomly amongst the remaining movies (this was repeated until a confirmed response was recorded).

Finally, after participants had responded to all 60 movies, the program ended, and the participants were debriefed and given the opportunity for comments and questions.

3.3. Results

Table 14 shows the number of times each participant pressed the spacebar per type of movement. The results indicate that participants did not often press the spacebar with fidgeting movements, but that they often pressed the spacebar with emblems, with the exception of one participant (P8). During debriefing this participant explained that he had lived in Italy for several years and recognized many of the emblems. He therefore treated them as 'not signs'. Even if we discard the results for this participant then participants pressed on signs (92% pressed) more often than on emblems (81% pressed; Fisher's Exact Test $p < 0.01$). If participant P8 is included the difference increases to 91% for signs versus 74% for emblems (Fisher's Exact Test $p < 0.01$). Table 14 further shows that, with the exception of participant P8, all participants responded more or less in the same way (the data does not warrant χ^2-tests between subjects, because many cells have an expected count of less than five). There is only one participant (P9) who pressed more often in response to the presentation of emblems than to the presentation of signs.

The results have also been split up per actor in Table 14, to check if production differences between actors had an effect on people's perception of the different types of movies and whether there is an interaction with movement type. A two-way ANOVA on the amount of times pressed by each participant confirms that movement type had a main effect ($F_{(2, 48)} = 166.19$, $p < 0.01$) but that the actors did not have an effect on the results ($F_{(1, 48)} = 0.40$, $p = 0.53$), nor is there an interaction effect ($F_{(2, 48)} = 1.63$, $p = 0.21$). Overall, participants responded in the same way to the movies by actor G and actor J, yet there is a noticeable difference in how often participants pressed the spacebar for the fidgets made by actor J (9%) and those made by actor G (1%) which is significant if it is singled out (ANOVA $F_{(1, 16)} = 11.53$, $p < 0.01$). However, this difference (9% versus 1%) is small compared to the differences between fidgeting movements and either signs or emblems. Furthermore, the difference is almost entirely caused by a single fidget, as can be seen in Table 17. Six participants pressed the spacebar when they saw the [Ear Grab] made by actor J.

Table 14. Responses of each participant to the movies of each movement type (respectively lexical signs, emblems, and fidgeting movements). Indicated are the number of times that a participant pressed the spacebar (because they judged the movement was a sign). These results have also been split up between movies with actor G and actor J. The number of movies is indicated with "N".

	Both Actors, N=20			Actor G, N=10			Actor J, N=10		
Participants	**Sign**	**Emb**	**Fid**	**Sign**	**Emb**	**Fid**	**Sign**	**Emb**	**Fid**
P1	20	14	1	10	7	0	10	7	1
P2	19	19	1	9	9	0	10	10	1
P3	19	18	1	10	9	0	9	9	1
P4	19	17	1	10	10	0	9	7	1
P5	19	14	1	10	8	0	9	6	1
P6	19	14	0	10	8	0	9	6	0
P7	17	16	3	8	8	1	9	8	2
P8	17	4	1	9	3	0	8	1	1
P9	15	18	0	9	9	0	6	9	0
Total	164	134	9	85	71	1	79	63	8
%	91%	74%	5%	94%	79%	1%	88%	70%	9%

Table 15. Responses to each lexical SLN sign across all nine participants. The number of participants who 'pressed' the spacebar is given (with a maximum of nine in all cases). These results are also split up between movies with actor G and actor J. Signs that are also in use as common Dutch gestures are indicated with '*'.

Signs by Actor G	**Pressed**	**Signs by Actor J**	**Pressed**
SHOVEL (SCHEP)	6	FEVER (KOORTS)	5
SAND (ZAND)	8	TIDY UP (OPRUIMEN)	7
BATHROOM (WC)	8	LOOK (KIJKEN) *	7
DADDY (PAPA)	9	DRAW (TEKENEN)	8
BROTHER (BROER)	9	BIKE (FIETS) *	8
SOUP (SOEP)	9	CAR (AUTO) *	8
COW (KOE)	9	DIRTY (VIES)	9
CAT (POES)	9	BATH (BAD)	9
HORSE (PAARD)	9	HEDGEHOG (EGEL)	9
TREE (BOOM)	9	SANDWICH (BOTERHAM)	9
Total	85/90		79/90
Average %	94%		88%

Table 16. Responses to each emblem across nine participants. The number of participants who 'pressed' the spacebar is given. Emblems that are also in use as lexical SLN signs are indicated with '*'. Emblems that have a commonly known offensive meaning in the Netherlands are indicated with '♪'.

Actor	Known Emblems	Pressed	Actor	Unknown Emblems	Pressed
G	Forearm Jerk ♪	3			
J	Fig * ♪	4			
G	Nose Thumb♪	5			
J	Flat-Hand Flick * ♪	5			
J	Hand Purse	6	J	Fingers Cross *	6
J	Eyelid Pull	6	J	Vertical Horn	6
G	Ring *	7	G	Teeth Flick	7
J	Thumb Up *	7	J	Chin Flick *	8
G	Temple Tap *	7	G	Cheek Screw *	8
J	Palm-Back V-Sign *	7	J	Ear Touch	8
G	Finger Tip Kiss *	8	G	Horizontal Horn	9
G	Cheek Stroke *	8	G	Nose Tap	9
	Total	73/108			61/72
	Average %	68%			85%

Table 17. Responses to each fidgeting movement across nine participants. The number of participants who 'pressed' the spacebar is given.

Fidgets by Actor G	Pressed	Fidgets by Actor J	Pressed
[Nail Clean]	1	[Ear Grab]	6
[Arm Fold]	0	[Hand Squeeze]	1
[Chin Rub]	0	[Nose Rub]	1
[Hair Brush]	0	[Chest Scratch]	0
[Lip Touch]	0	[Nuckle Crack]	0
[Table Drum]	0	[Teeth Cleaning]	0
[Eye Rub]	0	[Spectacle Push]	0
[Button Up]	0	[Time Check]	0
[Button Down]	0	[Ring Twist]	0
[Necklace Adjust]	0	[Clothes Adjust]	0
Total	1/90		8/90
Average	1%		9%

Table 18. Totals of responses across nine participants (including P8) and across eight participants (excluding P8) to each movement type.

	% Pressed	
Movement type	**Inc. P8**	**Ex. P8**
Signs	91%	92%
Emblems that are not commonly known	85%	91%
Emblems that are commonly known and not offensive	78%	86%
Emblems that are commonly known and offensive	47%	53%
Fidgeting	5%	5%

Participants did not press the spacebar equally often with each of the lexical SLN signs as can be seen in Table 15. Eleven signs were classified as a sign (spacebar pressed) by all nine participants. FEVER was most often classified as not a sign (spacebar not pressed), and this is in line with the results of Arendsen et al. (2007) in which a similar set of signs was used. They found that FEVER was the sign that was most often 'missed' (spacebar not pressed) when participants were instructed to 'press the spacebar as soon as you see the beginning of a sign'. The three signs that are also known as Dutch gestures (i.e. LOOK, BIKE and CAR), according to Andrea and De Boer (1993), do not appear to have elicited exceptional response behaviour.

The number of participants that classified an emblem as a SLN sign (pressed the spacebar) is given in Table 16. It is striking that the four emblems which were least often thought to be a SLN sign (the Forearm Jerk, the Fig, the Nose Thumb, and the Flat-Hand Flick) are not only four known Dutch gestures but also impolite, even vulgar gestures at which one can take offense. They are largely responsible for the difference in responses between commonly known (68% pressed) and unknown emblems (85% pressed) (Fisher's Exact Test p = 0.01).

If we compare the responses to the emblems that are not commonly known with the responses to lexical SLN signs (excluding LOOK, BIKE, and CAR) then there is no significant difference in how often participants classified them as SLN signs (Emblems: 85% vs. Signs: 92%; Fisher's Exact Test p = 0.101).

All of the fidgeting movements were classified by all participants or all but one, as 'not a SLN sign' (spacebar not pressed), with the exception of the [Ear Grab], see Table 17. The [Ear Grab] was even thought to be a sign more often than FEVER, one of the lexical SLN signs.

3.4. Discussion

Table 18 shows a summary of the results split up per type of movement. These results are given both including and excluding participant P8 who showed atypical behaviour, see above. With P8 excluded there is no difference between how often participants pressed in response to signs and to unknown emblems. Emblems that are commonly known in the Netherlands were less often judged to be a sign than unknown emblems which suggests participants used their knowledge of those emblems in this experiment. Offensive emblems show a remarkably low rate of pressing the spacebar. Fidgeting movements were judged to be signs in only five percent of all cases.

The perception of an intention to communicate - In a previous study of Arendsen et al. (2007) fidgeting movements were almost never classified as SLN signs by signers and non-signers alike. In the present experiment, where a third category of movements (emblems) was introduced next to fidgeting movements and lexical signs, non-signers again almost never judged fidgeting movements to be SLN signs. These findings, together with the finding in the current experiment that emblems were classified very often as SLN signs, are consistent with the idea expressed by Kendon (2004) that people are able to classify movements as those that are and those that are not intended to communicate, based only on their appearance. The way people do it cannot be deduced from our data. One possible interpretation might be that people are able to do this because they are familiar with the fidgeting movements. Following this strategy they may simply classify all movements that are not familiar fidgeting movements as movements that are intended to communicate. Yet, it is also possible that the intention to communicate is visible in the appearance of the signs and emblems themselves. Perceiving communicative intention may arise from a combination of visible factors such as the amount of visible action, a high complexity of the movement in terms of frequent motion boundaries (Rubin & Richards, 1985), the absence of some noticeable result branding the movement a practical action, or the iconicity that the movement affords. The latter refers to the interpretation of a movement using iconic strategies, such as the embodiment of some object, the enactment of some action, or the 2D trace of an object's silhouette or 3D model of its shape (Müller 1998). In case of a location on the body or the face the use of typical movement patterns for gestures (such as the repeated tapping of DADDY) or typical handshapes (such as the 'Y'-hand in DIRTY) that are atypical for fidgeting may also contribute to the 'gestural appearance'. If it is true that the perception of communicative intention arises from the factors described above then it may be logical that four out of nine participants did not press the spacebar with the sign FEVER. This sign is made by placing a hooked hand with the back of the fingers on the cheek and keeping it still for about a second. FEVER is made on the face (where many fidgeting movements are also made), lacks movement, and does not afford an obvious iconic interpretation. Therefore, the only visible factors that can contribute to the movement's communicative appearance are the shape of the hand and the way it is held at its location. It seems reasonable that for some observers this was not enough to see a communicative intention.

One fidgeting movement, an [Ear Grab], was exceptionally often mistaken for a sign. We checked the movie for any anomalies but we did not find anything strange that could explain the finding. However, there was an emblem in the material, the Ear Touch, which was very similar to the [Ear Grab] fidgeting movement[1], and this could have influenced participants' responses in that they were inclined to classify it as a sign as well.

No unique appearance for signs - A unique appearance for lexical signs as a specific class of gestures could not be demonstrated for non-signers with our material. The finding that SLN signs and emblems *not* commonly known in the Netherlands were equally often classified by non-signers as SLN signs, suggests that people are not able to distinguish lexical SLN signs from emblems if they do not know the (meaning of the) movements, and have to classify them based on their appearance only. We have also documented cases where emblematic gestures are also in use as SLN signs and vice versa. Further research will be needed to clarify whether other types of gestures are also indistinguishable from emblems and lexical signs based on appearances only.

This finding may have implications for studies that involve differentiation between lexical signs and other types of gestures, for example with regard to the classification of experimental material. Sign language is typically contrasted with 'co-speech gestures'[2] or 'non-linguistic gestures' in studies involving both sign language and gesture (e.g. Corina & Knapp 2007, MacSweeney et al. 2004, Emmorey 1999). However, to be able to classify a movement as, for example, an American Sign Language (ASL) sign a coder has to know ASL (or use a dictionary of signs) and recognize a gestural movement as having a certain meaning in ASL, at the same time classifying all other movements, that they do not recognize as having a specific ASL meaning, as 'non-linguistic' gestures. Hence, classification is determined by coders' knowledge of ASL and not by intrinsic characteristics of the movements themselves.

Commonly known and offensive emblems - With emblems commonly known in the Netherlands, participants may use their knowledge of those gestures and classify them as 'not a SLN sign'. Our results suggest that this indeed sometimes happened. This appeared to occur especially when the participants saw offensive gestures. It is remarkable that people seemed to think that vulgar or offensive gestures were probably not signs, especially since some of those gestures are actually in use as SLN signs. Perhaps it is believed that the language of a challenged minority (as is the common layman's perception of deaf and hard of hearing people) is above such profanity (one way of dealing with something unknown is to put it on a pedestal and glorify it).

Implications for automatic gesture and sign recognition - Our findings suggest that, on the surface, emblems and lexical signs resemble each other. Although future research is necessary to investigate if there are perhaps differences on a more detailed level between the appearances of signs and other gestures, we believe that it is very likely that algorithms for the recognition of lexical signs can be applied to the recognition of a set of conventionalized gestures (and vice versa). We currently find no reason to assume a priori that such transfers would not be worthwhile.

For automatic gesture or sign language recognition the 'spotting' of relevant gestures in a stream of continuous data that may also include irrelevant movement (such as fidgeting) is still an unsolved problem, although some progress has been made (Roh et al., 2008; Junker et al. 2008). The results of the current study suggest that, in contrast to movements such as fidgeting, signs and emblems share certain characteristics that make observers see that they are intended to communicate. In other words, signs and emblems have the appearance of 'gestures' in the broad sense. It remains a topic for further investigation whether this intentionality is visible because of people's sensitivity to certain general movement characteristics or whether people rely on learned repertoires of movements that are or are not intended to communicate. Both strategies, which are not mutually exclusive, may not only explain our current findings, but could also be applied to automatic 'gesture spotting'. General rules or criteria could be used to specify gestures and fidgeting, and a set of fidgeting movements as 'non-gestures' could be included in training data.

Notes

1. Both the [Ear Grab] fidgeting movement and the Ear Touch emblematic gesture were performed by the same actor.
2. The term 'co-speech gestures' is a fairly loosely defined term used by gesture researchers to refer to gestures that are used during talking, together with speech. Although emblematic gestures or even signs can also be used during talking the term 'co-speech gestures' is usually wielded to exclude such highly conventionalized gestures and instead focus on less conventionalized gestures.

Chapter 4

♥

When do people start to recognize signs?

We studied how much time people need to recognize a sign as it unfolds. Deaf and hearing signers were presented movie fragments with sequences of restpositions, fidgets, and monomorphemic signs. They watched these movies at normal playing speed and had to respond as soon as they recognized a sign, which they were able to do after around 850 ms, counting from the beginning of the sign. By subtracting participants' reaction times to seeing a motion boundary (average of 310 ms) we estimate that confident sign recognition starts after around 540 ms, in the sense that the necessary information has become available in the signal. If we think of a sign as consisting of three main movement phases (i.e. preparation, stroke and recovery) lexical recognition starts about 220 ms after the onset of the stroke. In a comparison with findings from an earlier experiment on detecting the beginning of a sign (Arendsen et al., 2007) lexical recognition was found to take about 90 ms longer than detection.

This chapter will be published as:
Arendsen, Jeroen, Andrea J. van Doorn, & Huib de Ridder. When do people start to recognize signs? *Gesture* (in press).

4.1. Introduction

Like any other gesture, a lexical sign of a signed language, such as Sign Language of the Netherlands (SLN), is a biological motion that unfolds in time. It has a beginning and an ending and as we move our hands (or other body parts) the meaning of the gesture becomes apparent at some point to the knowledgeable observer. Lexical signs have highly conventional meanings and together form the lexicon of a signed language, while other gestures may have meanings that are less conventional or that depend on context, such as co-occurring speech. Understanding the time course of the recognition of the (lexical) meaning in relation to the unfolding form of a gesture will enable a more exact interpretation of the results of gesture and sign language studies that involve both the recognition of meaning and other time-related phenomena, for example brain activity measurements or co-occuring speech events (e.g. Quek et al., 2002; Corina & Knapp, 2006; Wu & Coulson, 2007; Willems et al., 2007).

Knowing how fast humans can recognize a sign and respond to it is also useful for the development of efficient automatic sign recognition technology, which is the aim of one of our related projects (Lichtenauer et al., 2007; Spaai et al., 2008). Current automatic sign recognizers tend to wait until a sign has fully ended before starting the recognition process and their response feels delayed, even if their recognition process is fast. Certain delays may be acceptable but as long as we lack a human benchmark we cannot even determine the size of the experienced delay. We suspect much is to be gained if automatic recognition could start sooner, preferably immediately after the moment human signers start to recognize the lexical meaning.

Recently, Arendsen et al. (2007) have shown that people can discriminate fidgeting movements (fidgets) and signs based on appearance, even if one does not know sign language. This was based on the results of an experiment in which signers as well as non-signers were asked to watch movies containing fidgets and/or signs and then press the spacebar as soon as they saw the beginning of a sign. Even though all participants were able to detect the beginning of a sign quickly and reliably, non-signers were about 120 ms slower than signers. Furthermore, response time was strongly related to what is here termed the nucleus (stroke and/or hold) (following Kendon 2004: 112) and was influenced by three factors: it was shorter for signs with a highly marked handshape, signs that were one-handed, and if a sign was preceded by a fidget (Arendsen et al., 2007).

In the present study, we investigate whether the observations for sign detection also hold for the moment signers start to recognize the lexical meaning of a sign and what the relation is between sign detection and lexical recognition. To this end, an experiment was set up to gather response time (RT) measurements, in a similar manner as Arendsen et al. (2007), only with a lexical recognition task for participants. The same material was used.

The paper is organized as follows. First, a brief review is given on the timing of lexical recognition. Next, in the first study, the experiment on lexical sign recognition is described. Subsequently, in the second study, the results will be compared directly with those from the detection experiment (Arendsen et al., 2007) to gain more insight into the relation between lexical recognition and sign detection. Implications of the findings are discussed at the end of each study and summarized in the conclusions.

4.2. Time course of lexical sign recognition: literature review

Grosjean (1981) and Emmorey and Corina (1990) performed studies on the time course of lexical sign recognition using a *gating* paradigm (Grosjean, 1996): A sign is presented in segments of increasing duration starting at the beginning of the movie. Participants guess the meaning of the sign after each pass and rate their confidence. Grosjean (1981) not only collected data in this way but also offered an insightful analysis of the top-down and bottom-up processes involved in sign and word recognition. For example, context, such as a sentence, was assumed and found to help recognition of words (Grosjean, 1980) and signs (Clark & Grosjean, 1982) by its limiting effect on the number of likely candidates. The bottom-up process was assumed to consist mainly of narrowing-in on the correct lexical item. This implies that people continuously guess the meaning of a signal (sign or word) until at some point they guess its correct meaning but at a low confidence level (usually level 3 on a scale from 1, 'a wild guess', to 6 "very sure"). This is called the *isolation* point. Subsequently, the signal is analysed further and confidence about the guess rises until the choice is accepted and *recognition* is said to have occurred (level 5 is reached).

For mono-morphemic signs Grosjean (1981) found an average isolation time of about 400 ms with a fairly large range of 198 ms to 585 ms. He also examined the narrowing-in process in more detail by looking at the errors made in participants' guesses. With progressive gate widths certain errors were not made anymore, from which Grosjean inferred that a certain *feature* of the sign was isolated, thereby ruling out those errors typically related to lacking such a feature. Measured in this way, the handshape, the location, and the orientation of a sign were all isolated on average after about 320 ms, whereas the 'movement' was isolated after about 390 ms. Movement isolation was found to be a near perfect predictor ($r = 0.98$) of sign isolation times.

Emmorey and Corina (1990) reported, for mono-morphemic signs, an average isolation point of about 240 ms and an average recognition point of 310 ms. For morphologically complex signs (with roots and affixes) they found longer isolation times. Furthermore, Emmorey and Corina (1990) studied the influence of three sign characteristics on lexical recognition: Marked handshapes were found to lead to shorter isolation and recognition times (for late signers but not for native signers), signs made on the face had longer isolation and recognition times than signs made in neutral space, but no difference was found between signs with or without a handshape change. Furthermore, Emmorey and Corina (1990) studied, in the same manner as Grosjean (1981) after how much time features were isolated. They found that location and orientation were isolated after about 146 ms, handshape after about 172 ms, and movement after about 238 ms, which coincided with total sign isolation.

Both Grosjean (1981) and Emmorey and Corina (1990) pointed out that signers require less of a sign (Grosjean mentions 51% and Emmorey and Corina mention 34%) than listeners require of a word. Words are isolated on average after 330 ms, which is 83% of their duration (Grosjean, 1980). This difference was contributed by the near-simultaneous availability of phonological features in ASL signs: handshape, location, orientation and (a little later) movement. The first three are all available at the beginning of the nucleus of a sign (Kita et al. 1998). In case these are not enough by themselves, the availability of the movement feature is almost always found to immediately disambiguate the lexical entry being signed.

Emmorey and Falgier (2004) measured lexical recognition response times to make sure their stimulus signs had equal recognition times. They report response times with an average of about 926 ms (SD = 237 ms). These times are much longer than the isolation or recognition times found by Grosjean (1981) and Emmorey and Corina (1990). Emmorey and Falgier (2004) do not provide details on their measurement procedure. We assume that people observed the signs and responded to them under fairly 'normal' conditions. In that case longer times are to be expected for several reasons. For example, in gating studies the information from previous segments is repeated in all subsequent segments, and after the presentation of a segment participants can take as much time as they need to process the information they saw and to respond. Under more normal conditions of observation the information is presented only once and the time required for information processing and/or memory access will delay the response. Furthermore, producing a motor response also requires a certain 'reaction time' which adds to the delay.

In the present study, we aim to gather response times (RTs) for lexical recognition under conditions that resemble normal observation as closely as possible. These RTs will be compared with the results from the gating studies to gain further insight. To facilitate this comparison, participants' reaction times to a controlled visual event will also be measured, to estimate the part of the delay caused by the time necessary to produce a motor response.

4.3. Study 1. Response times for lexical sign recognition

In this study an experiment is described in which response times (RTs) were gathered in a lexical recognition task. We will report how fast signers can recognize the meaning of signs. Our previously mentioned sign recognizer (Lichtenauer et al., 2007) extracts only manual features from the video and not mouthing or facial gestures. To be able to relate our findings to that recognizer, we copied this restriction: only manual features are present in our signed material (see Vogler and Goldenstein (2008) and Von Agris et al. (2008) for approaches to integrate facial features into automatic sign recognition). Our results allowed us to investigate:

1. The effects of participant characteristics,
2. The relations between RTs and the movement phases of signs,
3. The influence of sign characteristics on lexical recognition RTs,
4. The effects of the embedding conditions.

The following participant characteristics are treated in this study: Hearing status, which was shown by Arendsen et al. (2007) to influence response times on sign detection; deaf signers took more time than hearing signers. Parental hearing status will be checked; Emmorey and Corina (1990) found slightly shorter isolation times for *native signers* (deaf parents) than for *non-native* signers (hearing parents). The extent to which a signed language is one's *native* language may influence one's sensitivity to aspects of the language. However, parental hearing status is not the only indicator of such sensitivity. One might also expect shorter response times for participants with an early age of SLN acquisition, a high SLN fluency, and a high usage frequency and these factors will be checked.

The RTs will be related to the movement phases of a sign, specifically to the beginning and the duration of the nucleus (i.e. stroke and/or hold, typically carrying the essential expressive information). For example, Arendsen et al. (2007) found a strong correlation between the onset of the stroke and sign detection RTs.

Sign characteristics have clearly visible consequences for the appearance of a sign and might therefore cause differences in how it is perceived. We studied the following characteristics: handshape-markedness, handedness, hand location, and several movement parameters (path movement, handshape change, hand orientation change, repetition).

To check whether lexical recognition is influenced by embedding conditions we studied the impact of the presence or absence of a preceding fidget (in the same location as the sign or in a different location), and of different rest positions (again in the same location as the sign or in a different location).

Experimental method

Participants - 32 signers participated and provided their age, gender, hearing status (a self-classification as deaf or hearing), parental hearing status (again a self-classification as deaf or hearing), age of SLN acquisition, fluency in SLN (fluent, good, reasonable, bad), usage of SLN (frequent (primarily or daily) or infrequent (regularly or exceptionally)). The participants and their characteristics are described in detail in Appendix 4-A.

The participants were connected to the Dutch Foundation for the Deaf and Hard-of-Hearing Child (Nederlandse Stichting voor het Dove en Slechthorende Kind) in Amsterdam, the Polano school in Rotterdam, or they were students or teachers of the SLN programme at the Hogeschool van Utrecht.

Table 19. Glosses of the 32 SLN signs (in small caps) used in the experiment.

One-handed Signs: Dutch gloss (English gloss)		Two-handed Signs: Dutch gloss (English gloss)	
ZAND (SAND)	VIES (DIRTY)	AUTO (CAR)	OPRUIMEN (TIDY UP)
SCHEP (SHOVEL)	KOORTS (FEVER)	MELK (MILK)	RAAM (WINDOW)
EUROPA (EUROPE)	MAMA (MOM)	FIETS (BIKE)	TELEVISIE (TELEVISION)
TEKENEN (DRAW)	PAPA (DAD)	SOEP (SOUP)	BOOM (TREE)
WC (RESTROOM)	KIJKEN (LOOK)	JARIG (BIRTHDAY)	FEEST (PARTY)
BROER (BROTHER)	KIP (CHICKEN)	PAARD (HORSE)	AANKLEDEN (GET DRESSED)
BAD (BATH)	MIS (MISSED)	BOTERHAM (SANDWICH)	POES (CAT)
AFDROGEN lichaam (TOWEL-OFF body)	TELEFOON (PHONE)	EGEL (HEDGEHOG)	KOE (COW)

Table 20. List of fidgets used in the experiment. [Brackets] are used to indicate fidgets. These fidgets were selected after they were observed to occur in signed interaction.

Fidget	Description
[Lip Touch]	Touch lips with side of closed hands/index finger
[Nose Rub]	Rub hand/index fingers underneath nose
[Chin Rub]	Rub hand/fingers along chin line
[Ear Grab]	Grasp Earlobe
[Hair Brush]	Brush hair with fingers
[Arm Fold]	Fold both arms over each other
[Hand Squeeze]	Squeeze Hands together
[Chest Scratch]	Scratch chest through clothing
[Table Drum]	Wrap fingers/knuckles on tabletop

Table 21. List of rest positions used in the experiment. Two hands on the table is considered to be the 'neutral' rest position and the others 'non-neutral'. The term 'non-neutral' is not intended to refer to an increased readiness to start signing but to an atypical rest position for a seated signer.

Rest position	Description
2H-Table (neutral)	hands resting on table, fingers brought together
1H-Space	hand held floating, elbow on table
1H-Face	head/chin resting on hand
1H-Body	hand on chest, elbow on table

Material - A set of 112 movie fragments (further abbreviated as 'movies') was reused from previous work by Arendsen et al. (2007). To construct the movies we used 32 signs, see Table 19, nine fidgets, see Table 20, and four rest positions, see Table 21.

The set of 32 signs, see Table 19, was chosen by Arendsen et al. (2007) to be representative in the sense that most of the possible variations in surface form should be present. The signs contained variation in handshape-markedness, handedness (one or two-handed), location (face, body, neutral space, or arm), handshape change presence, path movement presence, orientation change presence, and repetition presence. Handshape-markedness refers to the level of 'marking' within the phonological system that is required to describe the handshape (Van der Kooij, 2002). Handshapes were grouped into highly marked, unmarked and those that are in-between. Highly marked are those handshapes that are infrequently used in the SLN Lexicon (Nederlands Gebarencentrum, 2002), and are acquired last (Conlin et al., 2000). Unmarked handshapes are used frequently and learned first. None of the signs required a gaze shift (e.g. as in the SLN sign GOD) or a facial expression, because our aim is to isolate the manual features that contribute to the recognition of signs.

Each of the 16 one-handed signs was recorded five times in the following embedding conditions (to illustrate the embedding conditions the movies with the sign PAPA, which is a repeated tap on the chin with the index finger, have been included in the online edition of this paper[1]):

1. with a neutral rest position (two hands on the table),
   Movie 1. 2H-Table → PAPA. 0.4 Mb
2. with a non-neutral rest position in the same location as the sign,
   Movie 2. 1H-Face → PAPA. 0.4 Mb
3. with a non-neutral rest position in a different location than the sign, [Movie 3]
   Movie 3. 1H-Space → PAPA. 0.4 Mb
4. with a neutral rest position and a preceding fidget in the same location as the sign,
   Movie 4. 2H-Table → [Chin Rub] → PAPA. 0.5 Mb
5. with a neutral rest position and a preceding fidget in a different location than the sign.
   Movie 5. 2H-Table → [Table Drum] → PAPA. 0.5 Mb

The 16 two-handed signs were only recorded in isolation with a neutral rest position to limit the number of sequences. Finally, 16 'dummies', which were movies that did not contain a sign but only a sequence of two fidgets, were recorded to make sign presentation less predictable. An example of a dummy is shown in Movie 6.

Movie 6 (dummy). 2H-Table → [Nose Rub] → [Hair Brush]. 0.5 Mb

All these movies were used once as stimulus for each participant. During the recording of the movies the signer sat behind a table and she was instructed not to use mouthing, to keep looking into the camera and to keep a straight face. For further details on the recording sessions, see Arendsen et al. (2007).

Procedure - The experiment began with the participant sitting at a table in front of a laptop. On the screen, instructions were given at each stage of the experiment in SLN (video) and in (written) text. First, participants provided their characteristics (e.g. age, gender, hearing status, etc.). Then, participants were asked to provide the lexical meaning of 50 signs, among which were the 32 test signs, because if a sign was unknown to a participant there could be no reliable lexical recognition RT measurement. Participants saw a video of a sign and then had to choose its meaning from a list with ten items (containing the target meaning randomly positioned among nine non-target meanings randomly picked from the set of 50 sign meanings). They could also enter a different meaning manually, or select 'I do not know this sign'.

In the next stage, the reaction time of the participants to respond to a visual event was measured. Participants were instructed to watch a 3D rendering of an arrow that revolved around the vertical axis and to press the spacebar as soon as they saw a change in the direction of the movement (at an arbitrary moment the arrow would suddenly start to revolve around a horizontal axis, causing a visible motion boundary (Rubin & Richards, 1985)). This 'Motion boundary reaction time' was measured ten times for each participant, but only the last five measurements were used to calculate each participant's average.

In the final stage, participants watched the 112 movies described in the material. For each movie they had to press the spacebar to start the movie. Then they had to press the spacebar again as soon as they recognized a sign. At that moment the movie stopped and disappeared. The raw response time was recorded as measured from the start of the movie. After pressing they had to pick the target meaning from a list with ten items (the target meaning at a random position among nine random selections from the other 49 signs of the second stage) or they could choose 'again'. Participants were instructed not to press the spacebar if they thought a movie did not contain a sign, or in case they did see a sign but did not recognize its meaning. If they did not press the spacebar the movie ran to its end and they could choose 'I did not see a sign', 'I do not know this sign', or 'again'. Participants were instructed to use the option 'again' only if they pressed accidentally, if they made a mistake, or if they responded much more slowly than they normally would. In such cases the movie would be repeated later at a random position amongst the remaining stimuli.

Participants were informed that there were no compound signs. To ensure that participants did not misunderstand the task as having to respond as soon as they saw any hand movement they were also told the movies could contain 'other hand movements' which they should try to ignore. Participants first got nine practice movies (extra material, none of the 112 movies that followed) to get used to the routine. If the procedure was unclear to them their questions would be answered.

Coding and analysis

Movement phases and beginning of signs - For our analysis of the results we calculated response times from the beginning of the sign in the movie (signs were preceded by fidgets in one condition). Three coders used custom software developed at our department to code where each (movement phase of a) sign began and ended. This was done according to a coding scheme first proposed by Kendon (1980) and elaborated by McNeill (1992), Kita et al. (1998) and Kendon (2004), but with small modifications. The movement phases we coded were: liberation, preparation, stroke, hold, recovery and settling. Coders were instructed that the stroke was 'the expressive phase of the movement' and that the preparation was the 'strict preparation for the stroke (or independent hold) with movement towards initial location and formation of initial handshape and orientation. A sign was defined as beginning with the start of the preparation and ending with the end of the recovery. The stroke and (post-stroke) hold were taken together as the 'nucleus' (Kendon, 2004).

Using this coding scheme, the average duration of signs turned out to be 1580 ms (range 1059 to 2588 ms). Preparation accounted for about twenty percent of the duration of a sign (average 320 ms, range 164 to 534 ms), the nucleus for about fifty percent (average 775 ms, range 300 to 1640 ms), and recovery for thirty percent (average 485 ms, range 242 to 770 ms). Signs in isolation (N =64) lasted 1680 milliseconds on average, and signs preceded by a fidget (N = 32) lasted 1385 milliseconds on average (with equal relative durations of the separate movement phases). For more details and for the reliability of this coding process see Arendsen et al. (2007).

Defining response times - In the following several response times will be used:

- *Mb-reactiontime* is reaction time for seeing a motion boundary as measured in the moving arrow experiment.
- *Raw RT* is the response time measured from the onset of the sign in the movie. This is the moment in the movie when the spacebar was pressed minus the coded beginning of the preparation of the sign in that movie.
- *RT* is Raw RT minus Mb-reactiontime.

The mean Mb-reactiontime, averaged across all participants, on this task was 308 ms (Standard Deviation (SD) = 28 ms). There was no significant difference in Mb-reactiontime between deaf signers (Mean (M) = 318 ms, SD = 36 ms) and hearing signers (M = 304 ms, SD = 24 ms) (t = 1.31, two-tailed p = 0.200, *not significant (ns)*).

Reducing the data to correct hits on known signs - For the results analyses the following measurements were excluded:

- All responses to unknown signs (about 6% of the data).
- 'Accidental starts' (presses before any movement in the movie: nine cases)

- All responses to fidgets (participants pressed the spacebar to a fidget with 2% of the fidget presentations and then selected 'again' in all of these cases).
- 'Misses' (cases when a known sign was presented but a participant did not press the spacebar: 35 cases. Two signs created many of these 'misses': MIS (MISSED) accounted for 8 of the 35 misses and AFDROGEN lichaam (TOWEL-OFF body) accounted for 7 misses. Other signs only produced 3 or less misses. In case of a miss a movie ran to its end, and participants got three options: 'I did not see a sign', 'I do not know this sign', and 'again'. In the 35 cases these options were chosen 11, 4, and 20 times respectively.)
- Incorrect hits (cases when a participant pressed the spacebar to a known sign but selected 'again' instead of the target meaning (19 cases) and cases where a non-target meaning was selected (6 cases)).

This implies that only correct hits on known signs (2852 cases) were included.

Results

Response times for hits - For the 2852 correct hits (on known signs) the overall median *Raw* RT was 848 ms, with quartiles at 725 ms and 1010 ms. The overall median RT (Mb-reactiontime subtracted) was 538 ms with quartiles at 417 and 700 ms. In our analyses, the median will be used instead of the mean to counter the effect of outlying late responses, which is typical for RT distributions. The median RT, across participants, best represents the distribution of lexical recognition RTs for a particular movie in a single number. Similarly, the median RT, across movies, best represent the distribution of lexical recognition RTs for a participant in a single number.

If we then wish to compare two groups of movies, we take only each movie's median RT and assume that those median RTs have a normal distribution. We also assume that the differences (z-scores) between the median RTs are independent even though their absolute values are dependent because the same group of participants saw all movies. These conditions allow the use of an ANOVA to test the significance of differences in group averages.

Comparison between participants - For each participant the median RT and other descriptive variables were determined for all correct hits on signs, see Appendix 4-A. The average across participants' median RTs (across signs) was about 550 ms from the sign beginning with average quartiles of about 450 and 680 ms. Typically, the distribution of RTs has a strong central cluster (avg. kurtosis = 3.4) and is skewed by a few late responses (avg. skewness = 1.4). Deaf and hearing signers showed nearly identical RT distributions with equal averages of median RTs ($t = 0.10$, two-tailed $p = 0.93$, *ns*). Similarly, no effects were found for parental hearing status ($t = 0.57$, two-tailed $p = 0.58$, *ns*), age of SLN acquisition ($r = -0.01$, $p = 0.96$, *ns*), SLN fluency ($F_{(3, 28)} = 1.19$, $p = 0.33$, *ns*), and SLN usage ($F_{(1, 30)} = 1.74$, $p = 0.20$, *ns*), nor was there an effect of age or gender.

To check whether longer median RTs on the recognition task were perhaps caused by more general problems, for example a participant's inexperience with using a computer, we compared participants' median RTs with their Mb-reactiontimes, these were measured with similar procedures, but found no correlation ($r = 0.03$, $p = 0.88$, *ns*).

Onset of stroke and nucleus duration - For each of the 96 movies containing a sign the median RT, across all correct hits of all participants, was determined, see Appendix 4-B. Figure 10 shows that the timing of the onset of the stroke is strongly correlated with these median RTs ($r = 0.61$, $p < 0.001$).

Next, the median RT for each movie, with respect to the beginning of the stroke, was calculated by subtracting the duration of the corresponding preparation (average 320 ms, range 164 to 534 ms) from the median RTs. These 'median RTs calculated from the onset of the stroke (range 6 to 585 ms, quartiles of 161 and 277 ms) were not correlated to preparation duration ($r = -0.14$, $p < 0.16$, *ns*). In contrast, median RTs calculated from the onset of the stroke were positively correlated to nucleus duration (*r* = 0.53, $p < 0.01$), as shown in Figure 11, and these median RTs were always shorter than the corresponding nucleus duration. There was no correlation between the durations of the preparation and the nucleus of signs (*r* = 0.03, p = 0.76, *ns*).

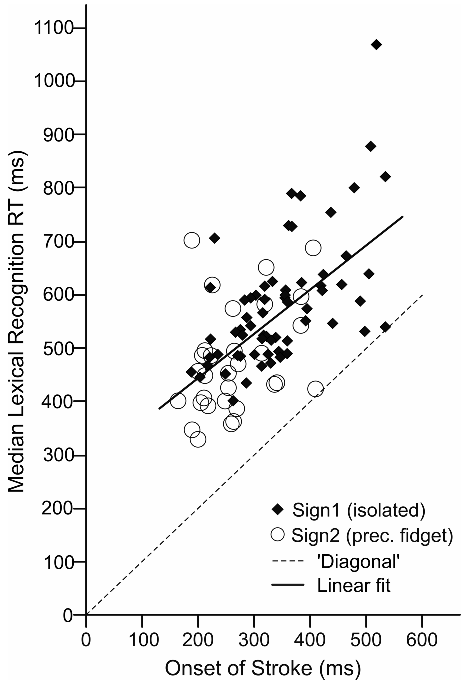


Figure 10. A scatterplot of median RT (across participants) for lexical recognition as a function of the times of onset of stroke for each production of a sign (sign1 = directly from a rest position; sign2 = preceded by a fidget). The solid line represents a least squares linear fit resulting from a regression analysis (the estimated linear relation has a slope of 0.84 (with a standard error of 0.11; a slope of one is inside the 95% confidence interval) and an intercept of 279 ms (std. error = 38 ms)). The dotted 'diagonal' line crosses zero on the axes and has a slope of exactly 1.

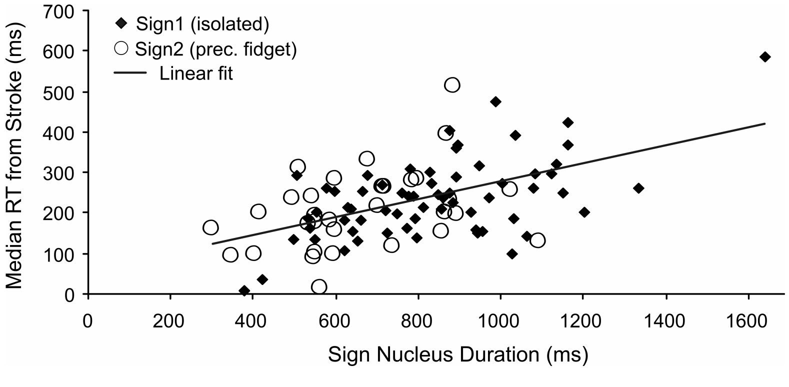


Figure 11. Median response time from the onset of the stroke as a function of the sign nucleus duration (sign1 = directly from a rest position; sign2 = preceded by a fidget). The full line represents a linear fit resulting from a regression analysis (with a slope of 0.22 (with a standard error of 0.04) and an intercept of 54 ms (std. error = 30 ms)).

A regression analysis with median RT (from the beginning of the sign) as the predicted variable and the onset of the stroke and the nucleus duration as predictors yields the following regression equation:

Median RT = (0.82 * Stroke onset) + (0.23 * Nucleus duration) + 109 ms. {a}

Combined in this way, the onset of the stroke and the nucleus duration explained about 54% (adjusted R^2) of the variance in median RTs for lexical recognition of signs. The onset of the stroke had the largest impact on RT (stroke onset: R^2-change = 36.8%, $F = 54.8$, $p < 0.01$; nucleus duration: R^2-change = 18.4%, $F = 38.3$, $p < 0.01$). In the further analysis of the results, where the effects of sign characteristics and embedding are studied, the time of the onset of the stroke in the signs will be taken into account.

Comparison between signs on characteristics - The following possible effects of sign characteristics on lexical recognition RTs were checked in planned comparisons: handshape-markedness, handedness (one or two-handed), location (face, body, neutral space, or arm), handshape change presence, path movement presence, orientation change presence, and repetition presence. The differences in RTs were tested with one-way ANCOVAs, including as covariate the onset of stroke (in other words, we tested whether these sign characteristics had an effect after removing the variance for which the stroke as a quantitative predictor (covariate) accounted). Only the data from the 32 movies with isolated signs and neutral rest positions was used in these tests.

Handshape-markedness had a significant effect on RT ($F_{(2, 28)} = 4.15$, $p < 0.05$). Signs with a highly marked handshape were recognized about 160 ms faster ($t = 2.83$, $p = 0.01$) than signs with an unmarked handshape and about 140 ms faster than signs with in-between handshapes ($t = 2.42$, $p = 0.02$). The difference between unmarked and in-between handshapes was not significant. Recognizing signs that included a change in orientation took about 100 ms longer than recognizing signs without such movement ($F_{(1, 29)} = 5.96$, $p < 0.05$). We did not find an effect of handedness ($F_{(1, 29)} = 3.78$, $p = 0.06$, *ns*), location ($F_{(3, 27)} = 1.39$, $p = 0.27$, *ns*), presence versus absence of a handshape change ($F_{(1, 29)} = 0.56$, $p = 0.46$, *ns*), presence versus absence of path movement ($F_{(1, 29)} = 1.32$, $p = 0.26$, *ns*), presence versus absence of movement repetition ($F_{(1, 29)} = 0.13$, $p = 0.91$, *ns*). Note that these statistical tests were all based on estimated marginal means, as calculated in the ANCOVA procedure, and not on the actual means of the median RTs.

Comparison between signs on embedding - To check the effects of different rest positions and preceding fidgets on lexical recognition RTs the results for the different movies of one-handed signs were compared. The effects of these embedding conditions on the median RTs for the movies were tested with one-way ANCOVAs with stroke onset as covariate.

We found that if a sign was preceded by a fidget it was recognized about 70 ms faster than when it was made directly from the neutral rest position ($F_{(1, 45)} = 4.65$, $p < 0.05$), although if non-neutral rest positions were included in this comparison than the difference dropped to about 50 ms, which was no longer significant ($F_{(1, 78)} = 3.05$, $p = 0.09$, *ns*). All other variations had no impact. Whether the fidget was made in the same location as the sign or somewhere else made no difference ($F_{(1, 45)} = 0.49$, $p = 0.49$, *ns*). Whether the hands were in a neutral rest position (two hands on the table) or in a non-neutral rest position with the signing hand held resting in signing space, on the body or on the face (chin) did not influence response times ($F_{(1, 45)} = 0.02$, $p = 0.89$, *ns*). And, finally, whether the non-neutral rest position had the same location as the sign or had another location did not have an effect ($F_{(1, 29)} = 0.72$, $p = 0.40$, *ns*).

Discussion and comparison with previous lexical recognition work

From the current experiment several findings emerge. Firstly, our results are comparable to the findings of the gating studies by Grosjean (1981) and Emmorey and Corina (1990) in the sense that a fairly small part of a sign typically suffices for lexical recognition. Next, we find that the movement phases of a sign, especially the onset of the stroke, may be used to predict when lexical recognition is possible. In addition, our results indicate effects of several sign characteristics and of conditions prior to the sign. Finally, these results bear some implications for automatic sign recognition. These items are discussed below.

Which earliest part of a sign suffices for lexical recognition? - Before we make comparisons it should be noted that in the current experiment participants were given some prior exposure of each sign to check their knowledge of the signs. Although the target signs were embedded in a larger group during this vocabulary check, it is still the case that each sign was seen by participants about ten minutes (on average) before their lexical recognition response time for that sign was measured. Furthermore, each one-handed sign was offered in five conditions. This prior exposure and repetition together might have had a shortening effect on RTs in comparison to other experiments.

As mentioned before, there are several important differences between our method and the gating paradigm. With the gating paradigm, people have time to process the information and to (motorically) respond *after* each movie presentation and this time is *not included* in the isolation or recognition point. The currently gathered Raw RTs *do include* the time needed for information processing and responding. By subtracting the Mb-reactiontime from the Raw RTs we aim to compensate the time required for a motor response. Still, people will require time for information processing. Another difference lies in the confidence people have in their recognition. In the gating studies, participants' confidence about their guess was still low at the isolation point (on average three on a scale from one to six (Grosjean, 1981)) and, by definition, high (five on the scale) at the recognition point. In the current experiment confidence was not measured. However, very few recognition errors were made, which suggests that most participants responded in a conservative way and wanted to be confident about recognition. Therefore, it is most appropriate to compare our findings with the recognition points in the gating studies, but unfortunately those were only published by Emmorey and Corina (1990) and not by Grosjean (1981).

From the results of the present study we estimate that the first 540 ms of a sign suffice, on average, for lexical recognition. Grosjean (1981) found an average isolation point of about 400 ms. Emmorey and Corina (1990) reported an average isolation time of about 240 ms and an average recognition time of 310 ms which is much shorter than the 400 ms found by Grosjean (1981). They did not explain this difference. A reason may be that the begin points are defined differently. Grosjean (1981) defined the beginning of the sign as "the moment the hand(s) appeared on the screen" and the end as "the moment the sign was no longer recognizable; i.e. when the hand(s) began to move back down to resting position". We are left guessing where the hands were before they appeared on the screen but assume some rest position below the line of recording. Emmorey and Corina (1990) report that the signer rested her hands on her lap with the video camera aimed at a level about 3 inches above her lap and then "a sign was considered to begin when the hand(s) entered signing space and to end when the hand(s) began to move out of the sign configuration and back down to resting position". From these definitions it is unclear if different begin points are used, and if so, how far apart they are. Our begin point was defined as the moment the hands leave the rest position (when there is no preceding fidget). This seems different from the begin point as defined in Emmorey and Corina (1990). We can estimate the differences, and align the current results and those from the gating studies with each other, by assuming that the point at which hand location, hand orientation, and handshape are isolated coincides, on average, with our onset of the stroke. In Grosjean (1981) that moment occurs on average at 320 ms, the same as our average onset of the stroke, which is coded as "the expressive phase of the movement" (Arendsen et al., 2007). Thus, the current results and those of Grosjean (1981) appear to use the same begin points and to be aligned already. However, Emmorey and Corina (1990) report that hand location, hand orientation, and handshape are already isolated at 170 ms from which we estimate a 150 ms (320 ms – 170 ms) difference in begin points between Emmorey and Corina (1990) and both Grosjean (1981) and the current results. These differences in begin points can be made irrelevant if we adjust all the findings by using the onset of the stroke as the 'zero' in the time dimension. In the current results, lexical recognition occurred about 220 ms after the onset of the stroke. In the results of Grosjean (1981) signs were isolated 80 ms after the onset of the stroke (our estimation). Emmorey and Corina (1990) reported that signs were isolated 70 ms after the onset of the stroke and recognized, with good confidence, 140 ms after the onset of the stroke (also our estimations). Thus, the current response times are about 80 ms longer than their average recognition point which may reflect time required to process the information. The current response times are thus about 140-150 ms longer than the previously reported isolation times, a difference that includes time for both processing information and gaining confidence.

Regarding sign duration, it should be noted that the signing in our material was probably performed more slowly than in the studies by Grosjean (1981) and Emmorey and Corina (1990). If signs last longer because they are performed more slowly it is possible that more time is required for recognition because information may become available at a slower pace (note that some factors, such as the length of post-stroke holds, may effect sign duration but not the speed at which the crucial information at the onset of the stroke becomes available). Average sign duration in our material was 1100 ms including preparation and nucleus (and excluding the recovery phase in compliance with duration calculation in the other studies). The average duration of the signs was 817 ms for Grosjean (1981) and 703 ms for Emmorey and Corina (1990), a difference of 114 ms which can be explained by the estimated 150 ms difference in begin points mentioned above, making it likely that signs were performed about equally fast in both studies. Our sign durations are considerably longer, which, as argued, may have had a lengthening effect on response times.

Emmorey and Falgier (2004) reported recognition times (M = 926 ms, SD = 237 ms) that are similar to our Raw RTs (M = 894 ms, SD = 251). One might have expected our RTs to be shorter because of contextual factors in our study (e.g. prior exposure and repetition of signs). However, their sign duration (M = 833 ms, SD = 269 ms) was again much shorter than our sign duration (M = 1100 ms, SD = 254). Therefore, it is possible that the signing in our material was performed more slowly, perhaps causing longer response times than would otherwise have been obtained. In the experiment of Emmorey and Falgier (2004) response times were measured that occurred 100 ms after the end of the nucleus. If we assume that the participating signers also needed about 300 ms to provide a motor response then these data suggest that, without any contextual aid, lexical recognition typically requires a somewhat larger part of the nucleus of the sign than we find in our setting.

Recognition times and coded movement phases - For the current material, the stroke and other movement phases of the signs in these movies were coded manually by three people. Between the coders there were often small differences and sometimes large differences (up to about 200 ms), see Arendsen et al. (2007). Despite this variance in the coding, the coded onset of the stroke (and to a lesser degree the duration of the nucleus) was found to be a good predictor of the median lexical recognition RT for a particular movie of a sign. This suggests an efficient approach for estimating the recognition times of a large set of signs, for example if you need to select a set of signs with equal recognition times as did Emmorey and Falgier (2004). In this approach you only need two or three people coding the movement phases. From their coding you can then estimate the time required for recognition of each sign using the regression equation provided in the results section (indicated with {a}). This may replace the time consuming gathering of recognition response times from a large group of signers. One step further would be to code the movement phases automatically instead of manually. However, this appears to be a difficult problem although there has been some progress on automatically recovering the temporal structure of gestures (Wilson et al., 1996) and signs (Parish et al., 1990).

Effects of sign characteristics - In our results, signs with highly marked handshapes were recognized faster than other signs matching the results from Emmorey and Corina (1990). Moreover, they found an interaction between handshape markedness and age of acquisition (native signers recognized signs with marked and unmarked handshapes equally fast while non-native signers recognized signs with marked handshapes faster). We did not find this interaction. Emmorey and Corina (1990) suggested that signs with highly marked handshapes can be recognized faster because highly marked handshapes rule out a larger proportion of non-target signs, i.e. there are less signs with such handshapes (or rather, the 'cohort size' is drastically reduced by the appearance of a highly marked handshape, see Klima and Bellugi, 1979). This seems a plausible explanation for our data as well.

Signs with a change in hand orientation had higher median RTs than signs without an orientation change. A possible explanation for this difference would be that an orientation change typically occurs in the unfolding sign after other informative elements of the sign (location, handshape, etc.) have already become visible and that people therefore have to wait for the appearance of the orientation change in the unfolding sign before they can recognize the sign. However, this does not explain why there is not a similar difference between signs with a handshape change or not, or signs that have a path movement or not, assuming that these kinds of movements occur at roughly the same time as a hand orientation change. In our results, signs with or without a handshape change had equal median RTs and this is in line with Emmorey and Corina (1990). Their first expectation was that people would require more time for signs with a handshape change because they would have to wait until the handshape change occurred to be able to identify the sign (following the same line of reasoning as used above for the effect of an orientation change). To explain the lack of an expected difference they speculated that in their gating study participants could predict a handshape change based on information present early in the signal. It may be the case that participants in the current test procedure could indeed pick up the relevant clues to predict the handshape change (and also the path movement), but not the orientation change. Therefore, our results suggest that this reasoning does not hold for hand orientation change.

Concerning the location of the sign our results are not in line with those of Emmorey and Corina (1990), who found that signs made near the face had longer isolation times than signs made in neutral space. We did not find an effect of location on median RTs but there was a trend that the less frequent locations (e.g. arms and face), which thus leave a more limited number of candidate signs for the identification process, had shorter median RT than the more frequent locations (e.g. neutral space). Therefore, our results are somewhat in line with Grosjean (1981), who found a positive correlation between frequency of location and isolation times.

Implications for automatic sign recognition - The current results show that automatic recognition should be possible shortly after the onset of the stroke of a sign, that is on average about 540 ms after the beginning of the sign. However, automatic sign recognition algorithms usually wait until a sign has ended, for example when a person's hands are back in some rest position (Lichtenauer et al., 2007; Lee et al., 2005; Zieren & Krais, 2005), before they start the recognition process that leads, after some processing time, to a response. Thus, assuming our expectations are shaped by human performance, anyone using such an application will experience a delay in the response to his or her signing. With the results of this study it is possible to estimate roughly the length of this experienced delay. If we allow a computer application to be as slow as a reasonable fraction of the human responses (for example the 75% percentile of human responses, that is three quarter of human responses is faster and one quarter of human responses is slower), then its response to a sign (in a task of lexical recognition) should be available after about 1000 ms, which is still during the sign. We estimate users will not experience a delay if the computer responds by the time the recovery phase of a sign starts, which was after 1100 on average in our material. The recovery was about 500 ms on average in on our material. If an application waits for a sign to end before attempting recognition this amounts to an estimated experienced delay of half a second, to which the necessary processing time for recognition and response is to be added.

A preferable approach for fast automatic sign recognition would be to start the recognition process during the sign. For a limited set of gestures, that approach was indeed attempted by Kim et al. (2007) who proposed a 'forward spotting scheme for simultaneous gesture segmentation and gesture recognition' instead of the usual 'backward spotting scheme' that first detects the end point, then traces back to the start point and sends the extracted gesture segment to the recognizer. Their approach appears to resemble that of humans who are constantly trying to guess what the meaning of a gesture or sign might be, given what they have seen of it as it unfolds. Kim et al. (2007) report that they were able to create a system with continuous online gesture recognition with a decent recognition rate and 'without any time delay'.

4.4. Study 2. Lexical recognition versus sign detection

In this study the current results on lexical sign recognition are compared with the results from a previous experiment on detecting the beginning of a sign (Arendsen et al., 2007). A specific concern is whether lexical recognition plays a role in sign detection by signers. The same stimulus material was used in a similar procedure with similar participants. The only difference was the task given to participants.

Method

In the current experiment participants were instructed "to press the spacebar as soon as you recognize a SLN sign" (lexical recognition). In the previous experiment they were instructed "to press the spacebar as soon as you see the beginning of a SLN sign" (sign detection).

The *Raw RT* measurements were used to investigate the effect of the different tasks of 'sign detection' and 'lexical recognition', thereby assuming that the participants from both experiments had the same average reaction times to events in the visual signal, that is to say, their visual information access times were assumed to be similar.

Results

The first comparison between the tasks was based on the median Raw RTs across all measurements from all signers for each experiment. From the sign detection experiment, in which participants performed two series of measurements, only data from the first series was included, because for lexical recognition only one series of measurements was taken. The measurements for lexical recognition had a median Raw RT of 848 ms (quartiles 725 and 1010 ms) and those for sign detection, by signers, had a median Raw RT of 754 ms (quartiles 639 and 902 ms). Thus, lexical recognition took about 90 ms longer and the distributions were spread about equally. For further comparison, sign detection by non-signers had a median Raw RT of 883 ms (quartiles 724 and 1106 ms)

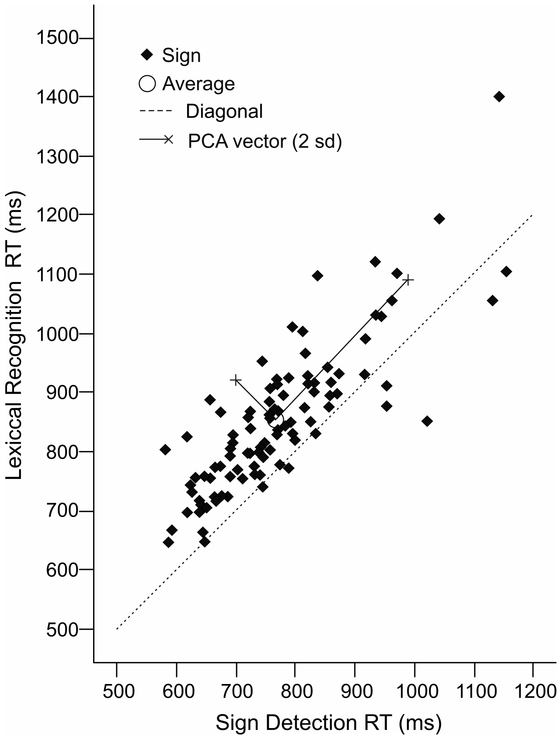


Figure 12. Median Raw RTs (across participants) for each movie for sign detection and lexical recognition. The averages are plotted together with the eigenvectors ('PCA vectors' in the figure) of the covariance matrix (with lengths of 2 standard deviations in the direction of each factor).

Next, the consistency of the difference between lexical recognition and sign detection was studied. Figure 12 shows, for each movie of a sign, the median Raw RTs (across signing participants) for lexical recognition, on the vertical axis, against the median Raw RTs for sign detection on the horizontal axis. We found a strong correlation ($r = 0.83$, $p < 0.001$) between signers' lexical recognition and sign detection Raw RTs. To compare, signers' lexical recognition and non-signers sign detection Raw RTs were also correlated ($r = 0.65$, $p < 0.001$), though somewhat less strongly, and the same was true for the correlation between signers' and non-signers' sign detection Raw RTs ($r = 0.66$, $p < 0.001$). To study the relation between signers' lexical recognition and sign detection RTs further, a principal components analysis (PCA) was performed. The PCA showed that the eigenvector of the covariance matrix with the highest Eigen value has a slope of 1.04, suggesting a linear relationship, in such a way that if the beginning of a sign is detected a certain amount of time faster than the beginning of another one, it is expected to be recognized faster by an equal amount of time. However, there were also some discrepancies. As can be seen in Figure 12 there are some points below the dotted diagonal line, each one representing a movie with a lower median Raw RT for lexical recognition than for sign detection.

The relation between lexical recognition and sign detection is examined further by comparing effect patterns. Table 22 summarizes the effects of sign characteristics and embedding on the response times in both experiments. The pattern of the differences in response times and their significance is remarkably similar. The only difference in results occurred for location, with a significant effect on sign detection, but not on lexical recognition. However, the trend for location was roughly the same with signs made on the face causing the shortest response times. Different embedding conditions and sign characteristics appear to have had very similar effects on sign detection and lexical recognition RTs.

Table 22. RT differences (ms) , for sign detection and lexical recognition, between signs with various characteristics and in various embedding conditions (based on estimated marginal means of median RTs). The significance of the differences is tested using one-way ANCOVAs of median Raw RTs with onset of stroke as covariate (or in case of more than two values with pairwise comparisons between RTs for a given value and the value set to zero). For each factor (e.g. 'preceding fidget') the RT for the *value* with the lowest RT (for preceding fidget: 'yes') is set to zero. The difference between the RT of the *value* with the lowest RT and the RTs of other *values* is then given for each value. To make a fair comparison the data for sign detection is only from the signers' first series of response times, not from subsequent series or from non-signers.
$^{*}p < 0.05$; $^{**}p < 0.01$.

Factor	Value	RT differences lex. recognition (ms)	p-value	RT differences sign detection (ms)	p-value
Handshape-markedness	Highly marked	0		0	
	In-between	147	0.016*	116	0.014*
$F_{(2, 28)}$	Unmarked	173	0.005**	142	0.003**
1 or 2 hands	1 hand	0		0	
$F_{(1, 29)}$	2 hands	88	0.042*	101	0.002**
	Face	0		0	
Location	Body	79	0.324	61	0.288
$F_{(3, 27)}$	Space	106	0.057	102	0.012*
	Arm	44	0.528	128	0.014*
Path movement	Yes	59	0.278	35	0.412
$F_{(1, 29)}$	No	0		0	
Orientation change	Yes	96	0.030*	80	0.020*
$F_{(1, 29)}$	No	0		0	
Handshape change	Yes	36	0.529	11	0.801
$F_{(1, 29)}$	No	0		0	
Repetition	Yes	13	0.781	0	
$F_{(1, 29)}$	No	0		25	0.493
Preceding fidget $F_{(1, 45)}$	Yes	0		0	
	No	67	0.036*	81	0.003**
Rest position	Neutral	0		0	
$F_{(1, 45)}$	Non-neutral	1	0.982	16	0.531

Discussion

Arendsen et al. (2007) found that the non-signers took about 120 ms more time on average than the signers to detect the beginning of a sign. They speculated that for sign detection the signers as well as the non-signers relied on the ability to see that a movement is intended to communicate, but that the signers were faster because they were more sensitive to prototypical forms of signs. In this speculation, lexical recognition did not play a relevant role in sign detection.

However, the finding of Arendsen et al. (2007) that the signers needed about 120 ms less time than non-signers to detect the onset of signs, might have more causes: Signers may have achieved shorter response times for sign detection than non-signers because they already recognized, to some extent, the lexical meaning of the sign. Grosjean (1981) and Emmorey and Corina (1990) demonstrated that lexical recognition is possible after only a small part of the sign has been presented, a part that is comparable with the part required by signers for sign detection (Arendsen et al., 2007). Moreover, it is very well possible, as proposed by Grosjean (1981), that signers use the information in the early part of the signal to make guesses about the lexical meaning with, initially, low confidence. Perhaps even low confidence recognition may suffice for detecting signs in the sense of discriminating them from fidgeting. Our current data do not contradict this possibility.

Our results show a strong, linear relation between the RTs of signers for lexical recognition and sign detection (Figure 12) and similar patterns of effects of sign characteristics and embedding conditions for both tasks (Table 22), in such a way that it suggests a consistent time difference of about 90 ms between sign detection and lexical recognition RTs. In addition, signers' lexical recognition RTs were found to be about 30 ms shorter than non-signers' sign detection RTs. In other words, signers can recognize the meaning of signs faster than non-signers can detect signs. Therefore, the fact that signers knew the meaning of most of the signs can explain (part of) the difference between signers and non-signers in sign detection RTs. It appears likely that during sign detection the observed movement is also analyzed, by signers, in terms of possible lexical meanings. Signers may project lexical items from their memory on the observed form of the movement. In case such a projection is successful, they may perhaps not have sufficient information to isolate the meaning of the sign (other projections may also still be possible), but they might have sufficient confidence to respond that the movement is a sign and not a fidget.

Conclusions

The current results show that, in a real-time recognition task with some context available, lexical recognition, with a certain degree of confidence, is typically possible shortly after the onset of the stroke of a sign. As the sign unfolds and shifts from the preparation to the nucleus signers start to recognize the lexical meaning. Overall, those findings are in line with the gating studies by Grosjean (1981) and Emmorey and Corina (1990) and with the data presented by Emmorey and Falgier (2004).

The lexical recognition response times were compared with the sign detection response times from Arendsen et al. (2007) and were found to be closely related. Lexical recognition took, on average, about 90 ms longer than sign detection and this difference was consistent across different (movies of) signs. In addition, sign characteristics and embedding conditions had similar effects on both response times. It appears likely that lexical recognition plays a role in sign detection by signers.

Notes

1. The online version is available at: http://www.ingentaconnect.com/content/jbp/gest/

Appendix 4-A. Individual Response Time Results

Table 23 lists the characteristics of each participant some personal data, the median Raw RT, the median RT with the 25% and 75% quartiles, and the number of measurements (N). Only the RTs for correct hits on a known sign are included.

Table 23. The characteristics of the participants and their results (all RTs in ms).

Hearing status	Parental hearing status	Age of SLN Acquisition	SLN Fluency	SLN Usage	Age	Gender	Med. Raw RT	RT Med. 50%	RT 25%	RT 75%	N
Deaf	Deaf	0-2 years	Fluent	Primarily	26	Male	731	426	356	497	86
Deaf	Deaf	0-2 years	Fluent	Regularly	39	Female	856	550	455	653	89
Deaf	Hearing	0-2 years	Fluent	Daily	46	Female	845	575	470	714	96
Deaf	Hearing	3-6 years	Fluent	Primarily	36	Female	911	550	410	684	96
Deaf	Hearing	3-6 years	Fluent	Daily	35	Female	863	544	403	673	96
Deaf	Hearing	3-6 years	Fluent	Daily	52	Female	1.100	788	619	934	96
Deaf	Hearing	7-16 years	Good	Daily	33	Female	1.071	679	542	917	89
Deaf	Hearing	7-16 years	Good	Daily	49	Male	942	659	575	820	96
Deaf	Hearing	17-29 years	Fluent	Primarily	35	Female	681	376	324	455	96
Deaf	Hearing	17-29 years	Good	Daily	34	Female	738	412	308	502	94
Hearing	Deaf	0-2 years	Good	Daily	24	Female	848	562	457	680	89
Hearing	Hearing	3-6 years	Good	Daily	26	Female	948	643	512	800	88
Hearing	Hearing	7-16 years	Some	Regularly	19	Female	904	608	458	735	68
Hearing	Hearing	17-29 years	Good	Daily	19	Female	856	536	466	669	96
Hearing	Hearing	17-29 years	Good	Daily	37	Female	821	500	405	620	95
Hearing	Hearing	17-29 years	Good	Regularly	26	Female	689	355	260	487	96
Hearing	Hearing	17-29 years	Good	Regularly	30	Female	606	319	248	456	95
Hearing	Hearing	17-29 years	Good	Regularly	36	Female	761	460	347	565	79
Hearing	Hearing	17-29 years	Average	Daily	27	Female	919	623	519	787	90
Hearing	Hearing	17-29 years	Average	Regularly	20	Female	1.012	764	638	925	74
Hearing	Hearing	17-29 years	Average	Regularly	29	Female	737	465	402	543	84
Hearing	Hearing	17-29 years	Average	Regularly	32	Female	747	434	348	545	95
Hearing	Hearing	17-29 years	Average	Regularly	33	Female	849	526	444	640	90
Hearing	Hearing	17-29 years	Average	Regularly	41	Female	751	446	365	555	96
Hearing	Hearing	>30 years	Good	Daily	44	Female	932	627	527	748	78
Hearing	Hearing	>30 years	Good	Daily	47	Male	897	560	470	655	96
Hearing	Hearing	>30 years	Good	Regularly	37	Female	676	388	325	481	95
Hearing	Hearing	>30 years	Average	Daily	46	Male	1.093	803	633	915	84
Hearing	Hearing	>30 years	Average	Regularly	35	Male	919	624	481	764	90
Hearing	Hearing	>30 years	Average	Regularly	42	Female	997	658	511	829	85
Hearing	Hearing	>30 years	Average	Regularly	52	Female	770	486	390	624	78
Hearing	Hearing	>30 years	Some	Exception	53	Male	1.082	738	590	943	77
Average					36		861	553	446	682	89

Appendix 4-B. Response Time Results per Movie

Each movie is identified in Table 24 by gloss, preceding fidget and rest position. Descriptive statistics of RTs are given for each movie: median and quartiles, and number of measurements (N). Only RTs for correct hits on a known sign are included. In addition the onset of the stroke is given for each movie.

Table 24. Response time results for each movie containing a sign (all times in ms).

Gloss	Preceding fidget	Rest position	Stroke onset	Nucleus duration	Med. Raw RT	RT Med. 50%	RT 25%	RT 75%	N
ZAND	-none-	2H-Table	231	988	1.005	705	579	809	30
ZAND	-none-	1H-Space	479	1.134	1.106	801	573	1.003	30
ZAND	-none-	1H-Face	367	894	1.058	726	607	897	30
ZAND	Lip Touch	2H-Table	225	867	913	619	511	781	30
ZAND	Table Drum	2H-Table	188	883	1.031	702	559	829	30
SCHEP	-none-	2H-Table	283	780	877	590	448	817	32
SCHEP	-none-	1H-Space	356	664	933	609	514	759	31
SCHEP	-none-	1H-Face	355	790	887	595	456	799	32
SCHEP	Nose Rub	2H-Table	222	718	777	488	393	593	32
SCHEP	Hand Squeeze	2H-Table	312	587	795	492	407	763	32
EUROPA	-none-	2H-Table	319	1.124	909	615	500	779	28
EUROPA	-none-	1H-Space	363	1.163	1.032	729	583	834	28
EUROPA	-none-	1H-Face	249	927	775	452	407	638	29
EUROPA	Chin Rub	2H-Table	208	783	805	487	376	635	28
EUROPA	Table Drum	2H-Table	404	795	1.012	688	554	1.013	29
TEKENEN	-none-	2H-Table	361	884	896	586	479	750	30
TEKENEN	-none-	1H-Space	294	827	902	594	448	741	30
TEKENEN	-none-	1H-Body	419	748	919	617	462	790	30
TEKENEN	Hair Brush	2H-Table	268	735	699	388	319	589	30
TEKENEN	Arm Fold	2H-Table	261	508	871	575	454	667	30
WC	-none-	2H-Table	359	654	799	490	403	604	32
WC	-none-	1H-Space	264	795	726	402	347	509	32
WC	-none-	1H-Body	315	552	830	518	423	615	32
WC	Hair Brush	2H-Table	254	533	725	428	331	492	32
WC	Hand Squeeze	2H-Table	188	596	649	348	257	482	32
BROER	-none-	2H-Table	236	597	805	488	399	568	32
BROER	-none-	1H-Body	280	847	830	525	411	619	32
BROER	-none-	1H-Face	189	714	758	456	346	665	32
BROER	Chest Scratch	2H-Table	217	549	670	394	305	443	32
BROER	Ear Grab	2H-Table	164	493	712	402	347	507	32

- Table continues on the next two pages -

Table 24. Continued.

Gloss	Preceding fidget	Rest position	Stroke onset	Nucleus duration	Med. Raw RT	RT Med. 50%	RT 25%	RT 75%	N
BAD	-none-	2H-Table	330	1.065	780	473	381	710	30
BAD	-none-	1H-Body	222	1.035	955	614	441	779	30
BAD	-none-	1H-Face	314	955	764	467	367	551	29
BAD	Chest Scratch	2H-Table	199	1.093	649	331	286	522	30
BAD	Chin Rub	2H-Table	264	877	799	496	406	600	30
AFDROGEN	-none-	2H-Table	275	1.079	860	535	394	852	22
AFDROGEN	-none-	1H-Body	221	1.336	817	483	341	679	23
AFDROGEN	-none-	1H-Face	272	813	836	487	415	660	23
AFDROGEN	Chest Scratch	2H-Table	247	857	708	403	339	557	23
AFDROGEN	Hair Brush	2H-Table	200	1.024	756	458	352	608	22
VIES	-none-	2H-Table	497	422	853	532	456	677	29
VIES	-none-	1H-Face	339	622	840	520	457	618	30
VIES	-none-	1H-Space	505	549	945	639	595	713	30
VIES	Chin Rub	2H-Table	336	402	742	434	359	591	29
VIES	Arm Fold	2H-Table	382	299	869	544	439	718	29
KOORTS	-none-	2H-Table	394	663	878	574	497	654	27
KOORTS	-none-	1H-Face	360	643	832	513	408	683	27
KOORTS	-none-	1H-Space	319	833	897	592	480	704	27
KOORTS	Lip Touch	2H-Table	318	712	889	584	420	668	27
KOORTS	Hand Squeeze	2H-Table	382	702	925	598	371	699	26
MAMA	-none-	2H-Table	333	505	929	624	458	772	31
MAMA	-none-	1H-Face	223	678	809	516	383	652	31
MAMA	-none-	1H-Space	267	578	821	529	412	675	31
MAMA	Ear Grab	2H-Table	204	550	721	399	337	491	31
MAMA	Chest Scratch	2H-Table	211	543	763	451	303	567	31
PAPA	-none-	2H-Table	316	762	870	566	479	704	31
PAPA	-none-	1H-Face	204	775	758	445	372	607	32
PAPA	-none-	1H-Space	217	877	775	468	402	616	32
PAPA	Chin Rub	2H-Table	211	598	807	495	378	559	32
PAPA	Table Drum	2H-Table	320	677	968	652	525	770	32
KIJKEN	-none-	2H-Table	317	719	845	524	421	673	32
KIJKEN	-none-	1H-Face	287	726	745	436	360	553	32
KIJKEN	-none-	1H-Space	276	637	774	485	388	648	32
KIJKEN	Lip Touch	2H-Table	259	594	665	359	303	437	32
KIJKEN	Hand Squeeze	2H-Table	349	545	761	440	371	596	32
KIP	-none-	2H-Table	422	791	914	608	473	872	32
KIP	-none-	1H-Face	385	862	927	623	456	778	32
KIP	-none-	1H-Space	326	773	793	488	395	709	32
KIP	Nose Rub	2H-Table	271	866	770	472	348	655	32
KIP	Hand Squeeze	2H-Table	210	892	719	408	267	526	32

Table 24. Continued.

Gloss	Preceding fidget	Rest position	Stroke onset	Nucleus duration	Med. Raw RT	RT Med. 50%	RT 25%	RT 75%	N
MIS	-none-	2H-Table	456	539	916	618	486	756	21
MIS	-none-	1H-Face	346	498	800	482	375	700	22
MIS	-none-	1H-Space	424	630	934	638	551	815	22
MIS	Nose Rub	2H-Table	253	413	761	455	334	538	23
MIS	Chest Scratch	2H-Table	340	348	728	436	385	546	23
TELEFOON	-none-	2H-Table	534	378	849	540	481	753	32
TELEFOON	-none-	1H-Face	301	535	827	488	381	635	31
TELEFOON	-none-	1H-Space	441	621	859	546	454	651	32
TELEFOON	Ear Grab	2H-Table	263	548	699	364	306	472	32
TELEFOON	Hand Squeeze	2H-Table	408	560	733	424	343	535	32
AUTO	-none-	2H-Table	355	972	900	593	474	754	31
MELK	-none-	2H-Table	330	1.030	833	514	449	579	32
FIETS	-none-	2H-Table	321	1.201	852	523	457	629	32
SOEP	-none-	2H-Table	392	938	854	552	472	691	22
JARIG	-none-	2H-Table	344	944	816	494	422	644	32
PAARD	-none-	2H-Table	287	1.002	875	558	506	683	27
BOTERHAM	-none-	2H-Table	294	1.152	865	542	449	668	32
EGEL	-none-	2H-Table	437	946	1.058	754	667	889	32
OPRUIMEN	-none-	2H-Table	509	898	1.196	877	730	1.192	30
RAAM	-none-	2H-Table	483	1.640	1.402	1.068	816	1.513	32
TELEVISIE	-none-	2H-Table	304	1.082	918	599	480	812	28
BOOM	-none-	2H-Table	383	876	1.103	785	667	904	26
FEEST	-none-	2H-Table	489	1.029	879	587	479	758	32
AANKLEDEN	-none-	2H-Table	367	1.163	1.098	789	499	1.055	32
POES	-none-	2H-Table	464	855	992	673	598	774	32
KOE	-none-	2H-Table	534	893	1.122	822	676	962	32

Chapter 5

♠

Acceptability of sign manipulations

This study contains the results of an experiment in which signers were asked to judge the acceptability of a set of sign manipulations. Signs were recorded with variations in different categories, in the temporal and spatial dimension. Participants varied much in tolerance, i.e. in the percentage of movies they judged to be acceptable, but their rankings of the acceptability of sign manipulations correlated well. On the level of dimensionality we found that temporal manipulations were highly acceptable while spatial (and spatiotemporal) manipulations were often judged unacceptable. Further division of the manipulations into categories, such as changes in hand orientation or movement direction, showed much variability and little regularity. The roles of phonology and iconicity in acceptability judgments were studied. Part of the variability in the acceptability of sign manipulations could be explained on the basis of the type of phonological error caused by each manipulation, and by considering a sign's iconicity and classifying whether manipulations are compatible or incompatible with that iconicity. Finally, human judgments were compared to acceptability ratings by three automatic sign recognizers.

To be submitted to Sign Language Studies.

5.1. Introduction

This paper presents an experimental study on the acceptability of sign manipulations. Lexical signs, like any linguistic utterance, can be produced with much variation. But which variations are acceptable? And what determines the acceptability of a given variation? How about the consistency within and between human judgments and the consistency between human judgments and scores obtained by automatic sign recognition? These questions became critical during the development of an Electronic Learning environment (ELo) for deaf and hearing impaired children to practice Sign Language of the Netherlands (SLN) signs (Spaai et al, 2008), a related project. To answer questions like these we set up an experiment in which 26 signers were instructed to judge the acceptability of a set of 131 manipulations of four lexical signs on a 5-point acceptability scale. The manipulations were in the temporal, spatial, and spatiotemporal dimension. In order to interpret the outcome of this experiment, we decided to take into account the phonological rules that might be violated by these manipulations as well as the possible impact of the iconicity of the four signs.

Acceptability is a concept used in linguistics to denote the intuitive judgments by users of a language on how acceptable a linguistic utterance is (Greenbaum, 1977; Sorace, 1996). The linguistic utterance can be a word, a sentence, a fragment of speech in a certain dialect, or any other piece of language. Acceptability is not the same as *grammaticality*. An utterance is grammatical if it obeys the rules of a formal linguistic theory (Greenbaum, 1977), such as grammatical rules for sentence construction or phonological rules for the construction of lexical items. In a sense, it takes a linguist and a chosen set of linguistic rules to determine grammaticality. However, this linguist's viewpoint can differ from the intuitive acceptability judgment of laymen. While determining the acceptability of an utterance in a given language, laymen may apply the rules they know of this language but at the same time they will also be influenced by other, 'extra-grammatical' factors. For example, if an utterance is placed after a set of acceptable utterances it will be judged less acceptable than if it is placed after a set of unacceptable utterances (Van Dijk, 1977). Moreover, acceptability judgments reflect people's beliefs about the forms they habitually use or ought to use, and their personal tolerance for other people using different forms (Greenbaum & Quirk, 1970). Differences in grammatical knowledge and extra-grammatical factors may cause variability in acceptability judgments: A single case of a produced utterance may be judged quite differently under different circumstances or by different judges (Sorace, 1996). Despite this potential for variability it has been found, for spoken language utterances, that the rank order of the acceptability of a set of produced utterances is largely the same across judges and across circumstances (e.g. Mohan, 1973; Cucchiarini et al., 2000; Moustroufas et al., 2007). In our study we investigated to what extent signers agreed on the acceptability of a set of sign language utterances.

Van der Kooij's (2002) SLN phonology was applied in order to interpret the acceptability judgments gathered experimentally. In other words, we tested if the acceptability judgments followed her rules. Van der Kooij (2002) and Crasborn (2001) analyzed the SignPhon (Crasborn et al., 2002) database containing Sign Language of the Netherlands (SLN) utterances to formulate phonological categories and rules of phonetic implementation for SLN. The resulting phonological proposals are mainly concerned with identifying what is 'distinctive' in the phonetic spectrum of a language (Crystal, 1980), but they are also intended to be predictive for the acceptability of variation (Van der Kooij, personal communication).

In sign language linguistics, phonology often has a problematic relationship with iconicity. A (semiotic) sign is considered iconic if it shares properties with its object (Peirce, 1868, 1965; Morris, 1946; Nöth, 1990), or rather, if people perceive such a shared property (Eco, 1976). Iconicity pervades gesture and sign language in many ways (Müller, 1998; Streeck, 2008). In her proposal for SLN phonology, Van der Kooij (2002) incorporated iconicity, but in a limited way (see also Van der Hulst & Van der Kooij, 2006): it is not an entire sign that is treated as iconic, but only some (sublexical) aspect of a sign, for example, the location 'temple' is related to 'thought activities'. In our study, we investigated the role of iconicity in judging the acceptability of variation in lexical signs, both where it concerns aspects of signs, in the sense of Van der Kooij (2002), and where it concerns the whole sign.

Man versus machine

In addition to asking humans to judge the acceptability of sign manipulations it is also possible to use automatic sign language recognition (Von Agris, 2008) to generate 'automatic acceptability ratings'. This can be applied, for example, in electronic learning environments where feedback is provided automatically to language learners. One example of such an electronic learning environment is ELo, to which our research is related (Lichtenauer et al. 2007; Spaai et al., 2008; Lichtenauer et al. 2008) but there are also other examples (Brashear et al., 2006; Zieren, 2007). In our study the human acceptability judgments were compared to automatic acceptability ratings from three types of sign recognizers used for ELo. Some aspects of the technology of automatic sign language recognition are relevant for the current research:

- A widespread method in automatic gesture or sign recognition is the use of 'time warping' or 'dynamic time warping' (Lichtenauer et al, 2007; Von Agris et al., 2008). This means that the signal of the 'test sign' (the input to be recognized or judged) is automatically aligned to the fixed time length of an expected sign model. Essentially, by using time warping temporal variation is disregarded. To our knowledge only anecdotal evidence exists to warrant this and empirical data on the acceptability of temporal variation would be reassuring. Thus, in the current research, the effects of temporal variation on acceptability were studied next to the effects of various forms of spatial and spatiotemporal variation.
- Sign recognizers appear to have very little explicit knowledge of what they are recognizing: they are typically trained with a fairly limited set of examples with limited variability (in comparison to the input received by human language learners) and they are typically not equipped with linguistic knowledge (see Vogler and Metaxas (2004) for an exception). In addition, they do not attempt to perceive iconicity. However, automatic sign recognizers do employ sophisticated computational strategies, for example to select distinctive features for a set of signs (Lichtenauer, 2007, 2008).
- ELo's sign recognizer (currently) observes neither facial expressions nor lip movements (for mouthing). This restriction was copied in the current research where we studied only variation of the manual aspects of signs and not variation in facial expressions or lip movements.

Research questions

We will measure which sign manipulations are judged acceptable or not by signers. Besides addressing the basic concerns regarding the consistency of acceptability judgments between signers, the following research questions will be treated:

1. Is temporal variation acceptable? How do temporal variations compare to various types of spatial and spatiotemporal variation? Are there differences between different types of temporal variation?
2. Are the effects of a certain sign manipulation (e.g. 'slower' or 'faster') or category of sign manipulations (e.g. 'changes in hand orientation') on the acceptability of this manipulation consistent for different signs or not?
3. Can acceptability be predicted from phonological specifications and phonetic implementation rules? Does iconicity play a role in acceptability judgments?

4. Do automatic acceptability ratings match human acceptability judgments?

The first and second group of research questions are addressed by the experiment reported in § 5.2. The third group of research questions are treated in § 5.3 and the final research question in § 5.4.

5.2. Experiment: Gathering acceptability judgments

Method

Participants - Acceptability judgments were gathered from 26 signers, 5 male and 21 female, who reported to be reasonable, good or fluent SLN signers (for details, see Appendix 5-A). Their age varied between 22 and 51 years (M = 35 years, SD = 8 years). Thirteen signers reported to be deaf, all congenitally. Family and SLN courses were mentioned as the primary source of SLN acquisition by seven and sixteen signers, respectively. Six of the eleven fluent signers reported starting their SLN acquisition at zero to two years old. Four of them had deaf parents and could therefore be considered *native signers*. Usage of SLN varied; five (deaf) signers cited SLN as their primary means of communication, twelve used SLN at least daily, and nine signers used SLN regularly. Twenty-three participants used SLN professionally, with seven being SLN teachers and four SLN interpreters.

Material - The material consisted of 131 movies of mild to extreme variations of four SLN signs: 30 movies of GORDIJN (CURTAIN), 35 movies of OVEN (OVEN), 34 movies of STAPEL (STACK), and 32 movies of ZAAG (SAW). Figure 13 presents images of these signs taken from the movies where the signs were recorded in their 'regular' form, i.e. how they were shown and described in a recent SLN dictionary (Nederlands Gebarencentrum, 2002). Furthermore, 21 movies of the SLN sign STRIJKIJZER (an IRON) were made with similar variations to be used as practice material.

Sign selection – The signs were selected from the vocabulary of 121 signs of the ELo sign recognition system (Spaai et al., 2008). Note that the ELo vocabulary contains many signs for visible objects because only signs were included that could be represented easily by a picture. For the current selection, compound signs, such as APPELMOES (APPLE SAUCE), were excluded because they consist of two distinct parts with their own parameters and a specific manipulation of a parameter would change each part in different ways. Signs with extremely high or low locations were excluded because they could get outside of the view of the camera if they were subjected to location manipulations. Of the remaining 66 signs, the five signs used in this study were selected randomly.

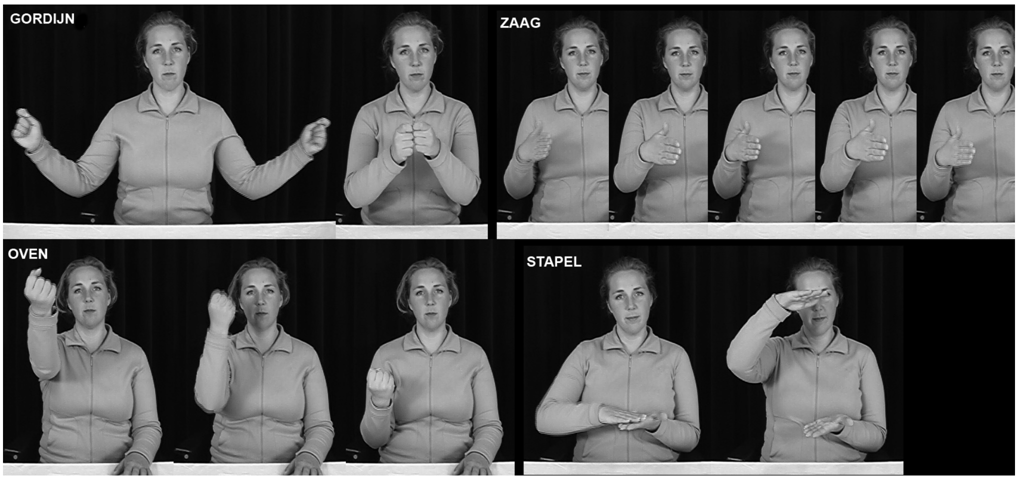


Figure 13. Images of the four test signs.

Table 25. Categories of sign manipulations in the temporal, spatio-temporal and spatial dimension. The numbers (#) can be used to look for more details in Appendix 5-C.

Dimension	Category	#	Manipulation	Dim.	Category	#	Manipulation
temporal	speed	1	slow	spatial	shift position	17	shift right
		2	fast			18	shift left
	hold structure	3	pre-stroke hold			19	shift up
		4	long post-stroke hold			20	shift down
		5	no post-stroke hold		handshape	21	finger selection
	other	6	monodirectional accent			22	finger configuration
						23	finger selection and configuration
spatio-temporal	scale	7	scale up			24	thumb position
		8	scale down		movement path	25	arc added to path
	symmetry	9	symmetry to one hand			26	straight path not arc
		10	symmetry to asymmetry			27	arc in wrong direction
		11	symmetry to alternating			28	setting reversal
		12	one hand to symmetry			29	setting changes reduced
	repetition	13	single to repetition			30	setting changes added
		14	repetition dropped		contact	31	contact to small distance
		15	repetition exaggerated			32	contact to large distance
	other	16	unfinished movement		hand orientation	33	45° relative orientation change
						34	90° relative orientation change
						35	135° relative orientation change
						36	180° relative orientation change
					movement direction	37	path at small angle
						38	path at large angle
					other	39	location hand:palm to neutral space
						40	shift dominant hand
						41	upside down

Phonological specification of the signs – Each of the four test signs was formally described by a phonological specification of the 'regular' form of the sign. Note that the term 'regular' is not intended to imply that this form was or should be the most acceptable form for the participants, it was just the form that we regard as the 'unmanipulated' sign. These specifications can be found in Appendix 5-B.

Sign manipulations – In total 41 types of manipulations were constructed from the four unmanipulated signs: 29 manipulations for GORDIJN (CURTAIN), 34 for OVEN (OVEN), 33 for STAPEL (STACK), and 31 for ZAAG (SAW). This large number of manipulations was included because the work is exploratory in nature. To ensure generalizability over the four signs, we attempted to manipulate each sign in the same way, if possible, on the same parameters. Table 25 shows all the categories of sign manipulations that were constructed. Only six out of the 41 manipulations were considered predominantly temporal. Ten other manipulations had a clear spatial and a clear temporal component. Scaling up, for example, increases the size of the movement path and also influences either the speed or the duration of the movement. The remaining 25 manipulations were considered to be predominantly spatial. It appeared to be much easier to imagine a great variety of spatial variations than to imagine a great variety of temporal ones.

In some cases, it turned out to be difficult to apply a controlled manipulation on a single feature of a sign. For example, movement direction and hand orientation are often strongly related features, depending on how hand orientation is specified. If hand orientation is specified in relation to the signer's body, as is the case in the KOMVA system (NSDSK, 1988), then the two features are independent. But it can also be specified in relation to other parameters of the sign, such as movement direction or location, as is the case in the phonological system proposed by Van der Kooij (2002) and Crasborn (2001). This latter system of specification was used for the treatment of sign features in the current investigation. An example of this treatment is the orientation of ZAAG (SAW), which is specified as 'the ulnar (pinky) side of the hand stands straight on the virtual sawing cut', which also means for the regular form that the fingertips are oriented in the same direction as the movement, i.e. almost straight forward (Van der Kooij, personal communication). Thus, to manipulate the movement direction of ZAAG (SAW) but not the hand orientation the hand must be aligned with the changed movement path with fingertips pointing in the direction of the movement (and not straight to the front).

In some cases, a sign could not be manipulated on the desired parameter in a way that still made sense. For example, consider the 'contact' parameter: In the regular form of the signs GORDIJN (CURTAIN) and STAPEL (STACK) the hands make contact and it makes sense to manipulate this aspect by maintaining some distance between the hands. However, OVEN (OVEN) and ZAAG (SAW) are made with one hand in neutral space and it appears impossible to manipulate them on the contact parameter in a way that makes sense. Thus, specific sign characteristics did not allow each manipulation to be applied to each sign. In addition, some manipulations were applied twice in different directions. For example, a change of 45º in the hand orientation in one direction was also done in the opposite direction, with the exception of OVEN (OVEN) where the hand is already fully supinated in the regular form of the sign. Therefore, the amount of manipulations was not equal for all signs. For a more detailed description of the sign manipulations of each sign, see Appendix 5-C.

Recording - A hearing, late signer performed the signing, based on written descriptions of the desired sign manipulations. She was instructed to look at the camera, to keep a straight face, and to leave out mouthing from the signing. (The recordings were immediately checked, by both signer and experimenter, to ensure that the desired manipulation had been produced. If this was not the case the sign manipulation was recorded anew.) The movies were recorded with three cameras simultaneously. One high-quality digital camera (DV, 3CCD, 720*576 pixels PAL, 25 fps) was used to record the stimulus material for human participants. In addition, two cameras (AVT GUPPY F-033c, CCD, 640*480 pixels, 25 fps), acting as a stereo-camera, recorded the material for the automatic sign recognition methods. An additional signer performed only the unmanipulated versions of the signs, and these were recorded with a single DV camera (these sign movies were used in the part of the procedure where recognition of the used signs was tested).

Procedure - The experiment was presented full screen on a laptop with a 17 inch screen. Experimental software was developed that led participants through the following steps:

- Entering personal data,
- Testing of recognition of the used signs,
- Rating the acceptability of the sign manipulations.

Participants' recognition of the signs was tested to verify whether the form of the sign that we considered to be the 'unmanipulated' sign was indeed a typical example of the intended sign for our participants. This was done by playing a movie of the sign and asking participants to provide the meaning of the sign or indicate 'I do not know this sign' (these movies were recorded with another signer than the one who was recorded in the movies described under 'material'). Three signers indicated that they did not know the sign STAPEL (STACK) because they did not recognize the way it was performed as a typical example, even though they could guess the meaning that was intended. In addition, all three signers believed the same alternative form of the sign to be the correct version. The data of these three signers for the sign STAPEL (STACK) was discarded, but their ratings for the other signs were kept and included in the results.

Before the actual acceptability ratings were collected, people were accustomed to the task by rating a set of movies for STRIJKIJZER (an IRON). During the training, participants followed the exact same procedure as during the actual measurements, no (additional) feedback was provided. After this training, the first series of measurements was recorded. Movies of the four test signs were offered in blocks consisting of all movies of one sign in a randomized order. The order of the blocks was also randomized. Each person rated each movie on a five point scale ranging from 'acceptable' to 'unacceptable' with 'doubt' in the middle. After a full series of movies was rated, the procedure was repeated to obtain a second rating by each person for each movie. For this second series, both the order of the signs and the order of the movies of each sign were newly randomized.

Results

In this section we first present our findings regarding individual participants and groupings of participants based on their characteristics. Based on these findings participants are subsequently treated as one group and the results (across that group) for each sign are given regarding the acceptability of the (categories of) sign manipulations.

Participants - Figure 14 indicates that participants showed a high variability in their usage of the scales and in their tolerance for variation. Tolerance, defined as the percentage of accepted sign manipulations (acceptable + acceptable/doubt) in the total amount of judgments, ranged from 70% to 13%. Some participants used only the extremes of the scale while others used all five points on the scale. Overall, the extreme points on the scale were used more often than the ones in the middle. Data summaries for each individual participant and an overview of their characteristics are given in Appendix 5-A.

The acceptability judgments were checked for consistency by determining the degree of concordance between participants' rankings of sign manipulations on the basis of their acceptability rating. Therefore, we first determined Kendall's rank correlation coefficient τ between the two series of ratings of a participant (intrarater concordance). Subsequently, we determined the interrater reliability (Kendall's τ) and Kendall's coefficient of concordance W for all participants. Finally, we determined Kendall's τ between the rating of each participant and the average rating of all the other participants (open correlation). Each of these measures was determined for the acceptability ratings of the sign manipulations of each sign separately and also for the entire set of sign manipulations (overall).

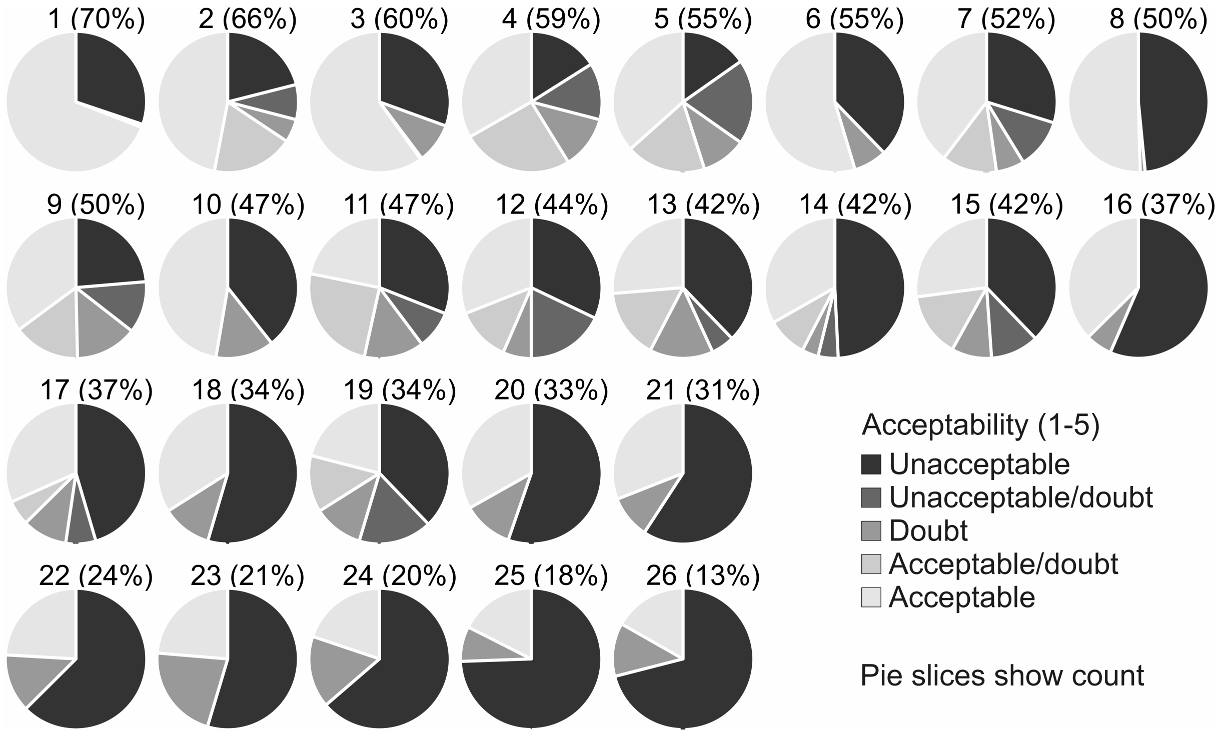


Figure 14. Distribution of acceptability judgments for each participant (across all signs and sign manipulations). Each pie represents one participant, identified by a number from 1 to 26 (these numbers can be used to find more information in Appendix 5-A). Between brackets each participant's tolerance is given.

Table 26. Participant rating concordances. Concordance between the two series of ratings of a participant (intrarater), between ratings of participant pairs (interrater), and between the rating of one participant and the average rating of all the other participants (open correlation) is expressed in Kendall's rank correlation coefficient τ. Kendall's coefficient of concordance W is a measure for the agreement among all participants' rankings.

			Intrarater τ		Interrater τ		Coefficient of Concordance			Open Correlation τ	
Items	**Raters**	**Items**	**avg**	**sd**	**avg**	**sd**	**W**	**χ2**	**p**	**avg**	**sd**
GORDIJN	26	30	.73	.11	.59	.11	.69	522	.000	.67	.08
OVEN	26	35	.70	.11	.52	.14	.61	543	.000	.61	.11
STAPEL	23	34	.80	.11	.67	.09	.78	589	.000	.69	.07
ZAAG	26	32	.72	.12	.59	.11	.70	561	.000	.67	.09
Overall	23	131	.74	.08	.58	.07	.71	2116	.000	.66	.05

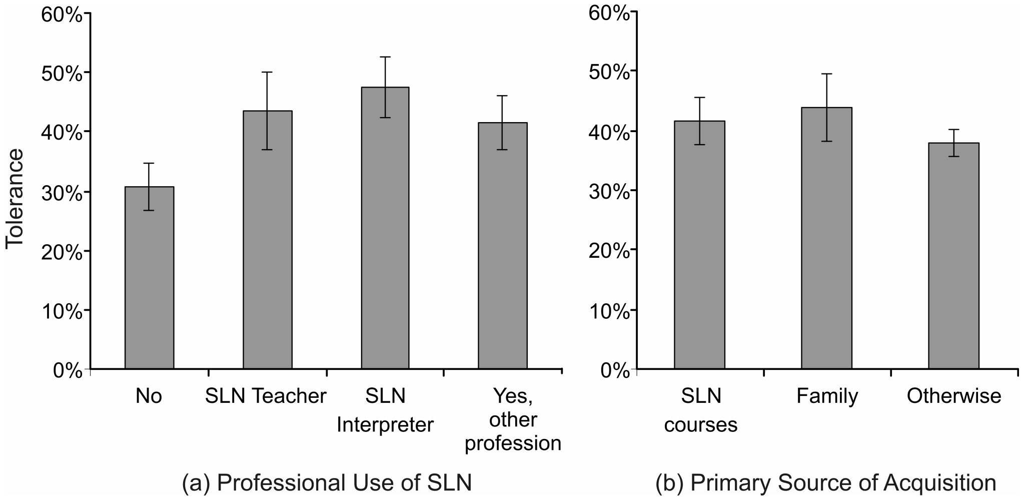
Table 26 summarizes the results. Since the minimum intrarater τ was 0.44, still being a significant value ($p < 0.01$), it is clear that all participants managed to provide consistent rankings of acceptability in their two series of ratings. Therefore, to determine interrater concordance, the ratings of the first and second series were added, creating new values of acceptability ratings on a scale from two (least acceptable) to ten (most acceptable).

The interrater concordance for the acceptability of the sign manipulations of the individual signs showed significant τ values, averaged across all participant pairs (with about 32 sign manipulations per sign a τ value of about 0.30 is significant with an alpha of 0.05 and a τ value of about 0.40 is significant with an alpha of 0.01). Participants agreed most with each other about the ranking of the sign manipulations of the sign STAPEL (STACK) ($\tau_{avg} = 0.67$) and least about those of OVEN (OVEN) ($\tau_{avg} = 0.52$). For OVEN (OVEN) five percent of the interrater τ's is lower than 0.30, the value required for significance ($p = 0.05$), with a minimum interrater τ of 0.04. This means that, especially for OVEN (OVEN), some participants did not agree with some other participants on the ranking of the acceptability of the sign manipulations. For GORDIJN (CURTAIN) and ZAAG (SAW) there were also a few insignificant interrater τ's, but less than 1%. However, the average interrater τ of all four signs is well above the level required for significance, indicating that, on average, participants did agree to a large extent with each other on the rankings. This conclusion is supported by the values of Kendall's coefficient of concordance W, which is a single measure for the interrater concordance of a group of participants (ranging from zero to one, with one being perfect agreement). We found significant values of W (range 0.61 – 0.78) for each sign and overall, see Table 26. This indicates that the participants shared an ordinal scale for the acceptability of these sign manipulations, in a comparable manner as people have been found to share an ordinal scale for the acceptability of spoken language utterances (Mohan, 1977; Cucchiarini et al., 2000; Moustroufas et al., 2007). Finally, the open correlations were about 0.66 with a minimum of 0.41, showing that it is possible to construct a ranking from the average ratings of a group of participants in concordance with the ranking of each individual participant. In conclusion, participants agreed largely on the ranking of the acceptability of sign manipulations and mainly differed in their criterion. Stated otherwise, they used the same acceptability hierarchy but differed in their threshold of what is judged acceptable or unacceptable. Some were stricter while others were more tolerant.

To investigate the effects of participant characteristics on their strictness in judgment we transformed their acceptability ratings into a new Tolerance variable. Tolerance was calculated for each participant as the fraction of 'acceptable' plus 'acceptable/doubt' ratings in their total amount of ratings. Ratings from both series were used separately. The average Tolerance was 42% with a standard deviation of 15% (see Table 30 in Appendix 5-A for individual participants' Tolerance values). We found that none of the participant characteristics had a significant main effect on their Tolerance (tested with one-way ANOVA's): Hearing status ($F_{(1,}$

$_{24)}$ = 0.118, *p* = 0.734, *ns*), parental hearing status ('deaf parents' coincides in our data with 'native signer') ($F_{(1, 24)}$ = 0.935, *p* = 0.343, *ns*), age of SLN acquisition ($F_{(4, 21)}$ = 0.766, *p* = 0.559, *ns*), SLN fluency ($F_{(2, 23)}$ = 0.394, *p* = 0.679, *ns*), and SLN usage ($F_{(2, 23)}$ = 0.569, *p* = 0.574, *ns*), professional use of SLN ($F_{(3, 22)}$ = 0.741, *p* = 0.539, *ns*), primary source of SLN acquisition ($F_{(2, 23)}$ = 0.155, *p* = 0.857, *ns*), gender ($F_{(1, 24)}$ = 0.061, *p* = 0.806, *ns*). Moreover, Tolerance did not correlate with age (Pearson *r* = 0.025, *p* = 0.904, *ns*). On the basis of remarks by signers and incidental observations during previous experiments we expected that interpreters might be more tolerant than teachers, given the nature of their respective professional involvements with SLN, but we found no more than a small, insignificant trend in the expected direction (Figure 15a). Furthermore, we suspected signers who cited 'family' as their primary source of acquisition might be more tolerant than those who would cite 'SLN courses', but again we found only a small, insignificant trend in the expected direction (Figure 15b).

In summary, participants gave consistent acceptability ratings in the sense that they agreed on the rankings of the acceptability of sign manipulations but they differed mainly in their overall tolerance for variation. Some participants accepted many sign manipulations while others accepted only few sign manipulations. We did not find a clear effect of participant characteristics on their tolerance. Therefore the results were collapsed across the participants (for the remainder of this chapter).



(a) Professional Use of SLN

(b) Primary Source of Acquisition

Figure 15. Mean Tolerance of participants grouped according to (a) their professional use of SLN and (b) their primary source of acquisition. Neither factor has a significant effect on Tolerance. The error bars indicate standard errors of the mean.

Sign manipulations - To investigate the acceptability of the different (categories of) sign manipulations we transformed the acceptability ratings into a new Acceptability variable. Acceptability was calculated as the fraction of 'acceptable' plus 'acceptable/doubt' ratings in the total amount of ratings of each sign manipulation of each sign. Ratings from both series of all participants were used. These Acceptability values are given in Appendix 5-C (Table 31 to Table 34).

Figure 16 shows, for each sign, a comparison of the means of the Acceptability of the sign manipulations in the temporal dimension, the spatial dimension, and both dimensions (spatiotemporal) plus the Acceptability of the unmanipulated signs (labelled 'none'). In a two-way ANOVA of the Acceptability of sign manipulations we found a main effect of manipulation dimension, ($F_{(3, 115)}$ = 21.993 *p* = 0.000) but not of sign ($F_{(3, 115)}$ = 0.957, *p* = 0.416) and no interaction ($F_{(9, 115)}$ = 0.288, *p* = 0.977). Post-hoc comparisons of the manipulation dimensions indicated that the Acceptability of temporal manipulations (M = 81%) was higher than the Acceptability of spatial manipulations (M = 30%) (Games-Howell, *p* = 0.000; see Field, 2005), and spatiotemporal manipulations (M = 43%) (Games-Howell, *p* = 0.000). Our temporal manipulations and our unmanipulated signs (M = 90%) were equally acceptable, (Games-Howell, *p* = 0.282). Similarly, there was also no significant Acceptability difference between spatial and spatiotemporal manipulations (Games-Howell, *p* = 0.394).

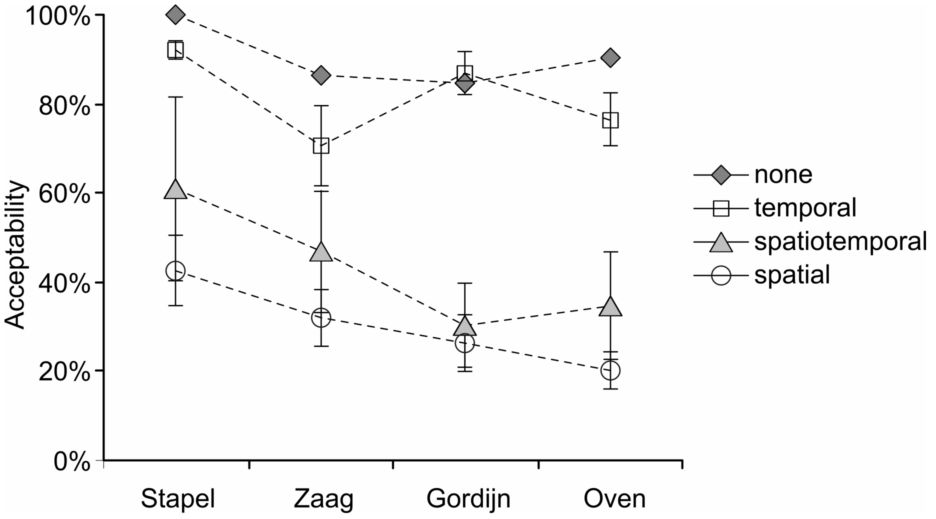


Figure 16. Mean Acceptability of sign manipulations of each sign grouped according to manipulation dimension. The error bars indicate standard errors of the mean.

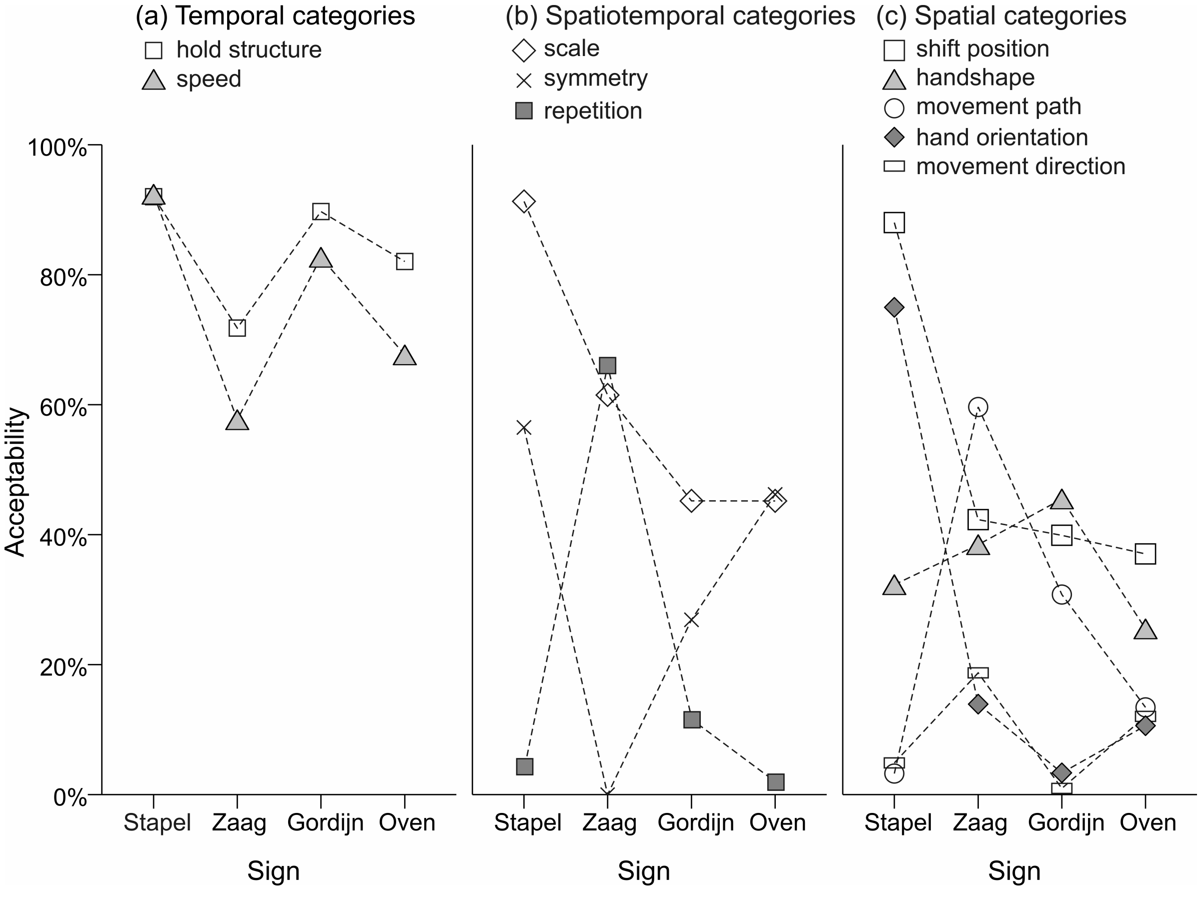


Figure 17. Mean Acceptability of sign manipulations of each sign grouped according to manipulation category (separate markers and lines) and manipulation dimension (separate graphs).

Each manipulation dimension can be divided into several manipulation categories. Figure 17 shows the results according to this division (excluding categories that did not have manipulations for all four signs). Figure 17a shows that, within the temporal dimension, the Acceptability of speed manipulations was somewhat lower than the Acceptability of hold structure manipulations, but this difference was not significant nor was there an effect of sign (two-way ANOVA: Manipulation Category $F_{(1,\ 12)} = 1.662$, $p = 0.222$; Sign $F_{(3,\ 12)} = 3.292$, $p = 0.058$; Sign*Manipulation $F_{(3,\ 12)} = 0.260$, $p = 0.853$). Figure 17b shows the manipulation categories within the spatiotemporal dimension. Here we found significant differences between the manipulation categories but not between signs nor was there a significant interaction (two-way ANOVA: Manipulation Category $F_{(2,\ 8)} = 4.688$, $p = 0.045$; Sign $F_{(3,\ 8)} = 0.783$, $p = 0.536$; Sign*Manipulation $F_{(6,\ 8)} = 2.109$, $p = 0.162$). Figure 17c shows the categories of spatial manipulation and here we found main effects of both sign and manipulation category as well as interactions (two-way ANOVA: Manipulation Category $F_{(4,\ 58)} = 5.967$, $p < 0.001$; Sign $F_{(3,\ 58)} = 2.830$, $p = 0.046$; Sign*Manipulation $F_{(12,\ 58)} = 2.323$, $p = 0.017$). For example, hand orientation manipulations had a high Acceptability for STAPEL (STACK) but not for the other signs, while movement path manipulations had a very low Acceptability for STAPEL (STACK) but a high Acceptability for ZAAG (SAW).

In most cases it is difficult to make general statements about the Acceptability of a certain category of variation (at this level of abstraction where we speak about ‘speed’, ‘repetition’, or ‘handshape’ manipulations) because there is much variability, across signs and across categories. This may be explained in part by the fact that each manipulation category contains a variety of sign manipulations that does not necessarily have to be the same for each sign (as we already noted in our description of the material). This can be seen in more detail, in Table 27, 28, and 29, where, respectively, the temporal, spatiotemporal, and spatial sign manipulation categories are broken down into (still abstract) sign manipulations. Note that these are not the results for the individual movies of sign manipulations, because in some cases a sign manipulation was applied once to a sign (e.g. ‘slow’ for all four signs), in some cases it could not reasonably be applied to a sign (e.g. ‘one hand to symmetry’ for GORDIJN (CURTAIN), a symmetrical sign), and in several cases a sign manipulation was applied twice to a sign (e.g. ‘repetition exaggerated’ was applied twice to ZAAG (SAW), once with two repetitions and once with three repetitions instead of the ‘regular’ single repetition). For the results at the level of individual movies, see Appendix 5-C.

Variability within a category of manipulations is clearly present, for example, in the speed category, where slow manipulations consistently have a lower Acceptability than fast manipulations, and in the hold structure category, where the addition of a pre-stroke hold leads to lower Acceptability than manipulations of the post-stroke hold. See Table 27 for both examples.

Variability across signs can be seen, for example, in the ‘one hand to symmetry’ manipulation (Table 28) which was entirely unacceptable for ZAAG (SAW) but had a medium Acceptability for OVEN (OVEN) and STAPEL (STACK). Likewise, the ‘45° relative orientation change’ (Table 29) had a high Acceptability for STAPEL (STACK) but a low Acceptability for the other three signs. Scaling down (Table 28) was acceptable for STAPEL (STACK) and ZAAG (SAW) but had a low Acceptability for GORDIJN (CURTAIN) and OVEN (OVEN).

There are also examples where variability is present within a category in the form of an interaction between sign manipulations and signs. See for example the shift position category (Table 29). The shifts in four different directions have almost equal overall Acceptability scores. However, shifting the position upwards is quite acceptable for GORDIJN (CURTAIN) but shifting downwards is not, while the reverse holds for ZAAG (SAW). Likewise, shifting to the right is much less acceptable than shifting to the left for OVEN (OVEN) while the reverse holds for ZAAG (SAW).

Table 27. Acceptability of temporal sign manipulations for each sign and across all signs (overall). If a sign manipulation was not applied to a sign the cell is empty. Overall acceptability was only calculated in case a sign manipulation could be applied to at least three signs (else the cell is empty).

Category	#	Manipulation	GORDIJN	OVEN	STAPEL	ZAAG	overall
none	0	none	85%	90%	100%	86%	90%
speed	1	slow	77%	58%	87%	38%	65%
	2	fast	88%	79%	98%	77%	86%
hold structure	3	pre-stroke hold	75%	69%	89%	48%	70%
	4	long post-stroke hold	94%	92%	96%	86%	92%
	5	no post-stroke hold	100%	85%	91%	81%	89%
other	6	monodirectional accent	.	.	.	92%	.

Table 28. Acceptability of spatiotemporal sign manipulations (specifications see Table 27).

Category	#	Manipulation	GORDIJN	OVEN	STAPEL	ZAAG	overall
none	0	none	85%	90%	100%	86%	90%
scale	7	scale up	69%	58%	94%	54%	69%
	8	scale down	21%	33%	89%	69%	53%
symmetry	9	symmetry to one hand	44%	.	.	.	.
	10	symmetry to asymmetry	29%	.	.	.	.
	11	symmetry to alternating	8%	.	.	.	.
	12	one hand to symmetry	.	46%	56%	0%	34%
repetition	13	unrepeated to repetition	12%	2%	4%	.	6%
	14	repetition dropped	.	.	.	27%	.
	15	repetition exaggerated	.	.	.	86%	.
other	16	unfinished movement	.	.	.	8%	.

Table 29. Acceptability of spatial sign manipulations (specifications see Table 27).

Category	#	Manipulation	GORDIJN	OVEN	STAPEL	ZAAG	overall
none	0	none	85%	90%	100%	86%	90%
shift position	17	shift right	42%	21%	89%	42%	49%
	18	shift left	35%	58%	80%	21%	48%
	19	shift up	64%	46%	89%	25%	56%
	20	shift down	19%	23%	94%	81%	54%
handshape	21	finger selection	.	4%	6%	10%	7%
	22	finger configuration	35%	11%	12%	0%	12%
	23	finger sel. and config.	100%	.	.	.	.
	24	thumb position	2%	53%	86%	72%	59%
movement path	25	arc added to path	27%	.	2%	.	.
	26	straight path not arc	.	44%	.	.	.
	27	arc in wrong direction	.	0%	.	.	.
	28	setting reversal	38%	12%	4%	.	18%
	29	setting changes reduced	.	6%	.	.	.
	30	setting changes added	.	.	.	60%	.
contact	31	contact to small distance	46%	.	85%	.	.
	32	contact to large distance	27%	.	48%	.	.
hand orientation	33	45° relative or. change	7%	17%	75%	18%	31%
	34	90° relative or. change	0%	6%	.	10%	5%
	35	135° relative or. change	.	6%	.	.	.
	36	180° relative or. change	.	14%	.	.	.
movement direction	37	path at small angle	2%	13%	10%	29%	15%
	38	path at large angle	0%	11%	0%	9%	6%
other	39	location hand:palm to neutral space	.	.	39%	.	.
	40	shift dominant hand	.	.	94%	.	.
	41	upside down	.	.	0%	.	.

As an example of variability within a category due to different manipulations being applied to different signs, see the repetition category (Table 28). Different repetition manipulations were applied depending on the specification of the (unmanipulated) signs. ZAAG (SAW) is specified with 'repetition' and therefore the sign manipulations 'repetition dropped' and 'repetition exaggerated' were applied; the other three signs have a regular stroke phase (no repetition) and therefore only the sign manipulation 'unrepeated to repetition' was applied. In the results there are large differences between these sign manipulations: 'unrepeated to repetition' has low Acceptability for each of the three signs it was applied to; 'repetition dropped' also had a low Acceptability but 'repetition exaggerated' had a high Acceptability.

Discussion

The results show that people agreed to a large extent with each other about the ranking order of the acceptability of variation in signs but differed in their individual strictness. This ranking order hardly revealed any regularity with respect to the categories of sign manipulations except that temporal variation was systematically found to be highly acceptable. Overall, temporal sign manipulations were judged to be just as acceptable as our unmanipulated signs; although there were some exceptions (e.g. a 'slow' ZAAG (SAW) had a medium acceptability). An explanation for this finding could be that the temporal manipulations (the speed and the hold structure) did not hurt the visibility of motion boundaries (Rubin & Richards, 1985). In most cases one can represent a sign by a beginning and an end state (Marr and Vaina, 1982; Van der Kooij, 2002), with motion boundaries being used to identify states (Rubin & Richards, 1985). Then, consider the introduction of a pre-stroke hold. Although it will likely increase the strength or the visibility of the motion boundary, a pre-stroke hold is (theoretically) not *necessary* for a visible motion boundary, because any force discontinuity or any start of movement of a part of the hand theoretically suffices (Rubin & Richards, 1985). Thus, a pre-stroke hold does not alter the way in which one might represent the movement (either in SLN Phonology or in State-Motion-State (Marr & Vaina, 1982)) because it does not change the states. If the motion boundaries remain visible and if the (spatial) characteristics of the states are untouched it is reasonable to expect very little impact on the acceptability of such temporal manipulations.

Variations that involved spatial aspects, or both temporal and spatial aspects, had a low overall acceptability but showed much variability and hardly any regularity at the level of fairly abstract categories such as 'scale', 'repetition', 'handshape', etc (see Figure 17). A more detailed division of these categories into less abstract sign manipulations, such as 'scale up', 'one hand to symmetry', '45° relative orientation change', etc. (see Table 27 to Table 29) showed that none of these sign manipulations had a consistent acceptability across our small sample of four signs, perhaps with the exception of manipulations of the movement direction, specifically 'path at large angle', which consistently had a low acceptability.

The variability that exists within categories and across signs may be better understood by studying the roles of phonology and iconicity in judgments of acceptability. As an example of a possible effect of phonology, consider the repetition manipulations: dropping a required repetition or adding an unspecified repetition are violations of the phonological specification which may explain their low acceptability; exaggerating the number of repetitions is not considered a phonological error but merely free phonetic variation. As an example of a possible effect of iconicity: the reason why it is acceptable to shift GORDIJN (CURTAIN) upwards but not downwards may lie in the sign's (iconic) enactment of 'closing the curtains' which is easily performed if you grab the curtains high but becomes difficult if you grab them low. Likewise, ZAAG (SAW) mimics the saw itself as well as the motion of the act of sawing, and the fact that it is hard to saw at an inconvenient height may cause the low acceptability of an upward shift. In the next section, such effects of phonology and iconicity are studied more systematically.

5.3. Phonology and Iconicity

Introduction

To investigate the impact of phonology and iconicity on acceptability, a framework was developed based largely on the proposal for SLN phonology by Van der Kooij (2002) and on discussions with her and her colleagues Johan Ros, Inge Zwitserlood and Onno Crasborn (although we drew our own conclusions from the discussions and any misinterpretations are therefore entirely our own). First, we identified factors in the area of phonology and iconicity that may influence the acceptability of sign variation and formulated hypotheses about this influence. Next, we applied a simplified classification scheme to our sign manipulations. Each movie of a sign manipulation was classified on two levels: the severity with which it breached the sign's phonological specification (*phonological error*) and its *iconic compatibility*. Then, we studied whether the Acceptability of the (movies of) sign manipulations, as calculated in the previous section, can be explained in terms of phonological error and iconic (in)compatibility.

Additionally, we investigated whether there is a difference between native and non-native signers in their sensitivity to phonology and iconicity and the possible impact on acceptability judgments. Differences between native and non-native signers with regard to the role of phonological information have been suggested previously. Emmorey and Corina (1990), for example, concluded that non-native signers, whom they found to be delayed in comparison to native signers in a task of lexical recognition, might be less able to apply phonological rules to rule out alternatives and hence quickly identify a sign. Hildebrandt and Corina (2002) reported that, overall, native and non-native signers showed the same preferences in making (phonological) similarity judgments, but they also noticed some minor differences between the groups with native signers apparently paying more attention than non-native signers to similarity in location (in some cases). Dye and Shih (2006) reported that native signers appeared to use phonological information in signs during lexical access but they could not find evidence for this with non-native signers. As far as we know, similar observations for iconicity have not been reported before.

Factors – The following factors were expected to influence the acceptability of variation in signs:

a. *Phonological specification* of a sign determining the phonological categories that should be visibly present in the sign production or not. We adopted the SLN Phonology by Van der Kooij (2002).
b. *Phonetic implementation rules* predicting the phonetic interpretation of the phonological specification. These rules include *default* phonetic implementations of phonological features and rules governing how a phonological feature should be implemented phonetically in the context of other phonological features. Phonetic implementation rules can be considered part of the SLN Phonology and are also described in Van der Kooij (2002).
c. *Default phonetic articulations*, which typically occur most frequently and are easily articulated (while remaining visually salient). These defaults are partly documented by Van der Kooij (2002). In some cases we relied on our own estimate of how a movement is produced most economically (see Appendix 5-B).
d. *Iconicity* or 'semantic motivation' of (aspects of) a sign being operational on the level of aspects, such as a motivated location (e.g. Van der Kooij 2002; Van der Hulst & Van der Kooij, 2006) as well as on the level of the whole sign (following strategies for iconicity in gesture as described by Müller, 1998).
e. Possible *modulations* of the sign, where additional or alternative aspects of meaning can be coded in the sign by changing aspects of the form. Macken et al. (1993) discussed, for American Sign Language (ASL), several ways in which the form and meaning of a sign

can be modified using strategies that rely on iconic interpretations. Zwitserlood (2003) also discussed, for SLN, hand configurations that can be treated as 'classifiers', which we also considered to be modulations. (Examples are given in Appendix 5-B).

f. *Native signers* and *non-native signers* might differ in the way in which phonology or iconicity mediates their acceptability judgments.

Hypotheses - Deviations from the specified, predicted, or default articulation of an aspect of the form of a sign were thought to be judged according to the following rules:

Regarding phonology and phonetic implementation:

1. Variations that are not in conflict with the above-mentioned specifications, or with any known phonetic implementation rule, will be referred to as *free phonetic variation*. These are expected to have no impact on acceptability. For example, if GORDIJN (CURTAIN) is made slower or faster than 'usual', and there is no known phonetic implementation rule that governs speed, then this variation is classified as harmless phonetic variation.
2. Violations of phonetic implementation rules or cases where the form of the sign deviates from the expected (default) phonetic interpretation will be referred to as *phonetic implementation errors*. These are expected to decrease acceptability. For example, if GORDIJN (CURTAIN) is not made in the center of the neutral space, which is the predicted phonetic implementation, but about 15 cm to the right then this is considered a phonetic implementation error.
3. Violations of the phonological specification, in the sense that a different value (including null values) of a feature is present will be referred to as *phonological errors*. These are expected to severely decrease acceptability (more than phonetic implementation errors). For example, if GORDIJN (CURTAIN) is repeated then the signal can be said to contain the phonological feature [repetition] which is not in the sign's specification and is therefore a phonological error. Phonological categories are by definition lexically distinctive and any phonological error may mean that the form of the sign shifts to another sign's form (if two signs differ only on a single feature and this feature is manipulated). We identified some such cases but they will not be treated separately here.

Regarding iconicity:

1. Variations that are fully compatible with the sign's iconicity will be referred to as cases with *iconic compatibility*. These are expected to have a high acceptability level.
2. Variations that are known modulations of (the meaning of) a sign are acceptable when used in the proper context but unacceptable if used otherwise. We will refer to such variations as cases with *'contextual iconic compatibility'*. For example, GORDIJN (CURTAIN) is usually made with a 'palm grasp' handshape but one may also use a 'pinch grip' handshape which may be considered a modulation that represents a light curtain. In our study, in which no context is offered, there is no indication of whether a certain modulation was intended or not. Assuming that people may differ in the extent to which they are willing or able to imagine a suitable context, we hypothesized that sign manipulations with contextual iconic compatibility will show a high variability across participants in acceptability judgments. We also expected to find a low acceptability level for these sign manipulations in comparison to those with (full) iconic compatibility. Note that some ways to modulate a sign are less frequent and less conventionalized than others. It is therefore possible that certain variations, which are classified as iconically incompatible, can be interpreted as a modulation in exceptional contexts.
3. Violations of iconic aspects of a sign will be referred to as cases with *iconic incompatibility*. Violations of the semantic motivation of the whole sign, whereby the gestural iconicity (in the sense of Müller, 1998) is put at risk also constitute *iconic incompatibility*. Iconic incompatibility is expected to severely decrease acceptability (more than contextual iconic compatibility). For example, an arc is supposed to be in the path movement of OVEN (OVEN) because it is an enactment (a form of iconicity) of the opening of an oven

door with a turning axis at the bottom of the door. Failing to produce this arc, or making the arc in the wrong direction, creates a gesture that is incompatible with this sign's iconicity.

Classification - Each movie of a sign manipulation was classified as belonging to one of three levels of phonological error, with increasing expected negative impact on acceptability:

1. Free phonetic variation
2. Phonetic implementation error
3. Phonological error

Each movie of a sign manipulation was also classified as belonging to one of three levels of iconic compatibility, again with increasing expected negative impact on acceptability:

1. Compatible
2. Contextually compatible (modulations)
3. Incompatible

The classifications of the movies of the four signs were made by the author and can be found in Appendix 5-C. based on the proposal for SLN Phonology by Van der Kooij (2002), on the sign specifications and considerations regarding iconicity and phonetic implementation (see Appendix 5-B), and, for iconic compatibility, on remarks made by several signing informants. It should be noted that the proposal for phonological categories for SLN by Van der Kooij (2002), even though it is quite extensive, sometimes leaves room for interpretation during classification.

Results

Figure 18 shows the effects of (a) phonological error and (b) iconic compatibility on the Acceptability of the sign manipulations of each of the four test signs. Separate two-way ANOVA's were performed (for phonological error and sign and for iconic compatibility and sign).

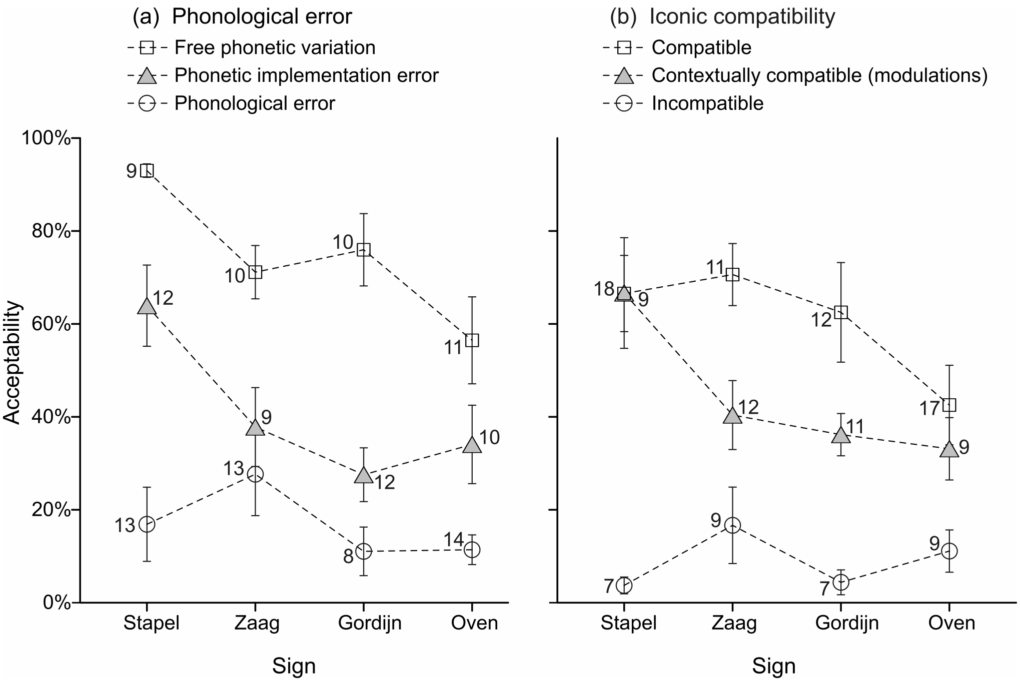


Figure 18. Mean Acceptability of sign manipulations of each sign grouped according to (a) phonological error and (b) iconic compatibility. The error bars indicate standard errors of the mean. The numbers next to the markers indicate the number of cases.

Phonological error and sign - We found a main effect on Acceptability of phonological error ($F_{(2, 119)} = 59.710$, $p < 0.001$) and of sign ($F_{(3, 119)} = 6.206$, $p = 0.001$) and no interaction ($F_{(6, 119)} = 2.116$, $p = 0.056$, *n.s.*), see Figure 18a. Post-hoc comparisons of phonological error indicated significant differences between each of the three levels (for all three pairs: Games-Howell, $p < 0.001$). The overall Acceptability was lower for phonological errors (Mean (M) = 17%) than for phonetic implementation errors (M = 41%) which in turn had a lower Acceptability than free phonetic variation (M = 73%). Despite the main effect of sign found in the ANOVA, post-hoc comparisons (Games-Howell) for sign did not indicate a significant difference between any of the pairs of signs. (Repeating these tests with just the 106 movies of spatial and spatiotemporal sign manipulations yielded the same main effects, but due to a lower mean Acceptability of free phonetic variations (M = 60%) they no longer differ significantly from the phonetic implementation errors (M = 41%).)

Iconic compatibility and sign - We found a main effect on Acceptability of iconic compatibility ($F_{(2, 119)} = 35.250$, $p < 0.001$) but not of sign ($F_{(3, 119)} = 2.350$, $p = 0.076$, *n.s.*) and no interaction ($F_{(6, 119)} = 1.607$, $p = 0.151$, *n.s.*), see Figure 18b. Post-hoc comparisons of iconic compatibility indicated significant differences between each of the three levels. Movies of sign manipulations with iconic incompatibility (M = 10%) were less acceptable than those with contextual compatibility (M = 43%) and those with iconic compatibility (M = 59%) (in both cases Games-Howell, $p < 0.001$). The difference between contextual compatibility and (full) iconic compatibility was also significant (Games-Howell, $p = 0.030$). (These tests were repeated with only spatial and spatiotemporal sign manipulations which also showed a main effect of iconic compatibility. However, (full) iconic compatibility (M = 42%) now had equal Acceptability as contextual iconic compatibility (M = 43%).)

Iconic compatibility and variability across participants – To investigate whether sign manipulations with contextual iconic compatibility have a high variability across participants in acceptability judgments, the variability across participants was determined for each movie of a sign manipulation as the standard deviation of the acceptability ratings given by all participants (the sum of a participant's ratings in the two series of judgments was used, resulting in a scale from 2 to 10). Iconic compatibility was found to have a main effect on this variability in acceptability judgments ($F_{(2, 128)} = 15.648$, $p < 0.001$) across participants. Sign manipulations with contextual iconic compatibility had a higher variability across participants (M = 2.46) than those with iconic incompatibility (M = 1.49) and those with (full) iconic compatibility (M = 1.87) (in both cases Games-Howell, $p < 0.001$). There was no significant difference between the latter two categories.

Phonological error, iconic compatibility and acceptability - Figure 19 shows Acceptability as a function of phonological error and iconic compatibility. Several observations can be made about the way phonological and iconicity influence acceptability. First, it should be noted that phonological error and iconic compatibility were not independent classification systems as is shown by the unequal distribution of cases across the nine possible combinations of the two variables (Pearson chi-square$_{(4)}$ = 36,316, p < 0.001). Most cases (65%) can be found 'on the diagonal' suggesting that phonological error and iconic compatibility are closely related to each other. This is supported by the fact that a case classified as 'free phonetic variation' while having 'iconic incompatibility' is rare (only 2 cases, or less than 2%). This dependency can be explained by the fact that the phonological theory we applied, in the sign specifications and in the classification of the sign manipulations, contained references to iconic aspects of the sign. Hence, if a sign manipulation violated a reference to a specified iconic aspect, it would also tend to constitute a phonological error. But note that not all considerations regarding iconicity were covered by the phonological specifications (see Appendix 5-B).

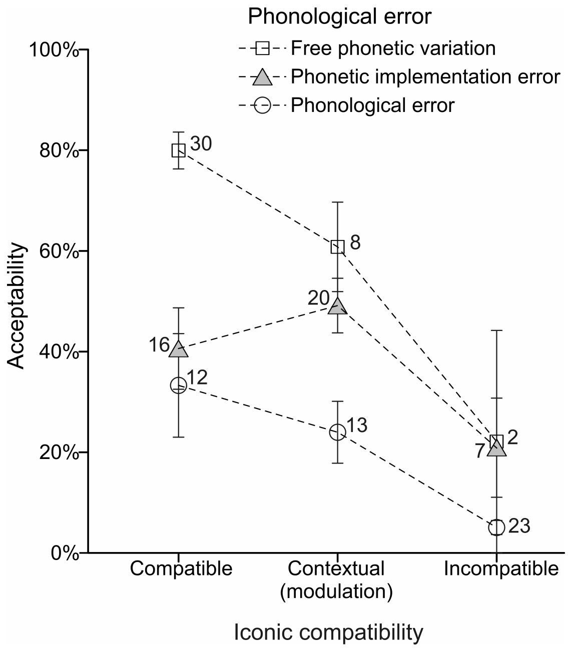


Figure 19. Mean Acceptability of sign manipulations as a function of phonological error and iconic compatibility. The error bars indicate standard errors of the mean. The numbers next to the markers indicate the number of cases.

Next, we found that, when combined, phonological error and iconic compatibility both had a significant effect on Acceptability without a significant interaction (two-way ANOVA; *phonological error* $F_{(2, 122)} = 11.682, p < 0.001$; *iconic compatibility* $F_{(2, 122)} = 11.455, p < 0.001$; *interaction* $F_{(4, 122)} = 1.947$, $p = 0.107$, *n.s.*). All pairwise post-hoc comparisons between the levels of phonological error as well as between the levels of iconic compatibility showed significant differences (Games-Howell, $0.000 < p < 0.030$). Together, the two factors explained 53% of the variance (adjusted R^2). Taken separately phonological error explained 44% of the variance (adjusted R^2) and iconic compatibility 32% (adjusted R^2). A stepwise regression analysis indicated that adding iconic compatibility, after entering phonological error, made a significant contribution to the explained variance (R^2 change = 8%, $F_{(change)} = 21.208, p < 0.001$)). Despite the lack of a significant interaction the following observations can be made about Figure 19:

- The distinction between phonetic implementation errors and free phonetic variation seems to be relevant in combination with (full) iconic compatibility and not if there is iconic incompatibility or contextual iconic compability. Conversely, the distinction between phonetic implementation errors and phonological errors appears to be mostly relevant if there is contextual iconic compatibility or iconic incompatibility; i.e. phonetic implementation errors and phonological errors yielded almost equally low Acceptability in case there is (full) iconic compatibility.
- The distinction between (full) iconic compatibility and contextual iconic compatibility appears to be relevant for free phonetic variations and less if there is a phonological error or a phonetic implementation error.
- A trend can be seen that sign manipulations with iconic incompatibility have very low Acceptability, even if they are not phonological errors; the Acceptability means of all three possible combinations with iconic incompatibility (the three markers on the right) were lower than the means of all six other combinations (although it should be noted that there were only 2 cases in the combination of 'iconic incompatibility' with 'free phonetic variation').

Native versus non-native signers – Participants reported several indicators of linguistic experience. Four of them could be classified as native signers (deaf parents, 'fluent' in SLN, and SLN acquisition started before the age of two). These were compared with nineteen other participants who were classified as non-native signers (hearing parents, varying SLN fluency and varying age of SLN acquisition, excluded were three signers who did not recognize the unmanipulated version of STAPEL (STACK) as such). For this comparison, separate Acceptability scores, one for the native and one for the non-native signers, were calculated for each of the 131 movies of sign manipulations.

For both native and non-native signers phonological error and iconic compatibility had significant main effects on Acceptability (two-way ANOVA for native signers: *phonological error* $F_{(2, 122)} = 7.673$, $p = 0.001$; *iconic compatibility* $F_{(2, 122)} = 12.480$, $p < 0.001$; *interaction* $F_{(4, 122)} = 1.511$, $p = 0.203$, *n.s.,* adjusted $R^2 = 47\%$; two-way ANOVA for non-native signers: *phonological error* $F_{(2, 122)} = 12.141$, $p < 0.001$; *iconic compatibility* $F_{(2, 122)} = 10.412$, $p < 0.001$; *interaction* $F_{(4, 122)} = 1.767$, $p = 0.140$, *n.s.* adjusted $R^2 = 52\%$). However, there was one difference between the native signers and the non-native signers: native signers judged sign manipulations with contextual iconic compatibility to be equally acceptable as those with (full) iconic compatibility (Games-Howell, $p = 0.954$, *n.s.*) while non-native signers did not. All the other pair wise post-hoc comparisons between the levels of phonological error and between the levels of iconic compatibility showed significant differences (Games-Howell, $0.000 < p < 0.012$), in line with the overall findings. (When these tests were limited to spatial and spatiotemporal manipulations the same main effects were found and the pattern of the results remained the same.)

Figure 20 shows that the native signers were somewhat more tolerant; they registered higher Acceptability scores in each class of phonological error and of iconic compatibility. A paired-samples t-test on the scores of both groups for each of the 131 movies confirmed the significance of this difference ($t_{(130)} = 4.270$, $p < 0.001$, $M_{native} = 48\%$, $SD_{native} = 37\%$, $M_{non\text{-}native} = 41\%$, $SD_{non\text{-}native} = 36\%$) (a result that remained unchanged when the test was limited to spatial and spatiotemporal manipulations).

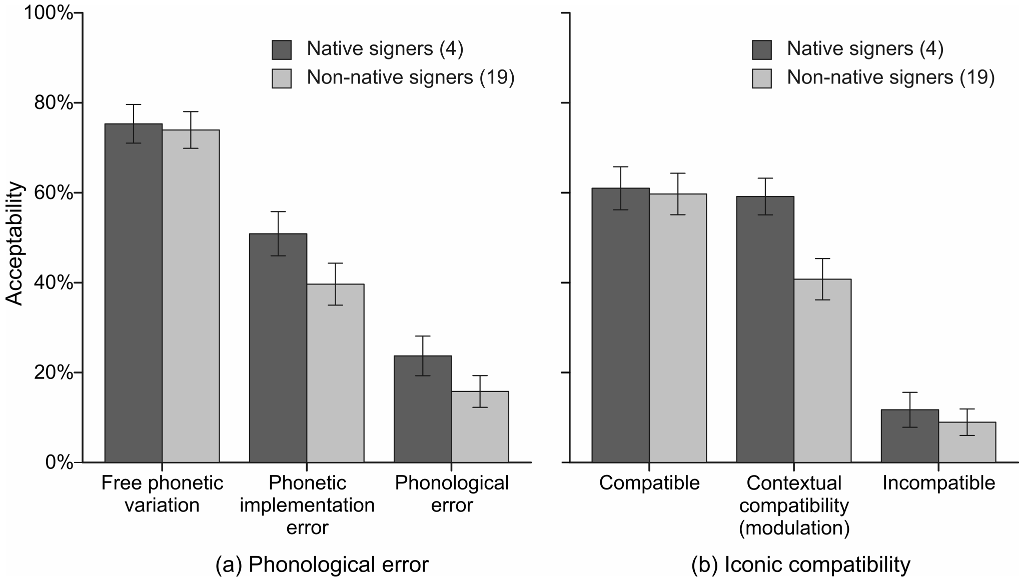


Figure 20. Mean Acceptability of sign manipulations as judged by four native signers and nineteen non-native signers grouped according to (a) phonological error and (b) iconic compatibility. The error bars indicate standard errors of the mean.

Discussion

In this section we showed that available knowledge about SLN phonology, specifically as proposed by Van der Kooij (2002), could be applied productively to predict a significant part of the variability in Acceptability judgments of our set of sign manipulations of four signs. Given the general nature of the proposed SLN phonology, it is reasonable to expect that this finding will generalize to other sets of signs, and perhaps also to other sets of variations of such signs. We have also shown that, next to phonology, iconicity plays an important role in the acceptability judgments of our material. Therefore, to predict the acceptability of variations in other sets of signs, it seems wise to include iconicity of the signs (if any) as well.

Phonological error and phonetic implementation error – The low Acceptability of (full) phonological errors is in line with the expectations formulated on the basis of the applied phonological theory (Van der Kooij, 2002). The separate treatment of phonetic implementation errors as an intermediate level of error severity appears to be useful to predict Acceptability more accurately. However, the impact of phonetic implementation errors was found to depend somewhat on iconic compatibility. Perhaps, in further research, a finer distinction should be made between different types of modifications (e.g. as described by Bellugi and Fischer (1972), Macken et al. (1993), or Zwitserlood (2003)); a distinction that takes the relationship between a modification and the sign's iconicity into consideration.

Iconic incompatibility - Our findings regarding the strong decrease of Acceptability due to iconic incompatibility (regardless of the level of phonological error), are in line with the expectations regarding the role of iconicity formulated by Van der Kooij (2002: chapter 6). She stated that it makes sense that iconicity takes precedence over phonetic implementation rules because iconicity is lexically specified whereas the rules are only general (she found similar results in a study on the Acceptability of variation in the exact place on the chest of a set of signs). In accordance with the expectation formulated in the introduction we found more variability across participants in their acceptability judgments of sign manipulations with contextual iconic compatibility, indicating that there may be differences in people's ability or inclination to imagine a suitable context for the sign manipulation.

Native signers' high tolerance – We found that native signers showed a higher overall tolerance for variation than non-native signers. We can only speculate why: Perhaps the phonological rules of signed languages are less restrictive than those of other languages with which the non-native signers had more experience, possibly causing those non-native signers to apply the SLN phonological rules they know too strictly. Or perhaps the participating native signers simply had a more tolerant personal attitude towards variation, confidently allowing more from the signer in our material.

Sorace (1996), who treats differences between native and non-native Acceptability judgments extensively, states that one might expect greater variability in the non-native Acceptability judgments. This could not be observed in our data.

Contextual iconic compatibility (modulations) and native signers - We found that native signers were tolerant for manipulations with contextual iconic compatibility, judging them equally acceptable as those with (full) iconic compatibility. In contrast, non-native signers judged manipulations with contextual iconic compatibility less acceptable than sign manipulations with (full) iconic compatibility. One explanation for this finding could be that modifying signs (and thereby creating contextual iconic compatibility) does not harm them but in fact enriches them in a (fairly advanced) grammatical way and, second, that the native signers were more aware of this than the non-native signers. Bellugi and Fischer (1972) provided a thorough description of many ways in which ASL users can (grammatically) modify signs to enrich their meaning, which they called 'incorporation'. They suggested this is one of the main strategies used by accomplished signers to reach a 'rate of propositions' comparable to that of speakers, despite the much lower production rate of signs (as compared with words). Likewise, Macken et al.

(1993) suggested that ASL is best treated as a combination of arbitrary conventional symbols and what they call 'richly grounding symbols' (signs with iconicity) whose meaning can be modified in accordance with their iconicity. More recently, the dissertation of Zwitserlood (2003) presented rules by which different hand configurations can signify different classes (e.g. of objects) in SLN. We do not know to what extent the modifications in our material (manipulations with contextual iconic compatibility) were grammatical but the high acceptability rating by native signers may be an indication that many were. Clahsen and Felser (2006) showed that non-native language users do not always have difficulty with grammar but typically do have problems with the more complex rules. If we assume that the rules by which modifications are allowed are fairly complex (or at least not easily acquired) then it is likely that they were often not part of the internalized grammar of the non-native signers. In short, it may be that the non-native signers did not know, in many cases, that a modification was allowed and therefore rated it 'unacceptable' more often than the native signers who did know.

A factor that may also have played a role in judging the acceptability of our manipulations with contextual iconic compatibility is that the form of the sign may suggest a context to those who use their imagination to try to interpret and judge it. Imagining a suitable context probably invokes a high degree of 'imagery'. Levelt et al. (1977) found that 'high imagery' material tended to be rated more acceptable than 'low imagery' material (they also highlighted the importance of 'trying to find a possible context' in a task of acceptability judgments). Perhaps native signers are more adept at imagining suitable contexts for modified sign forms (or more willing to use their imagination) and therefore experience a higher degree of 'imagery' as they observe them. Additional research that targets this aspect directly would be required to examine this possibility.

In summary, the two explanations offered rely on (grammatical) rules and conventions and on people's imagination, respectively. Obviously, these are not mutually exclusive. In any living language inventions (e.g. newly formed signs or ways to combine signs or modify them) become conventionalized, see for example Sexton's (1999) study on grammaticalization in ASL. Native signers may be more in touch with such processes than non-native signers. For example, we may note that poets are usually very accomplished (typically native) users of a language, and one can regard (experimental) poetry as an active search for innovations in a language, thereby enriching it. For signed languages such poets are also at work, see for example the work of the Dutch (SLN) poet Wim Emmerik (e.g. in Koenen et al., 1993 or in Emmerik and Meyer (2005)) or the American (ASL) poet Peter S. Cook (Cook, 1998).

5.4. Human versus machine ratings of acceptability

Introduction

In this section the 'human acceptability judgments' of 26 signers, as reported in the first part of this chapter, are compared with 'machine acceptability ratings' by three sign recognition methods (sign recognizers). These sign recognizers were all developed for an application called ELo (Electronic Learning environment) in which automatic sign recognition is used to provide feedback (based on a calculated acceptability rating) to deaf and hard of hearing children practising signs (Lichtenauer et al., 2007; Spaai et al., 2008; Lichtenauer et al., 2008). ELo was developed for young children, age 3-5 years, but similar applications have been developed for somewhat older children (Brashear et al., 2006) and for adult sign language learners (Zieren, 2007).

As Sorace (1996) pointed out and as we found in our data, human judgments of the acceptability of linguistic variation are not consistent in an absolute sense, but given a set of items humans do agree to a large extent on the rank order of their acceptability. One could think of a decision (by human or machine judges) to either accept or reject some set of sign variations as a stepwise procedure of (1) determining the relative acceptability of the sign variations (a ranking from highly acceptable to highly unacceptable) and (2) applying some threshold or 'strictness' parameter depending on context.

In automatic sign recognition, the second part of this process is already a common procedure (e.g. Lichtenauer et al. 2007, Von Agris et al. 2008). Sign recognizers can usually be configured by setting a threshold which determines at what level the raw 'likelihood' scores (the calculated similarity of the test sign to the target sign 'class') are split into 'positive' or 'negative' ('accepted' or 'rejected'). Being able to set an appropriate threshold is often important, for example to adjust the threshold for a child's age in an electronic learning environment (e.g. Lichtenauer et al. 2007). But to avoid mistakes in the sense of calling a properly produced sign unacceptable (false negative) while calling a badly produced sign acceptable (false positive) it is crucial that the scores of the sign recognizer for the sign productions are ranked in the same order as human judges would rank them (the first part of the procedure). In fact, in the example of the electronic learning environment, the only real task of the sign recognizer is to get this ranking right, because setting appropriate thresholds should be under the control of a teacher who knows the children.

Related work with automatic speech recognition – Automatic speech recognition methods (that employ similar techniques as automatic sign language recognition) are applied in spoken language learning applications to provide automatic ratings of the acceptability of words, sentences or overall pronunciation. Several studies report evaluations of such applications (Bernstein et al., 1990; Witt & Young, 2000; Neumeyer et al., 2000; Franco et al., 2000; Cucchiarini et al., 2000; Huang et al., 2007; Moustroufas et al., 2007; Cincarek et al., 2009). In all of these studies the automatic acceptability ratings are compared to human acceptability judgments by determining both human-human and human-machine correlations (the human-human correlations serve as a benchmark). For example, Moustroufas et al. (2007) report human-human 'open correlations', that is, the correlation between a rater and the average of all the others, with values ranging from 0.63 to 0.73 (for acceptability ratings of sentences). Moustroufas et al. (2007) determined human-machine correlations for various speech recognition methods and report that, at best, the speech recognizer gives acceptability ratings of sentences which correlate rather well with human ratings (about 0.49) but below the level of human-human correlations.

Method

Material - The material consisted of a selection of 68 movies from the original set of 131 movies. Movies of three of the four SLN signs were used: 21 movies of GORDIJN (CURTAIN), 23 movies of OVEN (OVEN) and 24 movies of ZAAG (SAW). Movies with variations on handshape and hand orientation were excluded from the original set of 131 movies of sign manipulations because the used sign recognition methods did not extract such features. Movies of the sign STAPEL (STACK) were also excluded because the feature tracking results of these movies contained many errors, and were therefore not suitable as input for the sign recognition methods to calculate automatic acceptability ratings (the recognizer actually calculates 'likelihood scores' which are used here as acceptability ratings).

Sign recognition methods – The three sign recognizers differed in how they were trained and how they calculated their acceptability ratings but they all had the following in common (Lichtenauer et al., 2008): Signs were recorded with two cameras (which act as a single 'stereo camera'). The video is analysed to detect the face and the hands and then a 3D location of the hands is determined from the difference between the cameras' images. Next, a set of 25 features is extracted, for example the left and right hand motions and the size changes (the handshape is not extracted). The signal is then 'time warped', meaning it is aligned with a fixed length feature model of the sign that ELo's recognizer is expecting to receive as input based on its training. All sign recognizers were trained on a set of positive examples (75 different people producing a target sign) and a set of negative examples (75 people producing 119 other signs). After time warping the sign recognizer calculates the acceptability rating: the likelihood that the input is an (acceptable) example of the expected sign.

SDTW - The first sign recognition method is called Statistical Dynamic Time Warping (SDTW) because it uses a byproduct of the SDTW procedure, namely the degree to which the input fitted the model, to calculate automatic acceptability ratings.

CDFD - The second sign recognition method uses the same time warping as SDTW, but to calculate the automatic acceptability ratings the Combined Discriminative Feature Detectors (CDFD) method (Lichtenauer et al., 2008) leaves many features out of the calculation. A reduced set of discriminative features is selected during training for each sign because, in the eyes of the machine, they are distinctive for this sign within the set of training signs.

Q-DFFM – This sign recognizer uses the same time warping as SDTW and CDFD and also the same method of selecting discriminative features as CDFD. However, calculating automatic acceptability ratings is now done differently. The method of 'Quadratic classification on a Fisher Mapping of Discriminative Features' (Q-DFFM) (Lichtenauer et al., 2008) takes dependencies between features into account, while CDFD treats features as independent from each other.

Results

The automatic acceptability ratings calculated by the sign recognizers did not use the same scale and the scales were not all linear. Therefore, we determined the ranking order of the ratings from low to high acceptability and this was done separately for each sign. The rank of each of the movies is given in appendix 5-C. Similarly the human Acceptability scores, as defined in the previous section, were used to determine a human acceptability ranking.

Figure 21 shows the human-machine ranking scatterplots for each sign and each sign recognition method (see Arendsen et al. (2008) for more details). The average ranking correlation (Kendall's τ) between machine and human rankings was 0.30 (SD = 0.20). For GORDIJN (CURTAIN) and ZAAG (SAW) the machine rankings correlate with the human rankings (range 0.33 to 0.51), but for OVEN (OVEN) none of the three machine rankings correlated to the human ranking (range -0.08 to 0.13). SDTW had a slightly higher average correlation (across signs) with human rankings (0.32) than Q-DFFM (0.31) and CDFD (0.25), but the performance of the different sign recognizers is quite variable between signs, with Q-DFFM scoring higher correlations than SDTW for ZAAG (SAW) and OVEN (OVEN).

If the human-machine ranking correlations are limited to the spatial and spatiotemporal sign manipulations (denoted by τs in Figure 21) then we see that the degree of correlation remains unchanged for ZAAG (SAW). For GORDIJN (CURTAIN) there are no longer (significant) correlations indicating they depended largely on human-machine agreement about the high acceptability of the temporal manipulations for that sign. In addition it can be observed that, for the spatial and spatiotemporal manipulations, the Q-DFFM method clearly outperforms the other methods (for all three signs).

To aid the interpretation of the human-machine ranking correlations we also determined, for the same set of movies, the human-human open correlations (between the rankings of one human and the average of the other humans) for each sign and these had an average of 0.68 (SD = 0.10, Minimum = 0.42) across signs and participants. In addition, the machine-machine ranking correlations were calculated for each sign and these had an average of 0.61 (SD = 0.12), taken across signs and sign recognizers.

Discussion

In this section we have evaluated whether the automatic acceptability ratings given by three different sign recognizers to a set of 68 movies of highly variable sign productions are ranked in the same way as the human acceptability ratings of these movies.

Overall, we have found low human-machine ranking correlations in comparison to human-human or machine-machine ranking correlations. However, there were considerable differences between sign recognizers and between signs: The Q-DFFM recognizer for ZAAG (SAW) and the SDTW recognizer for GORDIJN (CURTAIN) had human-machine ranking correlations that were within the range of the human-human ranking correlations, while none of the sign recognizers showed a clear human-machine ranking correlation for OVEN (OVEN).

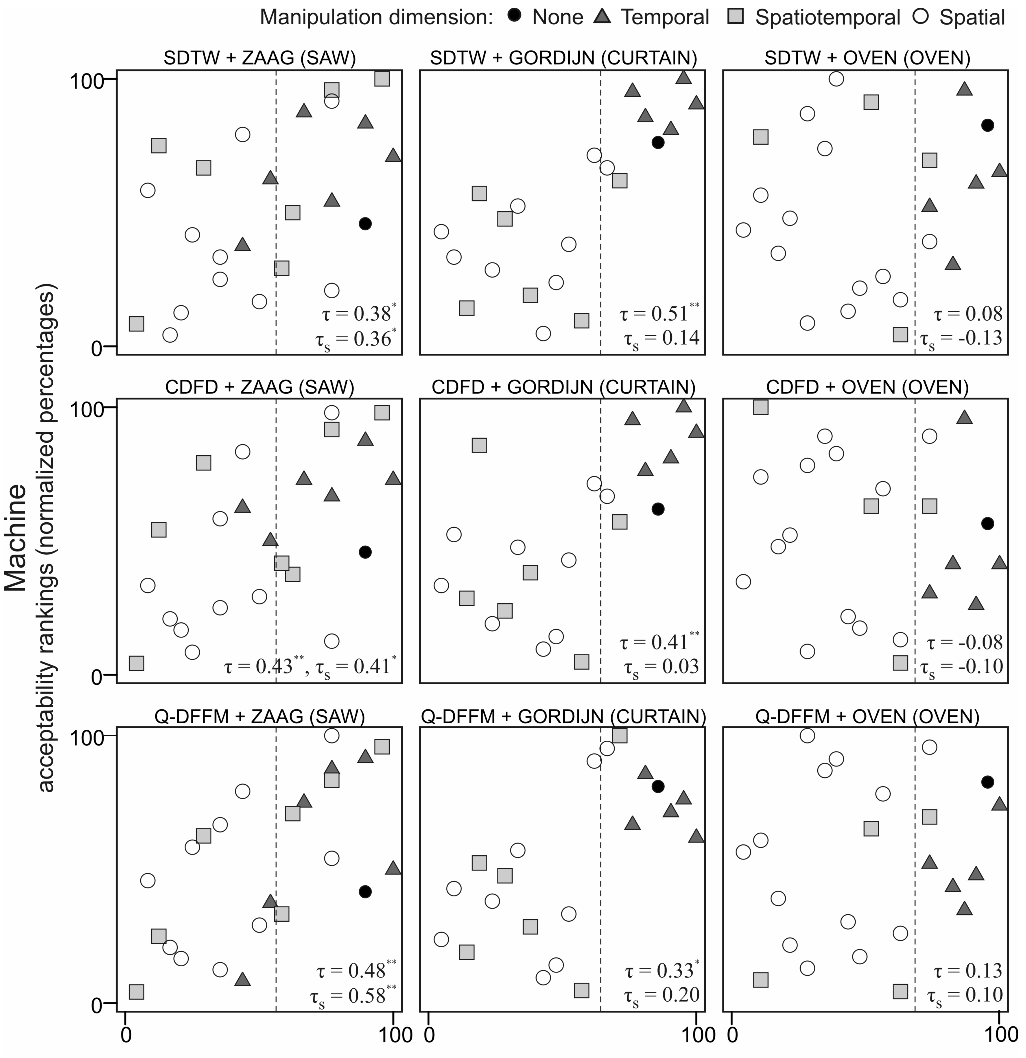


Figure 21. Human-machine ranking scatterplots. For each sign and each sign recognition method the human rankings are given on the X-axis and the machine rankings are given on the Y-axis. These rankings are given as normalized percentages because the number of cases differs between the signs (NSAW = 24, NCURTAIN = 21, NOVEN = 23). The marker style indicates the manipulation dimension. Kendall's ranking correlation coefficient τ is given for each of the ranking pairs in the lower right corner, with τ_s denoting the ranking correlation if only spatial and spatiotemporal manipulations are included. Significance of the correlations is indicated with * ($p < 0.05$) or ** ($p < 0.01$). The vertical dotted lines are the demarcations between movies with an Acceptability which is higher (right) or lower (left) than 50%.

The low human-machine correlations may suggest that the variation present in the training set was not representative for the actual space of acceptable variations. Humans may base their judgment about acceptable variations of a sign on knowledge that cannot be extracted from a set of positive examples with limited variation alone. Humans receive rich input to develop certain sensitivities to categories and margins. Sign productions can contain large variations that are still acceptable while it is equally possible that a sign feature must be produced within very narrow margins (Crasborn, 2001; Van der Kooij, 2002). Humans learn from a combination of highly variable positive examples, targeted negative examples ("no, you tap the side of your nose, not the tip."), and various rules they might be taught explicitly (e.g. a child may receive instruction that there is a certain distinction between handshapes that should be observed). In contrast, the sign recognizers were only trained with a set of positive, acceptable examples of each sign that contained limited variation. The training included neither targeted negative examples, because only the movies from other, non-target signs served as negative examples, nor explicit rules about phonological categories or phonetic implementation. Improving the quality of automatic acceptability ratings from sign recognizers may require that those sign recognizers are trained with different data. Adequate training data may have to include a greater variety of positive examples but perhaps also negative examples in the form of unacceptable variants of signs. Human acceptability judgments, if available, might also be taken into account when training with such examples.

If acquiring a large amount of positive and negative examples (and obtaining human acceptability judgments of them) is too expensive, explicit phonological knowledge, that is, rules about the acceptability of certain values of features or feature combinations, might be formulated in a way that it can be implemented in an automatic sign recognition system. Most existing sign recognition methods do not incorporate much phonological knowledge other than acknowledging in a very general sense that handshape, hand orientation, hand motion and hand location determine the meaning of sign language signs. However, there have been a few attempts to incorporate more phonological knowledge into automatic sign recognition (Derpanis et al., 2004, 2008; Vogler and Metaxas, 1999, 2004). In each of these cases, ASL was the language being recognized and ASL phonological proposals were used (Stokoe, 1960/2005; Liddell & Johnson, 1989). To incorporate phonological knowledge about SLN into the automatic recognition of SLN signs, the proposal by Van der Kooij (2002) could be used.

Another example of knowledge that is lacking in the calculations of sign recognizers is iconicity. Humans may perceive iconicity in signs (see the previous section on phonology and iconicity). Certain variations, which are perhaps small in terms of feature scores, may be judged as unacceptable because they were incompatible with a sign's iconicity. For example, participants commented about why it was acceptable to shift the position of the sign CURTAIN (which is iconic in the sense that it is an enactment of closing the curtains) upwards but not downwards ("you cannot close a curtain well if you grab it low") and why it was acceptable to shift the sign SAW (the hand models a saw blade while a sawing movement is made) downwards but not upwards ("you cannot saw well if you can't lean into it"). For the sign OVEN (OVEN) several variations in the movement path, such as the presence and direction of an arc in the motion, were incompatible with the sign's iconicity (enactment of opening an oven door) and this may have contributed to the poor human- sign recognizer ranking correlations for OVEN (OVEN).

But, how could a sign recognizer deal with iconicity? Edwards (1998) recognized the difficulties that iconic modifications pose to automatic sign recognition and proposed to explicitly try to recognize in parallel (a) which signs were made and (b) how they were made. Edwards (1998) proposed to specify beforehand which features may change in a meaningful way and program this into a feature-based grammar (Gazdar & Mellish, 1989). Several sources could be consulted about such modifications. For ASL, Bellugi and Fischer (1972), Klima and Bellugi (1979), Macken et al (1993) and Emmorey (2002) give overviews of modifications. For

SLN, Zwitserlood (2003) offers a good overview of meaningful modifications of handshapes and she notes that her findings might generalize to some extent to other signed languages.

5.5. Conclusions

The signers participating in our experiment were able to judge the acceptability of a set of sign manipulations in a consistent way: their rankings of the acceptability of sign manipulations correlated well (although they varied in their tolerance). We found that manipulations in the temporal domain, such as increasing or decreasing the speed or tampering with the hold structure, were typically highly acceptable while spatial (and spatiotemporal) manipulations were often judged unacceptable. Examining further subcategories of the manipulations, such as changes in hand orientation or movement direction, revealed that acceptability varied greatly, both between different subcategories of the same domain and between signs.

With regard to acceptability judgments, phonology and iconicity were both found to be useful concepts to take into consideration. Determining the type of phonological error caused by a manipulation may (partially) explain its acceptability. Likewise, considering a sign's iconicity and classifying whether a manipulation is compatible or incompatible with that iconicity can also (partially) explain its acceptability.

Automatic acceptability ratings by three sign recognizers were found, on average, to correlate poorly to human acceptability judgments (although some sign recognition methods performed well for some signs). These human-machine correlations were low in comparison to human-human correlations, possibly due to the recognizers' lack of phonological knowledge and disregard of iconicity.

Appendix 5-A. Participant data

Table 30. Participant characteristics and results. Tolerance is the percentage of accepted sign manipulations in the total amount of judgments (acceptable + acceptable/doubt) / N.

ID	Hearing Status	Parental Hearing Status	Age	Gender	Age of SLN Acquisition	SLN Source of Acquisition	SLN Fluency	SLN Usage	Professional Use of SLN	N	Acceptable	Accept./Doubt	Doubt	Unacc./Doubt	Unacceptabe	Tolerance
1	Deaf (birth)	Hearing	48	Female	0-2yr	SLN courses	fluent	primarily	SLN Teacher	194	136	0	2	0	56	70%
2	Deaf (birth)	Deaf	44	Male	0-2yr	Family	fluent	primarily	Yes, different	262	123	49	14	21	55	66%
3	Deaf (birth)	Hearing	32	Female	3-6yr	SLN courses	reasonable	daily	Yes, different	262	157	1	24	0	80	60%
4	Hearing	Hearing	37	Male	>29yr	SLN courses	reasonable	regularly	Yes, different	262	87	67	32	34	42	59%
5	Hearing	Hearing	27	Female	17-29yr	SLN courses	good	daily	SLN Teacher	262	96	48	27	51	40	55%
6	Hearing	Deaf	46	Female	0-2yr	Family	fluent	daily	SLN Interpreter	262	143	0	20	0	99	55%
7	Hearing	Hearing	24	Female	17-29yr	SLN courses	good	daily	SLN Interpreter	262	104	33	17	30	78	52%
8	Deaf (birth)	Hearing	35	Female	7-16yr	Family	good	daily	Yes, different	262	132	0	3	0	127	50%
9	Hearing	Hearing	38	Female	17-29yr	SLN courses	good	daily	SLN Interpreter	262	92	40	37	31	62	50%
10	Hearing	Hearing	26	Female	17-29yr	SLN courses	fluent	daily	SLN Teacher	262	124	0	35	0	103	47%
11	Hearing	Hearing	24	Female	17-29yr	SLN courses	good	regularly	SLN Teacher	262	57	65	36	23	81	47%
12	Deaf (birth)	Hearing	44	Female	17-29yr	SLN courses	reasonable	regularly	Yes, different	262	81	33	17	47	84	44%
13	Deaf (birth)	Deaf	40	Male	0-2yr	Family	fluent	daily	Yes, different	262	69	42	38	14	99	42%
14	Deaf (birth)	Hearing	22	Female	3-6yr	Family	fluent	daily	Yes, different	262	87	24	10	12	129	42%
15	Hearing	Hearing	32	Female	17-29yr	Otherwise	reasonable	regularly	Yes, different	262	71	39	24	29	99	42%
16	Deaf (birth)	Hearing	37	Female	3-6yr	Otherwise	fluent	primarily	SLN Teacher	262	98	0	16	0	148	37%
17	Hearing	Hearing	24	Female	17-29yr	SLN courses	good	regularly	No	262	83	15	27	18	119	37%
18	Deaf (birth)	Hearing	30	Female	7-16yr	Otherwise	fluent	primarily	SLN Teacher	262	89	0	30	0	143	34%
19	Hearing	Hearing	34	Female	17-29yr	SLN courses	good	regularly	Yes, different	262	55	34	30	44	99	34%
20	Hearing	Hearing	32	Female	17-29yr	SLN courses	fluent	regularly	SLN Interpreter	262	87	0	30	0	145	33%
21	Deaf (birth)	Deaf	40	Female	0-2yr	Family	fluent	daily	No	262	81	0	26	0	155	31%
22	Hearing	Hearing	44	Female	>29yr	SLN courses	good	regularly	No	262	63	0	35	0	164	24%
23	Deaf (birth)	Hearing	33	Male	0-2yr	SLN courses	fluent	primarily	Yes, different	194	40	0	44	0	110	21%
24	Hearing	Hearing	27	Female	3-6yr	Family	good	daily	Yes, different	262	52	0	43	0	167	20%
25	Deaf (birth)	Hearing	35	Female	>29yr	SLN courses	reasonable	regularly	Yes, different	262	46	0	21	0	195	18%
26	Deaf (birth)	Hearing	51	Male	7-16yr	SLN courses	good	daily	SLN Teacher	194	26	0	24	0	144	13%
									Average		88	19	25	14	109	42%
									Average (%)		34%	7%	10%	5%	43%	

Appendix 5-B. Sign specifications

The four test signs are described below by means of their phonological specification, and phonetic implementation and iconicity considerations. The specifications were provided by Els van der Kooij together with Johan Ros (although we also made a few choices in cases where alternatives were offered, and therefore take responsibility for any errors). Regarding iconicity considerations, here we have relied, next to the iconic aspects already present in the sign specifications, on the comments given by several signing informants. Regarding the phonetic implementation considerations, we have tried to find and apply all the relevant rules documented by Van der Kooij (2002), but any errors due to misinterpretation are entirely our own.

GORDIJN (CURTAIN)

Phonological specification

Manner: [symmetrical]
Selected fingers: [one]
Curve: [curve]
Finger configuration: [close]
Setting: [ipsi], [contra]
+ Handle (classifier)

Considerations regarding iconicity and phonetic implementation

- The action of the hands is an imitation or enactment (Müller, 1998) of grasping the two parts of the curtain and closing them by bringing them together.
- The handshape (a ‘money’-hand) is a commonly used handshape, but it does require an extensive specification (selected fingers, curve, finger configuration) (Van der Kooij 2002: 153-5). Of the group of ‘handle’ classifiers it is the ‘knuckle grasp’. The ‘+ Handle (classifier)’ indicates that alternative handling handshapes, such as a palm grasp (heavy grip with a fist with /adducted/ thumb) or a finger pinch may also be appropriate in context (Van der Kooij 2002: 144). In this case we consider the money-hand the phonetic default but the palm grasp is considered free phonetic variation with full iconic compatibility. The precision grip is considered a phonetic implementation error but with contextual iconic compatibility. Handshapes that do not ‘handle’ are iconically incompatible.
- Manner: [symmetrical] requires 2 hands moving in synchrony.
    - Dropping one hand is grammatical in [symmetrical] signs (Van der Kooij 2002: 270). Dropping is not permitted if the two hands are necessary because of an iconic relationship. In this case it is possible that a curtain consists of only one part, in which case it can be closed with just one hand. It is considered a phonetic implementation error with contextual iconic compatibility.
- Setting: [ipsi], [contra] with manner: [symmetrical] means the movement ends in front of the body making contact between the palm sides of the hands (given a default orientation).
- If the palm grasp is used then the thumbs must be /adducted/ because of the contact between the palm sides of the hands (Van der Kooij 2002: 115).
- The setting change can be achieved by moving the hand or by wrist flexion.
- The path movement is straight. This is the default phonological implementation but this may be reinforced by iconicity.

- The hand orientation is straight (by default); neither supine nor prone. This orientation is also required to enable grabbing; Full supination and full pronation (at 90°) are therefore iconically incompatible as well as phonological errors.
- Curtains hang down and are best grabbed high to close them. Grabbing low makes closing difficult or impossible, and a low location is therefore iconically incompatible. A curtain's location to the side (left or right) can be coded in the shifted location of the sign (to a side) and, although this should be reinforced with a glance in that direction, it is considered a modulation.
- The sign can be used as a verb phrase 'close curtain(s)' enabling several modulations:
  - The direction of the setting change can code closing/opening the curtains.
  - Closing and opening the curtain repeatedly can be coded by repeating the movement (setting change).
  - Closing the curtains, but not entirely, can be coded by leaving a certain distance between the hands at the end of the stroke.
- The size of the curtain can be coded in the size of the sign as a modulation.

Oven (oven)

Phonological specification
Selected fingers: [all]
Finger configuration: [close]
Setting: [distal], [proximal] evt. in combinatie met [high].[low]
Location: "virtual object; handle oven"
Relative orientation: [palm]

Considerations regarding iconicity and phonetic implementation

- Semantic motivation (iconicity): the action of the hand is an imitation or enactment (Müller, 1998) of opening the door of an oven by its handle ("virtual object; handle oven") by pulling it forward and down in an arc.
- The handshape depicts the grasping of the handle to enable pulling it (an open hand cannot grasp).
- Setting: [high].[low] implies a horizontal turning axis at the bottom of the door, at the center of an arched movement.
- The main orientation of the hand depicts the orientation of the handle, and, given the turning axis at the bottom of the door, is therefore expected to be horizontal.
- Relative orientation: [palm] is relative to the location "virtual object; handle oven" and can therefore be made with either arm [prone] or [supine].
- Finger configuration: [close] implies a /crossed/ thumb (Van der Kooij, 2000: 113).
- The default location of a one handed sign is near the midsaggital plane, in this case near the center of neutral space (Van der Kooij, 2000: 189).
- The sign can be used as a verb phrase ‘open/close oven’.
    - The direction of the setting change can code closing/opening the oven (in usage as a verb).
- The size of the oven (door) can be coded in the size of the movement
- The location (high, low, to a side) of the oven can be coded in the location of the sign (high, low, to a side), although eye-gaze should also be directed toward the location to confirm locative use.

Stapel (stack)

Phonological specification

Selected fingers: [all]
Relative orientation: [palm]
Location: [hand:broad]
Setting: [low], [high]
Manner: [symmetrical]

Note: Because the handshapes are both with a /B-hand/ this can be considered an 'unbalanced sign with identical handshapes', which Van der Kooij (2002) classifies as an in-between form of symmetry (notated with both features [symmetrical] and [hand:broad]. This is in line with modulations where the handshape of both hands changes (curved or hooked)

Considerations regarding iconicity and phonetic implementation

- Iconicity: At the end of the (stroke of the) sign the hands enclose a certain distance that is the imaginary stack of objects ("hold both ends"). At the begin of the stroke the hands (almost) touch, as if to indicate the point from which the distance is measured, or in other words 'the bottom of the stack' (another surface, such as the tabletop, may substitute for the weak hand and function as 'bottom of the stack' instead). In the classification of Müller (1998) the hands can be said to 'model' the stack, or perhaps more precisely, the hands model the creation of the stack. The setting change [low], [high] confirms the semantic motivation of the sign (the stack grows upward).
- The size of the stack can be coded in the size of the setting change
- The growing of an existing stack can be coded by starting the sign with a certain distance to the weak hand and letting that distance grow.
- The handshape (aperture, curving, fingerselection) can code the nature (size, roundness) of the objects in the stack.
- The location (high, low, to a side) of the curtain can be coded in the location of the sign (high, low, to a side)
- In unbalanced signs (like STAPEL (STACK), as specified by the location [hand:broad]), the default hand arrangement is that the strong hand is on top of the weak hand, as is the case here. Any other hand arrangements (underneath for eaxample) are phonetic implementation errors.
- The default location of the weak hand of an unbalanced sign is in the center of neutral space (Van der Kooij, 2000: 189).
- The handshape of the weak hand is by default a /B-hand/ in unbalanced signs, as it is here in GORDIJN (CURTAIN).
- Weak drop: in unbalanced signs with a /B-hand/ the weak hand can be dropped (Van der Kooij, 2000: 273).
- With selected fingers: [all] the thumb is by default /adducted/ (Van der Kooij, 2000: 115).

Zaag (saw)

Phonological specification

Manner: [repeated]
Selected fingers: [all]
Setting: [proximal],[distal]
Relative orientation: [ulnar]
Location: "virtual object; sawing cut"

Considerations regarding iconicity and phonetic implementation

- Semantic motivation (iconicity): the flat hand is an embodied (see Müller, 1998) representation of the (blade of) the saw while the movement is an enactment of the act of sawing.
    - The secondary orientation of the hand (held vertical, or horizontal) corresponds to the angle of sawing, which can be horizontal for example when sawing a tree. The movement depicts the trajectory of the saw during the sawing action.
    - The main orientation of the hand (ulnar relative to the sawing cut) is therefore strict, since a saw must move in the direction of the blade.
- Location: "virtual object; sawing cut" requires a relative positioning of the body to the movement that would normally allow sawing, in the sense that you have to apply force and/or lean into the movement (the setting change [proximal], [distal]).
- Manner: [repeated] does not imply any fixed number of repetitions (Van der Kooij 2002). A single repetition is considered the default implementation here.
- The manner is considered to be [bidirectional] because there is continuous contact with the location: "virtual object; sawing cut" (Van der Kooij 2002: 249)
    - The endposition, of the post-stroke hold, is [proximal]
- The hand orientation is straight (by default); neither supine nor prone.
- With selected fingers: [all] the thumb is by default /adducted/ (Van der Kooij, 2000: 115).
- The movement forward is, by default, directed towards the center of the neutral space in front of the body
- The sign can be used as a verb phrase ‘saw [object]’.
    - The nature and location of the object can also be coded in the ‘saw’ sign.
    - The angle of sawing can be coded by the angle of the hand, for example a horizontal sawing movement in case of sawing a tree.
    - The effort needed for sawing can be coded in the tenseness and speed of the movement (requires facial expression).

Appendix 5-C. Sign manipulations

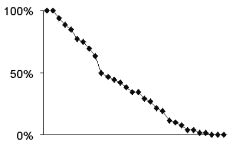
Descriptions of and results for the manipulation of each sign are given in Table 31 to Table 34. The manipulation descriptions (column 1) were also used as instructions during recording of the material and were sometimes more specific than the 'Manipulation' labels (see Table 25, the Manipulation number can be used for cross reference), because in some cases the details of a Manipulation depended on the form of the sign. Moreover, there are some cases where a single Manipulation (e.g. a 45° orientation change) was applied in different directions (e.g. for GORDIJN (CURTAIN): pronated and supinated) leading to two movies with different manipulation descriptions.

The column 'Phonology' shows whether a sign manipulation was considered to be (1) free phonetic variation, (2) a phonetic implementation error, or (3) a phonological error. The column 'Iconicity' shows whether a sign manipulation and the iconicity present in the sign were considered to be (1) compatible, (2) contextually compatible (e.g. modulations), or (3) incompatible.

Results are summarized in each table as the number of times each acceptability rating was given (across participants) to each sign manipulation. Acceptability was calculated by dividing the ratings 'acceptable' plus 'acceptable/doubt' by all ratings. The sign manipulations are ordered by Acceptability, with the highest Acceptability at the top. In the top left corner of each table the Acceptability of the sign manipulations is plotted as it decreases in the order of the table.

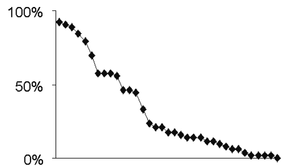
The final three columns show the automatic acceptability ratings as calculated by three different machine algorithms, SDTW, CDFD and QDFF. A high ranking means a high Acceptability. Only those movies that were part of the set that was judged automatically have a ranking in these columns (excluding for example all of the movies for STAPEL (STACK) which is why these columns are not shown for that sign)

Table 31. Sign manipulations for GORDIJN (CURTAIN): descriptions, labels, categories, and results.



GORDIJN Manipulation description (instruction to signer)	Manipulation ID	Manipulation	Manipulation Category	Manipulation Dimension	Phonology	Iconicity	N	Acceptable	Accept./Doubt	Doubt	Unacc./Doubt	Unacceptabe	Acceptability	SDTW	CDFD	QDFF
No post-stroke hold	5	no post-stroke hold	hold structure	temp	1	1	52	48	4	0	0	0	100%	19	19	13
Handshape: closed fist, thumb on side (As)	23	finger selection and configuration	handshape	spa	1	1	52	51	1	0	0	0	100%			
Long post-stroke hold	4	long post-stroke hold	hold structure	temp	1	1	52	48	1	2	0	1	94%	21	21	16
Fast movement	2	fast	speed	temp	1	1	52	42	4	6	0	0	89%	17	17	15
None	0	none	none	none	1	1	52	39	5	4	1	3	85%	16	13	17
Slow movement	1	slow	speed	temp	1	1	52	35	5	6	0	6	77%	18	16	18
Pre-stroke hold	3	pre-stroke hold	hold structure	temp	1	1	52	33	6	7	0	6	75%	20	20	14
Movement: scale up (200%)	7	scale up	scale	S-T	1	2	52	28	8	9	0	7	69%	13	12	21
Location +15 cm higher	19	shift up	shift position	spa	2	1	52	24	9	11	1	7	64%	14	14	20
Movement: Arc outward (+ wrist flexion)	25	arc added to path	movement path	spa	1	1	52	23	3	10	3	13	50%			
Movement without contact (5 cm distance)	31	contact to small distance	contact	spa	2	2	52	20	4	10	4	14	46%	15	15	19
1 Hand	9	symmetry to one hand	symmetry	S-T	2	2	52	14	9	5	5	19	44%	2	1	1
Location +15 cm right	17	shift right	shift position	spa	2	2	52	14	8	6	4	20	42%	8	9	7
Movement: medial-lateral [contra], [ipsi]	28	setting reversal	movement path	spa	3	2	52	16	4	6	6	20	39%	5	3	3
Location +15 cm left	18	shift left	shift position	spa	2	2	52	13	5	9	7	18	35%	1	2	2
Handshape: ring with index and thumb (pinch)	22	finger configuration	handshape	spa	2	2	52	12	6	10	3	21	35%			
1 Hand still	10	symmetry to asymmetry	symmetry	S-T	3	2	52	8	7	4	8	25	29%	4	8	6
Movement without contact (15 cm distance)	32	contact to large distance	contact	spa	2	2	52	12	2	5	8	25	27%	11	10	12
Movement: scale down (50%)	8	scale down	scale	S-T	1	2	52	10	1	10	4	27	21%	10	5	10
Location +15 cm lower	20	shift down	shift position	spa	2	3	52	7	3	5	8	29	19%	6	4	8
Repeated 2 times	13	single to repetition	repetition	S-T	3	2	52	4	2	3	4	39	12%	12	18	11
Hand orientation -45° pronated	33	45° relative orientation change	hand orientation	spa	2	1	52	2	3	5	5	37	10%			
2 Hands parallel/alternating	11	symmetry to alternating	symmetry	S-T	3	3	52	3	1	1	2	45	8%	3	6	4
Movement: Arc inward (+ wrist extension)	25	arc added to path	movement path	spa	2	1	52	1	1	4	3	43	4%			
Hand orientation 45° supinated	33	45° relative orientation change	hand orientation	spa	2	1	52	1	1	1	7	42	4%			
Handshape: closed fist, thumb extended (A)	24	thumb position	handshape	spa	3	3	52	1	0	2	7	42	2%			
Hand orientation + movement +20°	37	path at small angle	movement direction	spa	2	3	52	1	0	3	2	46	2%	7	11	9
Hand orientation -90° prone	34	90° relative orientation change	hand orientation	spa	3	3	52	0	0	1	2	49	0%			
Hand orientation 90° supine	34	90° relative orientation change	hand orientation	spa	3	3	52	0	0	0	5	47	0%			
Hand orientation + movement +45°	38	path at large angle	movement direction	spa	3	3	52	0	0	2	0	50	0%	9	7	5

Table 32. Sign manipulations for OVEN (OVEN): descriptions, categories, and results.



OVEN Manipulation description (instruction to signer)	Manipulation ID	Manipulation	Manipulation Category	Manipulation Dimension	Phonology	Iconicity	N	Acceptable	Accept./Doubt	Doubt	Unacc./Doubt	Unacceptabe	Acceptability	SDTW	CDFD	QDFF
Long post-stroke hold	4	long post-stroke hold	hold structure	temp	1	1	52	41	7	2	1	1	92%	15	10	17
None	0	none	none	none	1	1	52	43	4	4	0	1	90%	19	13	19
Handshape: Thumb adducted (As)	24	thumb position	handshape	spa	2	1	52	43	3	3	0	3	89%			
No post-stroke hold	5	no post-stroke hold	hold structure	temp	1	1	52	42	2	5	0	3	85%	14	6	11
Fast movement	2	fast	speed	temp	1	1	52	32	9	6	0	5	79%	22	22	8
Pre-stroke hold	3	pre-stroke hold	hold structure	temp	1	1	52	22	14	10	0	6	69%	7	10	10
Slow movement	1	slow	speed	temp	1	1	52	18	12	9	2	11	58%	12	7	12
Movement: scale up (200%)	7	scale up	scale	S-T	1	2	52	22	8	16	1	5	58%	16	15	16
Location +15 cm left	18	shift left	shift position	spa	2	2	52	24	6	9	2	11	58%	9	21	22
Handshape: Thumb inside fist E	24	thumb position	handshape	spa	2	1	52	25	4	2	3	18	56%			
2 Hands (symmetrical)	12	one hand to symmetry	symmetry	S-T	3	2	52	22	2	9	4	15	46%	1	1	1
Location +15 cm higher	19	shift up	shift position	spa	2	2	52	16	8	13	2	13	46%	4	3	6
Movement: straight, no arc	26	straight path not arc	movement path	spa	1	3	52	16	7	12	4	13	44%	6	16	18
Movement: scale down (50%)	8	scale down	scale	S-T	1	2	52	10	7	6	9	20	33%	21	15	15
Location +15 cm lower	20	shift down	shift position	spa	2	2	52	10	2	6	5	29	23%	5	4	4
Location +15 cm right	17	shift right	shift position	spa	2	2	52	8	3	15	4	22	21%	3	5	7
Finger configuration: hooked (not closed)	22	finger configuration	handshape	spa	3	1	52	5	6	2	5	34	21%			
Finger configuration: half open (not closed)	22	finger configuration	handshape	spa	3	1	52	4	5	9	4	30	17%			
Hand orientation 45° supinated	33	45° relative orientation change	hand orientation	spa	3	3	52	7	2	11	2	30	17%			
Hand orientation + movement -20° (outside)	37	path at small angle	movement direction	spa	2	1	52	4	4	4	7	33	15%	23	19	21
Handshape: Thumb extended (A)	24	thumb position	handshape	spa	3	3	52	4	3	2	10	33	14%			
Hand orientation -90° (pronated)	36	180° relative orientation change	hand orientation	spa	1	1	52	3	4	2	7	36	14%			
Hand orientation + movement +45° (inside)	38	path at large angle	movement direction	spa	2	1	52	3	4	7	3	35	14%	17	21	20
Movement: reversed direction	28	setting reversal	movement path	spa	3	2	52	5	1	9	7	30	12%	2	2	3
Hand orientation + movement +20° (inside)	37	path at small angle	movement direction	spa	2	1	52	3	3	4	3	39	12%	20	18	23
Movement: straight towards body	29	setting changes reduced	movement path	spa	3	3	52	3	2	2	3	42	10%	11	12	5
Hand orientation + movement -45° (outside)	38	path at large angle	movement direction	spa	2	1	52	3	1	1	3	44	8%	8	11	9
Hand orientation 0°	34	90° relative orientation change	hand orientation	spa	3	3	52	2	1	1	3	45	6%			
Hand orientation -45° (pronated)	35	135° relative orientation change	hand orientation	spa	3	3	52	2	1	3	6	40	6%			
Handshape: Selected fingers = 1 (closed ring)	21	finger selection	handshape	spa	3	1	52	0	2	9	7	34	4%			
Repeated 2 times	13	single to repetition	repetition	S-T	3	2	52	0	1	3	2	46	2%	18	23	2
Finger configuration: curved (not closed)	22	finger configuration	handshape	spa	3	1	52	0	1	9	6	36	2%			
Finger configuration: flat open	22	finger configuration	handshape	spa	3	3	52	1	0	0	1	50	2%			
Movement: straight down	29	setting changes reduced	movement path	spa	3	3	52	1	0	2	1	48	2%	13	17	14
Movement: arc in outward direction	27	arc in wrong direction	movement path	spa	1	3	52	0	0	0	2	50	0%	10	8	13

Table 33. Sign manipulations for STAPEL (STACK): descriptions, categories, and results.

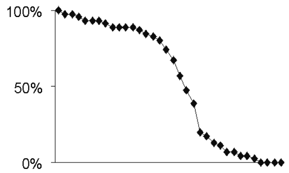
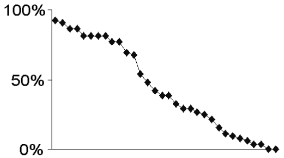
STAPEL Manipulation description (instruction to signer) 	**Manipulation ID**	**Manipulation**	**Manipulation Category**	**Manipulation Dimension**	**Phonology**	**Iconicity**	**N**	**Acceptable**	**Accept./Doubt**	**Doubt**	**Unacc./Doubt**	**Unacceptabe**	**Acceptability**
None	0	none	none	none	1	1	46	46	0	0	0	0	100%
Fast movement	2	fast	speed	temp	1	1	46	43	2	1	0	0	98%
Handshape: Thumb extended (B0)	24	thumb position	handshape	spa	3	1	46	43	2	0	0	1	98%
Long post-stroke hold	4	long post-stroke hold	hold structure	temp	1	1	46	42	2	2	0	0	96%
Movement: scale up (200%)	7	scale up	scale	S-T	1	2	46	40	3	0	0	3	94%
Location +15 cm lower	20	shift down	shift position	spa	2	2	46	39	4	1	0	2	94%
Switch weak-strong hand (left-right)	40	shift dominant hand	other (spatial)	spa	1	1	46	42	1	3	0	0	94%
No post-stroke hold	5	no post-stroke hold	hold structure	temp	1	1	46	38	4	2	0	2	91%
Pre-stroke hold	3	pre-stroke hold	hold structure	temp	1	1	46	40	1	3	0	2	89%
Movement: scale down (50%)	8	scale down	scale	S-T	1	2	46	34	7	2	1	2	89%
Location +15 cm right	17	shift right	shift position	spa	2	2	46	36	5	2	2	1	89%
Location +15 cm higher	19	shift up	shift position	spa	2	2	46	35	6	2	1	2	89%
Slow movement	1	slow	speed	temp	1	1	46	37	3	5	0	1	87%
Movement low-high without contact (5 cm distance)	31	contact to small distance	contact	spa	2	1	46	38	1	3	2	2	85%
Hand orientation -45° (fing-ori)	33	45° relative orientation change	hand orientation	spa	2	1	46	34	4	1	1	6	83%
Location +15 cm left	18	shift left	shift position	spa	2	2	46	32	5	2	1	6	80%
Handshape: Thumb crossed in palm (B)	24	thumb position	handshape	spa	2	1	46	31	3	7	1	4	74%
Hand orientation +45° (fing-ori)	33	45° relative orientation change	hand orientation	spa	2	1	46	28	3	9	0	6	67%
2 Handed symmetrical movement	12	one hand to symmetry	symmetry	S-T	3	1	46	19	7	8	2	10	57%
Movement low-high without contact (15 cm distance)	32	contact to large distance	contact	spa	2	2	46	14	8	7	5	12	48%
1 Hand (removal of weak hand)	39	location hand:palm to neutral space	other (spatial)	spa	2	1	46	13	5	12	2	14	39%
Finger configuration: hooked (bent) (MCP)	22	finger configuration	handshape	spa	3	1	46	7	2	7	5	25	20%
Finger configuration: clawed (klauw)	22	finger configuration	handshape	spa	3	1	46	6	2	5	1	32	17%
Hand orientation + movement +20° (palm-ori)	37	path at small angle	movement direction	spa	2	3	46	1	5	5	6	29	13%
Finger configuration: curved (all joints)	22	finger configuration	handshape	spa	3	2	46	4	1	6	4	31	11%
Handshape: 2 fingers selected	21	finger selection	handshape	spa	3	2	46	0	3	3	1	39	7%
Hand orientation + movement -20° (palm-ori)	37	path at small angle	movement direction	spa	2	3	46	3	0	4	8	31	7%
Repeated 2 times	13	single to repetition	repetition	S-T	3	1	46	2	0	5	5	34	4%
Movement: high-low	28	setting reversal	movement path	spa	3	3	46	2	0	0	1	43	4%
Movement: low-high with arc to side	25	arc added to path	movement path	spa	3	3	46	0	1	2	6	37	2%
Finger configuration: closed fist	22	finger configuration	handshape	spa	3	1	46	0	0	0	2	44	0%
Hand orientation + movement +45° (palm-ori)	38	path at large angle	movement direction	spa	3	3	46	0	0	0	1	45	0%
Hand orientation + movement -45° (palm-ori)	38	path at large angle	movement direction	spa	3	3	46	0	0	0	1	45	0%
Switch weak-strong hand (lower hand moves down)	41	upside down	other (spatial)	spa	3	3	46	0	0	1	6	39	0%

Table 34. Sign manipulations for ZAAG (SAW): descriptions, categories, and results.



ZAAG Manipulation description (instruction to signer)	**Manipulation ID**	**Manipulation**	**Manipulation Category**	**Manipulation Dimension**	**Phonology**	**Iconicity**	**N**	**Acceptable**	**Accept./Doubt**	**Doubt**	**Unacc./Doubt**	**Unacceptabe**	**Acceptability**	**SDTW**	**CDFD**	**QDFF**
Monodirectional accent	6	monodirectional accent	other (temp)	temp	3	1	52	46	2	3	0	1	92%	17	18	12
Repeated 3 times	15	repetition exaggerated	repetition	S-T	1	1	52	43	4	3	0	2	90%	24	24	23
None	0	none	none	none	1	1	52	42	3	7	0	0	87%	11	11	10
Long post-stroke hold	4	long post-stroke hold	hold structure	temp	1	1	52	40	5	6	0	1	87%	20	21	22
No post-stroke hold	5	no post-stroke hold	hold structure	temp	1	1	52	37	5	8	0	2	81%	13	16	21
Repeated 4 times	15	repetition exaggerated	repetition	S-T	1	1	52	36	6	5	2	3	81%	23	22	20
Location +15 cm lower	20	shift down	shift position	spa	2	2	52	30	12	8	0	2	81%	5	3	13
Movement: forward and down [high/low + proximal/distal]	30	setting changes added	movement path	spa	3	2	52	39	3	8	1	1	81%	22	24	24
Fast movement	2	fast	speed	temp	1	1	52	35	5	8	0	4	77%	21	18	18
Handshape: Thumb crosses palm	24	thumb position	handshape	spa	2	3	52	34	6	6	0	6	77%			
Movement: scale down (50%)	8	scale down	scale	S-T	1	2	52	27	9	10	0	6	69%	12	9	17
Handshape: Thumb extended	24	thumb position	handshape	spa	3	1	52	26	9	7	3	7	67%			
Movement: scale up (200%)	7	scale up	scale	S-T	1	2	52	20	8	7	3	14	54%	7	10	8
Pre-stroke hold	3	pre-stroke hold	hold structure	temp	1	1	52	13	12	12	3	12	48%	15	12	9
Location +15 cm right	17	shift right	shift position	spa	2	2	52	17	5	8	6	16	42%	4	7	7
Slow movement	1	slow	speed	temp	1	1	52	11	9	16	2	14	39%	9	15	2
Movement: forward and up [low/high + proximal/distal]	30	setting changes added	movement path	spa	3	2	52	11	9	8	3	21	39%	19	20	19
Hand orientation 45° supinated	33	45° relative orientation change	hand orientation	spa	2	2	52	9	8	5	8	22	33%			
Hand orientation + movement +20° (inside)	37	path at small angle	movement direction	spa	2	1	52	7	8	10	6	21	29%	6	6	3
Hand orientation + movement -20° (outside)	37	path at small angle	movement direction	spa	2	2	52	7	8	5	4	28	29%	8	14	16
Not repeated	14	repetition dropped	repetition	S-T	3	3	52	10	4	10	6	22	27%	16	19	15
Location +15 cm higher	19	shift up	shift position	spa	2	3	52	8	5	6	4	29	25%	10	2	14
Location +15 cm left	18	shift left	shift position	spa	2	2	52	6	5	6	4	31	21%	3	4	4
Hand orientation 90° supinated	34	90° relative orientation change	hand orientation	spa	3	2	52	5	3	2	1	41	15%			
Hand orientation + movement +45° (inside)	38	path at large angle	movement direction	spa	3	2	52	5	1	3	2	41	12%	1	5	5
Handshape: 2 fingers selected	21	finger selection	handshape	spa	3	2	52	4	1	4	3	40	10%			
Not repeated (half movement)	16	unfinished movement	other (sp-temp)	S-T	3	3	52	2	2	1	2	45	8%	18	13	6
Hand orientation + movement -45° (outside)	38	path at large angle	movement direction	spa	3	3	52	2	1	14	5	30	6%	14	8	11
Hand orientation -45° (pronated)	33	45° relative orientation change	hand orientation	spa	2	3	52	2	0	2	1	47	4%			
Hand orientation -90° (pronated)	34	90° relative orientation change	hand orientation	spa	3	3	52	2	0	0	0	50	4%			
2 Hands (symmetrical)	12	one hand to symmetry	symmetry	S-T	3	3	52	0	0	0	0	52	0%	2	1	1
Fingers spread extended	22	finger configuration	handshape	spa	3	3	52	0	0	0	2	50	0%			

Chapter 6

•

General discussion

In the previous chapters the results of a series of studies on the appearance of manual movements in gestures have been reported and discussed. In this final chapter some of these findings will be combined and several issues will be highlighted that may be relevant for our current understanding of how people perceive and interpret manual movements. In addition, a shortlist is provided with those insights that might be useful to inspire further developments in automatic gesture and sign language recognition. Finally, some ideas for further research are presented.

Looking for an intention to communicate – The current research provides empirical data on people's ability to discriminate signs (and other gestures) from human behaviour that is not intended to communicate, such as fidgeting. Kendon (2004) has previously cited anecdotal evidence that people appear to be able to see that a movement is intended to communicate, or in his words 'has an appearance of deliberate expression', even if they do not know what the exact meaning of the movement is. A first affirmation of such 'gesture detection' abilities was found when non-signers turned out to be just as capable as signers in detecting signs in material that contained both signs and fidgets (chapter 2). A second affirmation came when non-signers were again able to discriminate between signs and fidgets and when, in addition, emblems (which were not commonly known) were not treated differently than signs (chapter 3). These findings make it clear that non-signers, in both experiments, were not really able to see that some hand movement certainly *was* a lexical sign, but rather that it *could be* a sign, probably because they could see that it appeared to be intended to communicate. In chapters 2 and 3 we made some suggestions about how people were able to detect gestures, but this remains a topic for further research, which is discussed below.

A little bit of a sign is enough - We have measured how much time people require to respond to a sign as they watch its movement unfold, under fairly normal viewing conditions. This was done twice. In chapter 2 we asked people to respond as soon as they saw the beginning of a sign (sign detection) and in chapter 4 we asked them to respond as soon as they recognized the meaning of a sign (lexical recognition). In both cases we found that people were able to respond quickly and certainly did not have to wait until the sign had ended. From the response times we estimated at what moment the information necessary to respond had become available in the signal by subtracting the time people needed to provide a motor response to a visual event. In both cases the information was estimated to become available shortly after the onset of the stroke or the nucleus of the sign (we coded the movement phases of the signs) and both response times were highly correlated to the stroke onset. By using the same movies for both experiments we were able to compare the times required for sign detection and lexical recognition. These times were highly correlated and for lexical recognition people required about 90 ms more time than for sign detection. The lexical recognition response times, and our estimation of when the necessary information for lexical recognition became available, were in line with previous findings of two 'gating studies' (Grosjean, 1981; Emmorey and Corina, 1990) in which the viewing conditions are arguably less normal, see chapter 4. In these studies people were able to recognize the lexical meaning of signs using only a small part of the sign. Our findings show that the necessary information for lexical recognition is not only quickly present in the signal but it is also quickly processed and available under normal viewing conditions.

A robust temporal structure - Several studies in this dissertation contained aspects that gave us more insight into (a) how people retrieve the temporal structure (the sequence of relevant events) of movement and into (b) what the boundaries are between (phases of) movements. Overall, with our material, people appeared to have little difficulty with processing the temporal structure of hand movements. We found in chapter 5 that temporal sign manipulations, e.g. the speed of signs or their hold structure, had very little impact on their acceptability. In addition, we found in chapters 2 and 4 that the only effect of trying to camouflage the beginning of a sign with a preceding fidgeting movement at the same location was that it appeared to facilitate rather than hamper the perception of a sign. From these findings, it appears that retrieving the temporal structure of hand movements is a resilient process that is not easily disrupted by variation in the signal.

Work in various fields points to a fairly straightforward, coherent explanation of how humans are able to retrieve the temporal structure of hand movements in such a robust way. In the field of sign language phonology, movement is usually treated primarily as a transition between a beginning and an end state (Hayes, 1993; Uyechi, 1996; Van der Kooij, 2002). This is in line with theory in the field of visual perception: the state-motion-state (SMS) moving shape representation as proposed by Marr and Vaina (1982). In the case of sign language the 'moving shape' is the 'articulator' of the sign (Crasborn, 2001), which is usually the hand but which can also be a part of the hand (or even the whole arm). 'States' are moments in which (parts of) the shape are either absolutely or relatively at rest. Rubin and Richards (1985) created a more operational definition for such 'states' in the form of 'visible motion boundaries' which they defined as starts, stops and force discontinuities. According to Rubin and Richards (1985) visible motion boundaries, defined in this way, are fairly robust against variation. We speculated in chapter 2 that the camouflage failed because the movement boundaries, in the sense of Rubin & Richards (1985), remained clearly visible in the movies. Likewise, in chapter 5, the manipulations of the speed and hold structure of signs may not have threatened the visibility of a motion boundary.

The relation we found between the onset of the stroke and response times for sign detection (chapter 2) and lexical recognition (chapter 4) is compatible with the previous suggestion that our findings can be explained by assuming a robust temporal segmentation through visible motion boundaries. Newquist (1976) showed that motion boundaries, or 'breakpoints' as he called them, were the most perceptually salient parts of human behaviour. In our discussion of the findings we have also noted that the onset of the stroke appears to be a moment where much information becomes visible in the signal which can be used for sign detection and lexical recognition, and this is in line with other findings (Grosjean, 1981; Emmorey and Corina, 1990, Ten Holt et al., 2009a).

Dealing with variation –We found in chapter 5 that manipulations of spatial features of signs were often judged unacceptable but, moreover, that this acceptability was quite variable and depended on the signs. The study on phonology and iconicity showed that both ways of looking at variation in signs explained a significant part of the variability in acceptability judgments. If the form of a sign is not made within the boundaries of the expected phonological categories then this leads to judgments of 'unacceptable'. At the same time, signers appeared also to pay much attention to iconic interpretations. If a sign was made in such a way that it was incompatible with the sign's iconicity it was very often judged unacceptable, regardless of whether it was a phonological error or not. In contrast, if a sign was made in such a way that it could be interpreted as an iconic modification it was often judged as acceptable, especially by native signers.

This is in line with the 'dual nature' that many writers have ascribed to sign language signs and also to other gestures (Tervoort, 1953; Klima & Bellugi, 1979; Macken et al., 1993; Armstrong et al., 1995; Pietrandrea, 2002; Kendon, 2004). Signs as well as the emblematic gestures that we have used in our studies can all be regarded as semiotic 'signs' in the sense that they have a form (the signifier) that denotes some meaning (the signified) (Nöth, 1990), but they are dualistic in the way they signify. On the one hand, lexical signs as well as emblems are highly conventionalized and some signs and emblems may even be considered entirely arbitrary symbols (leaving aside whether they were arbitrary at the time of their formation). As in any collection of symbols that together form a semiotic system, lexical (e.g., SLN or ASL) signs need to have distinctive characteristics so as not to be confused with each other. This leads to a system of contrasts and, for a lexicon, to phonological categories. On the other hand, many signs, as well as many other gestures, are iconic too, in the sense that people perceive them as sharing some property with the objects or actions they denote.

Although there is general agreement that iconicity is an important element in the *formation* of signs, there are different views about the role of iconicity in the actual *usage* of signed languages. Cuxac (1999, 2000) is one of the few to ascribe a prominent role to iconicity in the usage of signed languages and he proposes that iconicity plays a central role in many aspects of French Sign Language. Many other authors, who are mostly concerned with ASL or British Sign Language, suggest a very limited role for iconicity. Klima and Bellugi (1979) showed that iconicity does not mean that the meaning of signs can be easily guessed, or, in their words, most signs are not 'transparent'. They also sketched how many signs appear to lose some of their iconicity as they mature, for example because their form is slightly changed to conform to the phonological categories of a language or because the action they denote is no longer a part of normal, everyday culture. Some findings even seem to show that iconicity does not play an important or active role in the perception of iconic signs (Siple et al., 1982; Newport & Meier, 1985) or in their production (Marshall et al., 2004; Emmorey et al., 2004) at all.

The current findings are an indication that, at least when signers have to make acceptability judgments of highly variable sign manipulations, iconicity does play a role. A possible explanation might be that people only rely on iconicity in signs when needed. If one regards iconicity as a fairly complex strategy to interpret the meaning of signs it might then be reasonable to expect that iconicity only plays a role if a straightforward symbolic interpretation is not possible or does not suffice, for example if one has to judge the acceptability of a sign manipulation or if one sees an unknown sign or gesture. Perhaps in the studies by Siple et al. (1982), Newport & Meier (1985), Marshall et al. (2004) and Emmorey et al. (2004) people did not have to rely on iconicity even though they were dealing with iconic signs. There are several indications that seeing iconicity is indeed a fairly complex matter and not something that necessarily occurs whenever you look at an iconic sign. For example, Eco (1976) provided a solid argument, based on many observations and examples, that people must first learn cultural conventions before they are able to see shared properties in iconic signs (any semiotic sign that relies on iconicity) because they do not actually share properties. Another example, with ASL signs, is the finding by Tolar et al. (2008) who reported, in line with Eco's view, that the ability to see the iconicity of iconic signs is not present in very young children but gradually develops during the preschool years.

Intelligent automatic gesture and sign language recognition – If one wishes to develop intelligent automatic gesture recognition, intelligent in the sense of mimicking human capabilities (Turing, 1950), then several of our findings might pose challenges that need to be overcome. Here is a list of some capabilities that computing machines may still have to develop before people will call them intelligent gesture watchers:

- Correctly retrieve the temporal structure of a sequence of movements and separate the movements. Allow restpositions at variable locations.
- Discriminate between movements that are intended to communicate (attend these) and other movements. Ignore fidgeting.
- Respond as fast as humans, preferably well before the gesture has ended.
- Allow and disallow variation according to the distinctive categories and rules, if any, of the gesture system (in case of lexical signs this is the phonology).

- Be able to interpret a sign or gesture production as a combination of (a) some gesture from a repertoire and (b) a meaningful (iconic) modification of that gesture.

This list could be longer but it is restricted here to issues related to the current findings. Humans are, for example, also able to interpret gestures they have never seen before through iconic strategies (Müller, 1998; Streeck, 2008).

Several authors in the field of automatic gesture and sign language recognition have already made remarks about the challenges listed above, suggested ideas or even reported some progress in those areas (Edwards, 1988; Parish et al., 1990; Brashear et al., 2006; Kim et al., 2007; Roh et al., 2008; Junker et al., 2008; Holt et al., 2009b). However, it is outside the scope of this chapter to discuss these contributions here.

6.1. Further research

Seeing an intention to communicate - One of the burning questions that have remained unanswered by this research, in the field of gesture studies, is which visible characteristics of an action make people see that the action is intended to communicate, which is what we have used as a definition of gesture throughout this dissertation. We have offered some suggestions in the discussions of chapters 2 and 3, and Kendon (2004) also offers some suggestions, but it will take more extensive and careful research to find the answer to this question. Part of the answer probably lies in people's experience. It appears reasonable to expect that people are often able to perceive the intention of action because they are familiar with the action (they perform the action themselves and/or have seen other people perform it). This might be the easiest explanation of why people were able to see that fidgeting movements were not intended to communicate: they knew that these movements (or movements that are very similar) were not addressed to other people. Yet, such experience does not offer a very satisfactory explanation of how people can observe a sign or an emblem they have never seen before and, despite not knowing what it means, can still tell that it is intended to communicate. One could argue that if people see an unknown action they can not associate with some probable intention, they simply assume that this action must be intended to communicate, as a kind of default interpretation. However, it may also be possible that those actions which are seen by people as intended to communicate also share certain visible characteristics and that this may even be, to some degree, a universal phenomenon. This does not have to imply that there is a single set of visible 'gesture' characteristics and that one can determine the 'gesturalness' of a movement by checking whether it has those characteristics, but perhaps various combinations of visible characteristics exist which give a movement a 'gestural' appearance. Context undoubtedly will play a role in these perceptions. For example, Gerwing and Bavelas (2004) demonstrated that the form of gestures tends to change (to a less pronounced form) with repeated use in a conversation. Yet it seems unlikely that context alone suffices to explain the ease with which people are able to discriminate between movements that are intended to communicate and those that are not, especially given that people were able to do so in our experiments where few, if any, contextual clues were given that could help them in this respect.

The question how people can see that an action is intended to communicate is tied to the more general question of how people are able to 'see' any intentions behind other people's actions, a topic which is receiving a growing amount of scientific attention, especially since the discovery of 'mirror neurons' led to a proposed neural mechanism for this capability, namely that we can use our own motor programs to interpret the actions we observe in others (e.g., Rizzolatti & Arbib, 1998; Baldwin and Baird, 2001; Kelly et al., 2007). Perhaps an interesting way to study people's capability of seeing that a movement is intended to communicate is by having people participate who have been diagnosed as suffering from delusions of communication, a group that has been found by Bucci et al. (2008) to often perceive 'incidental movements', without communicative intention, as meaningful gestures. If one could recruit patients suffering from such delusions to various degrees then they could be shown movies of human behaviour that include gestures, fidgeting, practical actions and so forth. If they are instructed to 'press the spacebar as soon as you see a movement that is intended to communicate' we could observe which actions are most often seen as intended to communicate. One might also observe whether such participants share an ordinal scale of movements that are least and most likely to be seen as intended to communicate and only differ in their sensitivity or their threshold or if different participants single out different incidental movements as gestures.

Other or additional material – The current findings have been the result of experiments with carefully constructed material. In each case a signer was asked to produce signs, emblematic gestures or fidgeting in a particular way and under favorable recording conditions (e.g. good lighting conditions and good color contrasts between skin and clothing). Doing the experiments again with the same procedures but with changes in the material may already provide additional insights:

- People's ability to discriminate movements that are intended to communicate, as found in chapters 2 and 3, could be further analyzed by offering them an even more mixed set of actions to observe. It would be interesting to include other sorts of gestures besides signs and emblematic gestures. Perhaps gestures that are used to regulate the flow of the communication, such as 'beats' in the sense of McNeill (1992), will not be so easily discriminated from fidgeting.
- To study effects of context on the perception of an intention to communicate one could perform experiments like those in chapters 2 and 3, but use recordings of gestures or signs in actual conversations by unsuspecting people, instead of instructing and recording a signer in a studio. The conversation itself as well as other factors, could then serve as a context.
- Context may also have an effect on how fast people can recognize the meaning of a sign (Clark & Grosjean, 1982). If experiments like those in chapters 2 and 4 were repeated with signs in a conversational context, the various effects of context on response times could be studied.
- The finding, in chapter 5, that native signers found sign manipulations with contextual iconic compatibility highly acceptable, rests on the acceptability judgments of only four native signers and sign manipulations of only four signs. A replication of the experiment that focuses on iconic compatibility would benefit from a larger number of native signers as participants and from more signs, including also signs that are not iconic by themselves (but which can nevertheless be modified using iconic strategies, see Macken et al. (1993)).

Multimodal HCI – Automatic gesture and sign language recognition can be applied as a means of HCI that is independent from other HCI means, but it might also contribute to the overall HCI performance if gesture recognition is integrated with other HCI means. Manual gestures could be integrated with speech, facial expressions, mouth gestures, body postures, gaze direction, etc. Such an integration of modalities into multimodal HCI is not easy and has its own problems (e.g. Jaimes & Sebe, 2007), but it may be helpful, and in some cases even necessary, to interpret gestures correctly. Speech or facial expressions may provide the clues that are necessary to be able to interpret a gesture as a combination of some gesture from a repertoire and a meaningful modification of that gesture. For example, in signed languages, modifications regarding the size of an object (e.g. a big ball versus a small ball) can be indicated by inflating (big) or deflating (small) the cheeks whilst also expanding or decreasing the size of the manual gesture (e.g. the hands model or hold a big or a small ball in front of the body). Another example in SLN is that directing the gaze to a particular location while also performing a sign in or towards that location, is an indication of 'locative use', meaning that the object and its location are both specified by the sign. Conversely, the sign made in that location is of course also helpful to interpret the gaze towards that location.

We received some indication, in chapter 5, of the importance of context to interpret signs correctly, when we observed that modifications to signs that were compatible with the sign's iconicity were judged to be quite acceptable, although in this task we were only dealing with isolated, though highly variable productions of lexical signs. It is likely the role of context in interpretation gets larger as the utterances get larger, such as signed sentences or entire turns in a discourse. For larger utterances, the information in one modality may serve as the context in which to interpret other elements. If someone changes his body posture or repositions his head during talking and gesturing this may serve as a discourse marker or as an indication that he is switching narrative perspective. Such cues can be very important to correctly interpret speech and gestures from a speaker or to interpret sign language (e.g. Emmorey et al., 2000). Conversely, if someone makes certain gestures during speaking this can indicate how his speech should be interpreted (e.g. Kendon, 1995).

The way in which humans can combine speech, gestures and other communicative behaviour offers us very powerful strategies to communicate meaning. Not being able to use that power may well limit our appreciation of the communication. Therefore, to get a truly satisfactory user experience out of automatic gesture or sign recognition (and the same may hold for speech recognition) it may prove to be necessary to combine the various technologies to track and interpret human behaviour into multimodal interfaces.

References

Andrea, P., & H. P. de Boer. (1993). *Het gebarenboekje* (Second ed.). Baarn: De Fontein.

Arendsen, J., J. F. Lichtenauer, A. J. van Doorn, E. A. Hendriks, G. Spaai, M. J. T. Reinders, & H. de Ridder (2008). Acceptability Ratings by Humans and Automatic Gesture Recognition for Variations in Sign Productions. *Proceedings of the 8th IEEE International Conference on Automatic Face and Gesture Recognition*, 6-11.

Arendsen, J., A. J. van Doorn, & H. de Ridder. (2007). When and how well do people see the onset of gestures? *Gesture, 7*(3), 305-342.

Armstrong, D. F., W. C. Stokoe, & S. E. Wilcox. (1995). *Gesture and the nature of language.* Cambridge: Cambridge University Press.

Baldwin, D. A., & J. A. Baird. (2001). Discerning intentions in dynamic human action. *Trends in Cognitive Sciences, 5*(4), 171-178.

Bellugi, U., & S. Fischer. (1972). A comparison of sign language and spoken language. *Cognition, 1*(2/3), 173-200.

Bernstein, J., M. Cohen, H. Murveit, D. Rtischev, & M. Weintraub. (1990). Automatic evaluation and training in English pronunciation. *Proceedings International Congress on Spoken Language Processing (ICSLP) '90*, Kobe, 1185-1188.

Bowden, R., D. Windridge, T. Kadir, A. Zisserman & M. Brady. (2004). A linguistic feature vector for the visual interpretation of sign language. *Lecture Notes in Computer Science*, 3021, 390-401.

Brashear, H., V. Henderson, K. H. Park, H. Hamilton, S. Lee, & T. Starner. (2006). American sign language recognition in game development for deaf children. *Paper presented at the Eighth International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS 2006.*

Bucci, S., M. Startup, P. Wynn, A. Baker, & T. J. Lewin. (2008). Referential delusions of communication and interpretations of gestures. *Psychiatry Research, 158*(1), 27-34.

Clahsen, H., & C. Felser. (2006). How native-like is non-native language processing? *Trends in Cognitive Sciences, 10*(12), 564-570.

Clark, L. E., & F. Grosjean (1982). Sign recognition processes in american sign language - the effect of context. *Language and Speech, 25*, 325-340.

Conlin, K. E., G. Mirus, C. Mauk, & R. P. Meier. (2000). Acquisition of first signs: Place, handshape, and movement. In C. Chamberlain, J. P. Morford & R. I. Mayberry (Eds.), *The acquisition of linguistic representation by eye* (pp. 51-69). Mahwah, NJ: Erlbaum.

Corina, D. P., & H. Knapp (2006). Sign language processing and the mirror neuron system. *Cortex, 42*(4), 529-539.

Cook, P. S. (1998). *United States of ASL poetry and other tales* (VHS-video; English). Chicago, IL: P.C. Productions.

Crasborn, O., H. van der Hulst, & E. van der Kooij. (2002). SignPhon: a phonological database for sign languages. Sign Language & Linguistics, 4(1/2), 215-228.

Crasborn, O. (2001). *Phonetic implementation of phonological categories in sign language of the netherlands.* Utrecht: LOT.

Crystal, D. (1980). *A first dictionary of linguistics and phonetics*. London: André Deutsch Limited.

Cucchiarini, C., H. Strik, & L. Boves. (2000). Different aspects of expert pronunciation quality ratings and their relation to scores produced by speech recognition algorithms. *Speech Communication, 30*(2-3), 109-119.

Cuxac, C. (1999). French Sign Language: Proposition of a Structural Explanation by Iconicity. In Braffort A. e. al. (Eds.), *Gesture-Based Communication in Human-Computer Interaction, LNAI 1739* (pp. 165-184): Springer

Cuxac, C. (2000). *La Langue des Signes Française. Les Voies de l'Iconicité* (Vol. No 15-16). Paris: Ophrys.

Dell, C. (1977). A primer for movement description using effort-shape and supplementary concepts (revised edition ed.). New York: Dance Notation Bureau Press.

Derpanis, K. G., R. P. Wildes, & J. K. Tsotsos. (2008). Definition and recovery of kinematic features for recognition of American sign language movements. *Image and Vision Computing, 26*(12), 1650-1662.

Derpanis, K. G., R. P. Wildes, & J. K. Tsotsos. (2004). Hand gesture recognition within a linguistics-based framework. *Lecture Notes in Computer Science 3021*, 282-296.

Dusan, S., & L. R. Rabiner. (2005). Can Automatic Speech Recognition Learn more from Human Speech Perception? In C. Burileanu (Ed.), *Trends in Speech Technology* (pp. 21-36). Bucharest, Romania: Romanian Academy Publisher.

Dye, M. W. G., & S.-I. Shih. (2006). Phonological priming in British Sign Language. In L. M. Goldstein, D. H. Whalen & C. T. Best (Eds.), *Laboratory Phonology 8*. Berlin: Mouton de Gruyter.

Eco, U. (1976). *A Theory of Semiotics*. Indiana University Press.

Edwards, A. D. N. (1998). Progress in Sign Language Recognition. *In Gesture and Sign Language in Human-Computer Interaction: International Gesture Workshop, Bielefeld, Germany, September 1997*. Proceedings (pp. 13-21).

Ekman, P. (1999). Emotional and conversational nonverbal signals. In L. S. Messing & R. Campbell (Eds.), *Gesture, speech, and sign* (pp. 45-55). New York: Oxford University Press.

Ekman, P., & W. Friesen. (1969). The repertoire of non-verbal behaviour: Categories, origins, usage and coding. *Semiotica, 1*(1), 49-98.

Emmerik, W., & G. Meyer (2005). Bewogen. Filmgedichten in gebarentaal. Amsterdam: Rubinstein. (with DVD).

Emmorey, K. (2002). *Language, cognition, and the brain: Insights from sign language research*. Mahwah, NJ: Lawrence Erlbaum and Associates.

Emmorey, K. (1999). Do signers gesture? In L. S. Messing & R. Campbell (Eds.), *Gesture, speech, and sign* (pp. 133-160). Oxford: Oxford University Press.

Emmorey, K., & D. Corina. (1990). Lexical recognition in sign language - effects of phonetic structure and morphology. *Perceptual and Motor Skills*, 71(3), 1227-1252.

Emmorey, K., & B. Falgier (2004). Conceptual locations and pronominal reference in american sign language. Journal of Psycholinguistic Research, 33(4), 321-331.

Emmorey, K., T. Grabowski, S. McCullough, H. Damasio, L. Ponto, R. Hichwa, & U. Bellugi. (2004). Motor-iconicity of sign language does not alter the neural systems underlying tool and action naming. *Brain and Language, 89*(1), 27-37.

Emmorey, K., S. McCullough, & D. Brentari. (2003). Categorical perception in american sign language. *Language and Cognitive Processes*, 18(1), 21-45.

Emmorey, K., B. Tversky, & H. A. Taylor. (2000). Using Space to Describe Space: Perspective in Speech, Sign, and Gesture. *Spatial Cognition and Computation, 2*(3), 157-180.

Field, A. (2005). *Discovering Statistics Using SPSS*. London: Sage Publications.

Franco, H., L. Neumeyer, V. Digalakis, & O. Ronen. (2000). Combination of machine scores for automatic grading of pronunciation quality. *Speech Communication, 30*(2-3), 121-130.

Gazdar, G., & C. Mellish (1989). Natural Language Processing. In *Prolog: An Introduction to Computational Linguistics.* Wokingham, England: Addison-Wesley. pp. 504.

Gerwing, J., & J. Bavelas. (2004). Linguistic influences on gesture's form. *Gesture*, 4(2), 157-195.

Greenbaum, S. (Ed.). (1977). *Acceptability in language*. The Hague: Mouton.

Greenbaum, S., & R. Quirk. (1970). *Elicitation Experiments in English: Linguistic Studies in Use and Attitude*. London: Longman.

Grosjean, F. (1996). Gating. Language and Cognitive Processes, 11(6), 597-604.

Grosjean, F. (1981). Sign & word recognition: A first comparison. Sign Language Studies, 28, 195-220.

Grosjean, F. (1980). Spoken Word Recognition Processes and the Gating Paradigm. *Perception & Psychophysics, 28*(4), 267-283.

Haber, R. N., & M. Hershenson. (1980). The psychology of visual perception (second ed.). New York: Holt, Rinehart and Winston.

Hall, J. (2004). Cicero and Quintilian on the oratorical use of hand gestures. *The Classical Quarterly, 54*, 143-160. doi:10.1093/cq/54.1.143

Hildebrandt, U., & D. Corina. (2002). Phonological similarity in American Sign Language. *Language and Cognitive Processes, 17*(6), 593-612.

Huang, W. C., T. L. Chang-Chien, & H. P. Lin. (2007). An intelligent multimedia E-learning system for pronunciations, *Lecture Notes in Computer Science 4570*, pp. 84-93.

Jackendoff, R. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution.* Oxford/New York: Oxford University Press.

Jaimes, A., & N. Sebe. (2007). Multimodal human-computer interaction: A survey. *Computer Vision and Image Understanding, 108*(1-2), 116-134.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics* (14), 201-211.

Junker, H., O. Amft, P. Lukowicz, & G. Tröster. (2008). Gesture spotting with body-worn inertial sensors to detect user activities. *Pattern Recognition, 41*(6), 2010-2024.

Kelly, S. D., S. Ward, P. Creigh, & J. Bartolotti. (2007). An intentional stance modulates the integration of gesture and speech during comprehension. *Brain and Language, 101*(3), 222-233.

Kendon, A. (2004). *Gesture: Visible action as utterance.* Cambridge: Cambridge University Press.

Kendon, A. (1994). Do gestures communicate? A review. *Research on Language and Social Interaction*, 27(3), 175-200.

Kendon, A. (1992). Some recent work from italy on quotable gestures ('emblems'). Journal of Linguistic Anthropology, 2(1), 72-93.

Kendon, A. (1987). Speaking and Signing Simultaneously in Warlpiri Sign Language Users. *Multilingua*, 6(1), 25-68.

Kendon, A. (1995). Gestures as illocutionary and discourse structure markers in southern italian conversation. Journal of Pragmatics, 23(3), 247-279.

Kendon, A. (1984). Did gesture have the happiness to escape the curse at the confusion of babel? In A. Wolfgang (Ed.), Nonverbal behaviour: Perspectives, applications, intercultural insights (pp. 75-114). Lewiston, NY: C.J. Hogrefe.

Kendon, A. (1980). Gesticulation and Speech: Two Aspects of the Process of Utterance. In H. R. Key (Ed.), *The Relationship of Verbal and Nonverbal Communication* (pp. 207-227). The Hague: Mouton and Co.

Kim, D., J. Song, & D. Kim (2007). Simultaneous gesture segmentation and recognition based on forward spotting accumulative HMMs. *Pattern Recognition, 40*(11), 3012-3026.

Kiraly, I., J. Bianca, P. Wolfgang, G. Aschersleben, & G. Gergely. (2003). The early origins of goal attribution in infancy. *Consciousness and Cognition*, 12(4), 752.

Kita, S., I. van Gijn, & H. van der Hulst. (1998). Movement phases in signs and co-speech gestures, and their transcription by human coders. *Lecture Notes in Computer Science*, 1371, 23.

Klima, E., & U. Bellugi. (1979). *The Signs of Language.* Cambridge, MA: Harvard University Press.

Koenen, L., T. Bloem, & R. Janssen. (1993). Gebarentaal - De taal van doven in Nederland. Amsterdam: Nijgh & Van Ditmar.

Kurzweil, R. (1999). *The Age of Spiritual Machines. When computers exceed human intelligence.* London, UK: Penguin Books Ltd.Laban, Rudolf, & F.C. Lawrence. (1947). *Effort.* London: Macdonald and Evans.

Laban, R. (1975). Laban's principles of dance and movement notation (Second edition, annotated and edited by Roderyk Lange ed.). London: Macdonald and Evans Ltd.

Lakoff, G. (1987). Women, fire, and dangerous things. Chicago: The University of Chicago Press.

Lee, S., V. Henderson, H. Hamilton, T. Starner, H. Brashear, & S. Hamilton (2005). A gesture-based american sign language game for deaf children. *ACM Int. Conf. on Computer-Human Interaction, CHI 2005,* Portland, Oregon, USA.

Levelt, W. J. M., J. A. W. M. Gent, A. F. J. Haans, & A. J. A. Meijers. (1977). Grammaticality, paraphrase, and imagery. In S. Greenbaum (Ed.), *Acceptability in language.* The Hague: Mouton.

Lichtenauer, J. F., E. A. Hendriks, & M. J. T. Reinders. (2008). Sign Language Recognition by combining Statistical DTW and Independent Classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 30*(11), 2040-2046.

Lichtenauer, J. F., G. A. ten Holt, M. J. T. Reinders, & E. A. Hendriks. (2007). Person-Independent 3D Sign Language Recognition. *Paper presented at the 7th International Workshop on Gesture in Human-Computer Interaction and Simulation (GW2007).*

Liddell, S. K., & R. E. Johnson (1989). American Sign Language: The phonological base. *Sign Language Studies 64*, 195–277

Lippmann, R. P. (1997). Speech recognition by machines and humans. *Speech Communication, 22,* 1-15.

Macken, E., J. Perry, & C. Haas. (1993). Richly grounding symbols in ASL. *Sign Language Studies* (December).

MacSweeney, M., R. Campbell, B. Woll, V. Giampietro, A. S. David, P. K. McGuire, G. A. Calvert, & M. J. Brammer, (2004). Dissociating linguistic and nonlinguistic gestural communication in the brain. *NeuroImage, 22*(4), 1605-1618.

Marr, D., & H. K. Nishihara. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London. Series B. Biological sciences*, 200(1140), 269.

Marr, D., & L. M. Vaina. (1982). Representation and recognition of the movements of shapes. *Proceedings of the Royal Society of London - Biological Sciences*, 214(1197), 501.

Marshall, J., J. Atkinson, E. Smulovitch, A. Thacker, & B. Woll. (2004). Aphasia in a user of British Sign Language: Dissociation between sign and gesture. *Cognitive Neuropsychology, 21*(5), 537-554.

McNeill, D. (1992). *Hand and mind: what gestures reveal about thought.* Chicago: University of Chicago Press.

Melinger, A., & W. J. M. Levelt. (2004). Gesture and the communicative intention of the speaker. *Gesture*, 4(2), 119-141.

Michotte, A. (1962). *The perception of causality*. New York: Basic Books.

Mitra, S., & T. Acharya (2007). Gesture recognition: A survey. *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews, 37*(3), 311-324.

Mohan, B. A. (1977). Acceptability testing and fuzzy grammars. In S. Greenbaum (Ed.), *Acceptability in language*. The Hague: Mouton.

Morris, D., P. Collett, P. Marsh, & M. O'Shaughnessy (1979). *Gestures. Their origins and distributions*. London: Jonathan Cape.

Morris, C.W. (1946). Signs, Language, and Behaviours. In: *Writings on the General Theory of Signs*. The Hague: Mouton, pp73-398

Mount, A. (2008). Intentions, gestures, and salience in ordinary and deferred demonstrative reference. *Mind and Language, 23*(2), 145-164.

Moustroufas, N., & V. Digalakis. (2007). Automatic pronunciation evaluation of foreign speakers using unknown text. *Computer Speech and Language, 21*(1), 219-230.

Müller, C. (1998). Iconicity and gesture. *Paper presented at the Oralité et gestualité. Communication multimodale, interaction. Actes du colloque Orage 98,* Besancon.

Nederlands Gebarencentrum. (2002). *Standaard lexikon nederlandse gebarentaal. Deel 1*. Bunnik: Nederlands Gebarencentrum.

Neumeyer, L., H. Franco, V. Digalakis, & M. Weintraub. (2000). Automatic scoring of pronunciation quality. *Speech Communication, 30*(2-3), 83-93.

Newport, E. L., & R. P. Meier. (1985). The Acquisition of American Sign Language. In D. I. Slobin (Ed.), *The crosslinguistic study of language acquisition* (Vol. 1). Hillsdale, NJ: Lawrence Erlbaum Associates.

Newtson, D., & G. Engquist. (1976). The perceptual organization of ongoing behaviour. *Journal of Experimental Social Psychology*, 12(5), 436-450.

Nöth, W. (1990). *Handbook of Semiotics*. Bloomington: Indiana University Press.

NSDSK. (1988). Notatie-systeem voor nederlandse gebaren. Amsterdam: Nederlandse Stichting voor het Dove en Slechthorende Kind.

Ong, S. C. W., & S. Ranganath (2005). Automatic sign language analysis: A survey and the future beyond lexical meaning. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 27*(6), 873-891.

Parish, D. H., G. Sperling, & M. S. Landy. (1990). Intelligent temporal subsampling of american sign language using event boundaries. *Journal of Experimental Psychology*, 16, 282-294.

Pavlovic, V. I., R. Sharma, & T. S. Huang (1997). Visual interpretation of hand gestures for human-computer interaction: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 19*(7), 677-695.

Peirce, C. S. (1965). *Collected Papers of Charles Sanders Peirce*. C. Hartshorne & P. Weiss (Eds.). Vols 1-8. Cambridge: The Belknap Press of Harvard U P.

Peirce, C. S. (1868). On a New List of Categories. *Proceedings of the American Academy of Arts and Sciences, 7*, 287-298.

Pietrandrea, P. (2002). Iconicity and arbitrariness in Italian Sign Language. *Sign Language Studies, 2*(3), 296–321.

Poizner, H., U. Bellugi, & V. Lutes-Driscoll. (1981). Perception of american sign language in dynamic point-light displays. *Journal of Experimental Psychology-Human Perception and Performance*, 7(2), 430-440.

Quek, F., R. Bryll, C. Kirbas, H. Arslan, & D. McNeill (2002). A multimedia system for temporally situated perceptual psycholinguistic analysis. *Multimedia Tools and Applications, 18*(2), 91-114.

Rizzolatti, G., & M. A. Arbib. (1998). Language within our grasp. *Trends in Neurosciences, 21*(5), 188-194.

Roh, M.-C., B. Christmas, J. Kittler, & S.-W Lee. (2008). Gesture spotting for low-resolution sports video annotation. *Pattern Recognition, 41*(3), 1124-1137.

Roodenburg, H. (2004). *The Eloquence of the Body*. Zwolle: Waanders Publishers.

Rosch, E. (1978). Principles of categorization. In B. B. Lloyd & E. Rosch (Eds.), *Cognition and categorization* (pp. 27-48). Hillsdale, NJ: Erlbaum.

Rothpletz, A. M., D. H. Ashmead, & A. M. Tharpe. (2003). Responses to targets in the visual periphery in deaf and normal-hearing adults. *Journal of Speech, Language, and Hearing Research*, 46(6), 1378.

Rubin, J. M., & W. A. Richards (1985). *Boundaries of visual motion*. MIT (CSAIL) report, AI Memo 835. (ftp://publications.ai.mit.edu/ai-publications/pdf/AIM-835.pdf)

Sacks, H., & E. A. Schegloff. (2002). Home position. *Gesture*, 2(2), 133-146.

Schifferstein, R., & P. Hekkert (Eds.). (2008). *Product experience*. Amsterdam: Elsevier.

Sharma, R., M. Yeasin, N. Krahnstoever, I. Rauschert, G. Cai, I. Brewer, et al. (2003). Speech-gesture driven multimodal interfaces for crisis management. *Proceedings of the IEEE, 91*(9), 1327-1354.

Siple, P., F. Caccamise, & L. Brewer. (1982). Signs as pictures and signs as words: Effect of language knowledge on memory for new vocabulary. . *Journal of Experimental Psychology: Learning, Memory, and Cognition, 82,* 619-625.

Sorace, A. (1996). The Use of Acceptability Judgments in Second Language Acquisition Research. In *Handbook of Second Language Acquisition* (pp. 375-409): Academic Press.

Spaai, G. W. G., C. Fortgens, M. Elzenaar, E. Wennes, J. F. Lichtenauer, E. A. Hendriks, H. de Ridder, J. Arendsen, & G. A. ten Holt (2008). Een computerprogramma voor het leren van actieve en passieve gebarenschat aan ernstig slechthorende en dove kinderen. *Logopedie en Foniatrie* (2), 42-50.

Starner, T., J. Weaver, & A. Pentland. (1998). Real-time american sign language recognition using desk and wearable computer based video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(12), 1371-1375.

Stiefelhagen, R., H. K. Ekenel, C. Fuegen, P. Gieselmann, H. Holzapfel, F. Kraft, et al. (2007). Enabling multimodal human-robot interaction for the Karlsruhe humanoid robot. *IEEE Transactions on Robotics, 23*(5), 840-851.

Stokoe Jr, W. C. (1960/2005). Sign language structure: an outline of the visual communication systems of the American deaf. 1960. *Journal of deaf studies and deaf education, 10*(1), 3.

Streeck, J. (2008). Depicting by gesture. *Gesture, 8*(3), 285-301.

Ten Holt, G. A., A. J. van Doorn, H. de Ridder, M. J. T. Reinders, & E. A. Hendriks. (2009a). Which Fragments of a Sign Enable Its Recognition? *Sign Language Studies, 9*(2), 211-239.

Ten Holt, G. A., J. Arendsen, H. de Ridder, A. J. Koenderink-van Doorn, M. J. T. Reinders, & E. A. Hendriks. (2009b). Sign Language Perception Research for Improving Automatic Sign Language Recognition. *Paper presented at the IS&T/SPIE Electronic Imaging Conference 2009, Human Vision and Electronic Imaging XIV.*

Ten Holt, G., P. Hendriks, & T. Andringa. (2006). Why don't you see what I mean? Prospects and limitations of current automatic sign recognition research. *Sign Language Studies, 6*(4), 416-437.

Tervoort, B. T. M. (1953). *Structurele analyse van visueel taalgebruik binnen een groep dove kinderen* (Proefschrift Amsterdam, U.v.A. ed.). Amsterdam: Noord-Hollandsche Uitgevers Maatschappij.

Tolar, T. D., A. R. Lederberg, S. Gokhale, & M. Tomasello. (2008). The development of the ability to recognize the meaning of iconic signs. *Journal of Deaf Studies and Deaf Education, 13*(1), 71-86.

Troscianko, T., A. Holmes, J. Stillman, M. Mirmehdi, D. Wright, & A. Wilson. (2004). What happens next? The predictability of natural behaviour viewed through cctv cameras. *Perception*, 33, 87-101.

Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind, LIX*(236), 433-460.

Uyechi, L. (1996). *The geometry of visual phonology*. Stanford, CA: CSLI.

Van Dijk, T. A. (1977). Acceptability in Context. In S. Greenbaum (ed.), *Acceptability in Language.* The Hague, Mouton: 39-61.

Van der Hulst, H., & E. van der Kooij (2006). Phonetic implementation and phonetic pre-specification in sign language phonology. In L. Goldstein, D. H. Whalen & C. T. Best (Eds.), *Papers in laboratory phonology* (Vol. 8, pp. 265-286). Berlin/New York: Mouton de Gruyter.

Van der Kooij, E.(2002). *Reducing phonological categories in sign language of the netherlands. Phonetic implementation and iconic motivation.* PhD Dissertation, Leiden University.

Vogler, C., & S. Goldenstein. (2008). Facial movement analysis in ASL. *Universal Access in the Information Society, 6*(4), 363-374.

Vogler, C., & D. Metaxas (2004). Handshapes and movements: Multiple-channel american sign language recognition. *Lecture Notes in Artificial Intelligence (Subseries of Lecture Notes in Computer Science), 2915,* 247-258.

Vogler, C., & D. Metaxas. (1999). Parallel hidden Markov models for American sign language recognition. *Proceedings of the Proceedings of the IEEE International Conference on Computer Vision*, 116-122.

Von Agris, U., J. Zieren, U. Canzler, B. Bauer, & K.-F. Kraiss. (2008). Recent developments in visual sign language recognition. *Universal Access in the Information Society, 6*(4), 323-362.

Willems, R., A. Özyurek, A., & P. Hagoort (2007). When language meets action: The neural integration of gesture and speech. *Cerebral Cortex, 17*(10), 2322-2333.

Wilson, A. D., A. F. Bobick, & J. Cassell (1996). Recovering the temporal structure of natural gesture. *Paper presented at the International Conference on Face and Gesture Recognition '96*. IEEE.

Witt, S. M., & S. J. Young. (2000). Phone-level pronunciation scoring and assessment for interactive language learning. *Speech Communication, 30*(2-3), 95-108.

Wu, Y. C., & Coulson, S. (2007). How iconic gestures enhance communication: An ERP study. *Brain and Language, 101*(3), 234-245.

Xiong, Y. G., & F. Quek (2006). Hand motion oscillatory gestures and multimodal discourse analysis. *International Journal of Human-Computer Interaction, 21*(3), 285-312.

Zhang, L.-G., Y. Chen, W. Gao, G. Fang, & X. Chen. (2004). A vision-based sign language recognition system using tied-mixture density hmm. Paper presented at the *ICMI'04 - Sixth International Conference on Multimodal Interfaces.*

Zieren, J., & K.-F. Kraiss. (2005). Robust person-independent visual sign language recognition. *Lecture Notes in Computer Science*, 3522, 520.

Zwitserlood, I. (2003). *Classifying hand configurations in Nederlandse gebarentaal (sign language of the Netherlands).* LOT dissertation series; 78. Utrecht: LOT.

Summary

This dissertation presents the results of a series of studies on the appearance of manual movements in gestures. The main goal of this research is to increase our understanding of how humans perceive signs and other gestures. Generated insights from human perception may aid the development of technology for recognizing gestures and sign language automatically with cameras and computers. One example of an application of automatic gesture recognition that has played a role in shaping the research in this dissertation is *ELo*, an Electronic Learning environment for deaf and hearing impaired children to practice Sign Language of the Netherlands (SLN) signs (Spaai et al, 2008). The questions addressed in the research focus on a number of aspects including temporal processing of signs, discrimination of gestures from other human behaviour, and how humans handle variation in signs.

In chapter 2 we studied if and when people detect the beginning of a sign by presenting movie fragments consisting of sequences of rest positions, fidgets, and signs to deaf signers, hearing signers and non-signers. Participants were instructed to respond as soon as they saw that a SLN sign had begun. All participants showed themselves highly capable of responding to sign beginnings. Signs that are two-handed, performed in signing space, have a highly marked hand shape, and contain path movement were discriminated best. Considering a sign as having a preparation, a stroke, and a recovery, response times showed strong clusters around 500 milliseconds after the beginning of sign preparation, or 200 ms after the onset of the stroke. The non-signers needed more time before responding; deaf signers took more time than hearing signers. Response time was influenced by three factors (shorter for signs that have a highly marked hand shape, are one-handed, and are preceded by fidgets). The results showed that it is possible for people to discriminate fidgeting and signs based on appearance, even if one does not know sign language. No single feature of the movement appeared necessary to detect the beginning of a sign. In most cases visual information available up to an early stage of the stroke was sufficient but in some cases the information in the preparation was enough.

In the experiment described in chapter 3, we studied whether there are visible differences between lexical signs, emblems (i.e. highly conventionalized gestures) and fidgeting. To focus on the appearance of the movements instead of their meaning we selected non-signers as participants. They were shown movies with a single lexical sign, an emblem, or a fidgeting movement. They were instructed, as in the experiment reported in chapter 2, to press the spacebar as soon as they judged the movement to be a sign. Participants were found to be well able to let the fidgeting movements pass without pressing, but to press almost equally often in response to lexical signs as to emblems. Emblems that were commonly known in the Netherlands elicited pressing less often than emblems not commonly known. However, this difference was entirely due to four emblems with an offensive meaning which many participants did not judge to be SLN signs. These results showed that, based solely on appearances, non-signers are typically not able to discriminate signs from emblems, but they are typically able to discriminate between fidgeting and movements that are intended to communicate (emblems and SLN signs).

In chapter 4 we studied how much time people need to recognize a sign as it unfolds. Deaf and hearing signers were presented movie fragments with sequences of restpositions, fidgets, and mono-morphemic signs. They watched these movies at normal playing speed and had to respond as soon as they recognized the lexical meaning of a sign, which they were able to do after around 850 ms, counting from the beginning of the sign. By subtracting participants' reaction times to seeing a motion boundary (average of 310 ms) we estimated that confident sign recognition starts after around 540 ms, in the sense that the necessary information has

become available in the signal. If we think of a sign as consisting of three main movement phases (i.e. preparation, stroke and recovery) lexical recognition starts about 220 ms after the onset of the stroke. By comparing the raw data from the experiments described in chapter 2 and 4, lexical recognition was found to take about 90 ms longer than detection.

Chapter 5 contains the results of an experiment in which signers were asked to judge the acceptability of a set of sign manipulations. Signs were recorded with variations in different categories, in the temporal and spatial dimension. Participants varied much in tolerance, i.e. in the percentage of movies they judged to be acceptable, but their rankings of the acceptability of sign manipulations correlated well. On the level of dimensionality we found that temporal manipulations were highly acceptable while spatial (and spatiotemporal) manipulations were often judged unacceptable. Further division of the manipulations into categories, such as changes in hand orientation or movement direction, showed much variability and little regularity. The roles of phonology and iconicity in acceptability judgments were studied. Part of the variability in the acceptability of sign manipulations could be explained on the basis of the type of phonological error caused by each manipulation, and by considering a sign's iconicity and classifying whether manipulations are compatible or incompatible with that iconicity. Finally, human judgments were compared to acceptability ratings by three automatic sign recognizers.

In chapter 6 the findings of the experiments in the previous chapters are integrated and discussed in a more general sense. Several issues are highlighted that may be relevant for our current understanding of how people perceive and interpret manual movements. In addition, a shortlist is provided with those insights that might be useful to inspire further developments in automatic gesture and sign language recognition. Finally, some ideas for further research are presented.

Samenvatting

Dit proefschrift bevat de resultaten van een reeks studies naar de verschijningsvorm van handbewegingen in gebaren. Het primaire doel van dit onderzoek is het uitbreiden van onze kennis met betrekking tot de menselijke waarneming van gebaren, zowel gebarentaal als anderszins. De verkregen inzichten vanuit de menselijke perceptie kunnen gebruikt worden ter bevordering van de ontwikkeling van technologie voor het automatisch herkennen van gebaren met behulp van camera's en computers. Een voorbeeld van een toepassing van automatische gebarenherkenning, die bij het huidige onderzoek ook een rol heeft gespeeld, is *ELo*, een Elektronische LeerOmgeving voor dove en slechthorende kinderen om Nederlandse Gebarentaal (NGT) te oefenen (Spaai et al, 2008). De vragen die tijdens het onderzoek zijn behandeld richten zich op een aantal aspecten waaronder de temporele verwerking van gebaren, het onderscheid maken tussen gebaren en ander menselijk gedrag, en de menselijke verwerking van variatie in gebaren.

In hoofdstuk 2 is bestudeerd of en hoe snel mensen het begin van een gebaar kunnen detecteren door aan dove en horende NGT gebruikers en mensen zonder NGT ervaring filmfragmenten te laten zien bestaande uit opeenvolgingen van rustposities, gefrunnik en NGT gebaren. De deelnemers werden geïnstrueerd om te reageren zodra zij het begin van een NGT gebaar zagen. Alle deelnemers bleken in staat om te reageren als er inderdaad een NGT gebaar begon. Gebaren die worden gemaakt met twee handen, in de neutrale ruimte, met een zwaar gemarkeerde handvorm en met een verplaatsing over een pad werden het best onderscheiden. Als men een gebaar beschouwt als opgebouwd uit een voorbereiding, een kern, en een terugtrekking, vertoonden de gemeten responsie-tijden sterke clusters rond 500 ms na het begin van de voorbereiding van het gebaar, oftewel 200 ms na het begin van de kern. De mensen zonder NGT ervaring hadden meer tijd nodig om te reageren; dove NGT gebruikers namen meer tijd dan horende NGT gebruikers. Responsie-tijden werden beïnvloed door drie factoren (korter voor gebaren met een zwaar gemarkeerde handvorm, met één hand en bij een voorafgaande frunnikbeweging). De resultaten toonden aan that mensen in staat zijn om onderscheid te maken tussen gefrunnik en gebaren op basis van hoe deze bewegingen eruit zien, zelfs als men geen gebarentaal kent. Geen van de kenmerken van de beweging leek, op zichzelf, noodzakelijk om het begin van een gebaar te detecteren. In de meeste gevallen was de informatie die beschikbaar kwam tot aan een vroeg gedeelte van de kern voldoende, maar in sommige gevallen was zelfs de informatie in de voorbereiding al genoeg.

In het experiment beschreven in hoofdstuk 3 onderzochten wij of er zichtbare verschillen zijn tussen lexicale gebaren, emblemen (d.w.z. zwaar geconventionaliseerde gebaren) en gefrunnik. Om de focus te leggen op de verschijningsvorm van de bewegingen in plaats van hun betekenis werden mensen zonder NGT ervaring als deelnemers geselecteerd. Deze mensen kregen filmfragmenten te zien met daarin een enkel NGT gebaar, embleem of frunnikbeweging. Zij werden geïnstrueerd, net zoals in het experiment uit hoofdstuk 2, om op de spatiebalk te drukken zodra zij oordeelden dat de beweging een NGT gebaar was. Deelnemers bleken nagenoeg even vaak te drukken in reactie op lexicale NGT gebaren als op emblemen. Emblemen die algemeen bekend zijn in Nederland lokten minder vaak drukken uit dan emblemen die niet algemeen bekend zijn. Echter, dit verschil werd geheel veroorzaakt door vier emblemen met een beledigende betekenis, waarvan veel deelnemers niet oordeelden dat het NGT gebaren waren. Deze resultaten tonen aan dat, indien men alleen kan afgaan op de verschijningsvorm, mensen zonder NGT ervaring gewoonlijk niet in staat zijn om onderscheid te maken tussen NGT gebaren en emblemen, maar wel tussen gefrunnik en bewegingen die bedoeld zijn om te communiceren (emblemen en NGT gebaren).

In hoofdstuk 4 is bestudeerd hoeveel tijd mensen behoeven om een NGT gebaar te herkennen naarmate het zich ontvouwt. Dove en horende NGT gebruikers kregen filmfragementen te zien met opeenvolgingen van rustposities, gefrunnik en niet-samengestelde NGT gebaren. Zij bekeken deze filmfragmenten op normale afspeelsnelheid en moesten reageren zodra zij de lexicale betekenis van het gebaar herkenden, hetgeen men in staat was te doen na een tijd van rond de 850 ms, gerekend vanaf het begin van het gebaar. Door de reactietijd van deelnemers op het zien van een bewegingsgrens (gemiddeld 310 ms) hiervan af te trekken is ingeschat dat men het gebaar na ongeveer 540 ms met enige zekerheid begint te herkennen, in de zin dat dan de daarvoor benodigde informatie beschikbaar is gekomen in het signaal. Als wij een gebaar beschouwen als zijnde samengesteld uit drie hoofdfases van beweging (voorbereiding, kern en terugkeer) dan start lexicale herkenning ongeveer 220 ms na het begin van de kern. begin van de kern. Door het vergelijken van de ruwe data uit de experimenten beschreven in hoofdstukken 2 en 4 werd gevonden dat lexicale herkenning circa 90 ms langer duurt dan het detecteren van het begin van een gebaar.

Hoofdstuk 5 bevat de resultaten van een experiment waarin NGT gebruikers werd gevraagd van een groot aantal gebarenmanipulaties te beoordelen of deze acceptabel waren. Gebaren waren opgenomen met variaties in verschillende categorieën, in de temporele en de spatiële dimensie. De deelnemers verschilden veel in hun tolerantie, d.w.z. in het percentage filmfragmenten dat zij acceptabel vonden, maar hun rangschikkingen van de mate waarin gebaarmanipulaties acceptabel waren vertoonden goede correlaties. Op het niveau van dimensionaliteit werd gevonden dat temporele manipulaties zeer vaak acceptabel werden gevonden terwijl spatiële (en spatiotemporele) manipulaties vaak onacceptabel werden gevonden. Verdere onderverdelingen van de manipulaties naar categorieën, zoals veranderingen in handoriëntatie of bewegingsrichting vertoonden veel variabiliteit en weinig systematiek. Tevens werd de rol van fonologie en iconiciteit bestudeerd in het beoordelen of iets acceptabel is. Een deel van de variantie in hoevaak een gebaarmanipulatie acceptabel werd gevonden kan niet alleen verklaard worden aan de hand van de aard van de fonologische overtreding die door iedere manipulatie werd veroorzaakt, maar ook door de iconiciteit van het gebaar in ogenschouw te nemen en te classificeren of manipulaties wel of niet in overeenstemming zijn met die iconiciteit. Tot slot zijn de menselijke beoordelingen vergeleken met acceptabiliteitsbeoordelingen zoals gegenereerd door drie automatische gebaarherkenners.

In hoofdstuk 6 worden de bevindingen uit de experimenten in de voorgaande hoofdstukken geïntegreerd en bediscussieerd in meer algemene zin. Diverse punten worden er uit gelicht die wellicht relevant zijn voor ons huidige begrip van de menselijke waarneming en interpretatie van handbewegingen. Daarnaast is een selectie gemaakt van die inzichten die wellicht bruikbaar zijn als inspiratiebron voor verdere ontwikkelingen in de automatische gebaarherkenning. Tot slot worden enkele ideeën voor verder onderzoek gepresenteerd.

Publications

Arendsen, Jeroen, Andrea J. van Doorn, & Huib de Ridder. (submitted). Can non-signers differentiate between SLN signs, emblems and fidgeting? *Proceedings of GW2009, Lecture Notes in Computer Science* (submitted).

Arendsen, Jeroen, Andrea J. van Doorn, & Huib de Ridder. (in press). When do people start to recognize signs? *Gesture*. (in press).

Holt, Gineke A. ten, Jeroen Arendsen, Huib de Ridder, Andrea J. Koenderink-van Doorn, Marcel J. T. Reinders, & Emile A. Hendriks. (2009). Sign Language Perception Research for Improving Automatic Sign Language Recognition. *Proceedings of the IS&T/SPIE Electronic Imaging Conference 2009, Human Vision and Electronic Imaging XIV.* , 72400C-72400C-72412.

Arendsen, Jeroen , Jeroen F. Lichtenauer, Andrea J. van Doorn, Emile A. Hendriks, Gerard Spaai, Marcel J.T. Reinders, & Huib de Ridder. (2008). Acceptability Ratings by Humans and Automatic Gesture Recognition for Variations in Sign Productions. *Proceedings of the 8th IEEE International Conference on Automatic Face and Gesture Recognition*, 6-11.

Lichtenauer, Jeroen F., Gineke A. ten Holt, Emile Hendriks, Marcel J. T. Reinders, Abbie S. J. Vanhoutte, Irene Kamp, Jeroen Arendsen, Andre J. van Doorn, Huib de Ridder, Elmar Wenners, Marielle Elzenaar, Gerard Spaai, Connie Fortgens & M. Bruins (2008). A learning environment for sign language. *Proceedings of the 8th IEEE International Conference on Automatic Face and Gesture Recognition*, 88-89.

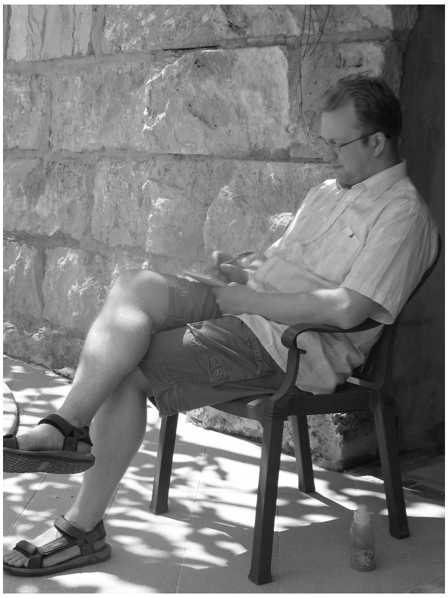
Spaai, Gerard W. G., Connie Fortgens, Marielle Elzenaar, Elmar Wennes, Jeroen Lichtenauer, Emile A. Hendriks, Huib de Ridder, Jeroen Arendsen, & Gineke ten Holt. (2008). Een computerprogramma voor het leren van actieve en passieve gebarenschat aan ernstig slechthorende en dove kinderen. *Logopedie en Foniatrie*(2), 42-50.

Arendsen, Jeroen, Andrea J. van Doorn, & Huib de Ridder. (2007). When and how well do people see the onset of gestures? *Gesture, 7*(3), 305-342.

Spaai, Gerard, Jeroen F. Lichtenauer, Emile A. Hendriks, Hanneke de Ridder, Jeroen Arendsen, Huib de Ridder, Connie Fortgens, Marjan Bruins, & Mariëlle Elzenaar. (2005). Elo: An electronic learning environment for practising sign vocabulary by young deaf children. *Proceedings of the International Congress for Education of the Deaf (ICED), July 2005.*

Arendsen, J. (2005). Sign Language Perception: Detecting a Sign. *Proceedings of the International Congress for Education of the Deaf (ICED), July 2005.*

About the Author



Jeroen Arendsen, was born on June 22, 1972, in Uden, the Netherlands and grew up in Best. He received his preparatory education (VWO) at the Jacob-Roelandslyceum in Boxtel, finishing in 1990. He studied industrial design engineering (IDE) at Delft University of Technology (DUT), obtaining a master of science degree in 1998 after a brief intermezzo during which he held his first position at KPN Research as a junior researcher. He then worked for six years as an interaction designer and usability specialist in the industry (KPN Telecom, Philips), gaining experience in the application of speech recognition and the methods of user centered design (UCD). In 2004 he returned to the DUT faculty of IDE, accepting a position as promovendus (aio), and performed research until 2009 to obtain his PhD degree.

In May 2009, Jeroen Arendsen accepted a position as a scientist at TNO Human Factors in Soesterberg, combining academic interests with UCD practice in various R&D projects. Current research interests include gesture and speech technology, human computer interaction, social robotics and creating entertaining user experiences.

Acknowledgements

If you are working on sign language there is often a reason for having chosen to do so. It is not unlikely you grew up with sign language because you or someone in your family (or a close friend) relied on it as a primary means of communication due to hearing loss. Usually one acquires a language before making it one's business. For me, that is not the case. Beyond the phrase 'what would you like on your sandwich?' which I can sign with only a few small errors, I have not mastered *Nederlandse Gebarentaal* (NGT) or, to use the English name of the language, Sign Language of the Netherlands (SLN). In my personal opinion this requires some explanation, and in my case it can be explained by recounting how the perception of sign language signs came to be the main part of my PhD work, and also by explaining the broader perspective of this dissertation on the perception of different kinds of gestures.

The perception of signs and other gestures became my cherished PhD topic through a few choices of my own but certainly also through coincidence and the good offices of others. One choice of my own was to decide I wanted to do a PhD. After my graduation from IDE in 1998, I had drifted from an R&D environment at KPN Research to operational business at KPN and Philips. But research had always remained on my mind, and when, due to reorganizations, the good people of Philips provided a financial reward for anyone pursuing his career elsewhere I started to look around.

Jans Aasman, who was my professor during my graduation as a Master student and a source of inspiration in my first working years at KPN Research (which is now TNO ICT), recommended me to Huib de Ridder. At the time I did not understand why he seemed to hand me over to his colleague, but in my first weeks on the job I find out why: He moved to America. Professor Huib de Ridder was kind enough to offer some options to a young man without any thorough scientific experience or shiny publications on his resume. Specifically, there were options in the field of gesture recognition, an area of interest to the ICT Delft Research Centre of Delft University of Technology, in which Huib de Ridder was cooperating with several groups at the EWI faculty. One of those groups was the ICT department, where Emile Hendriks was writing a proposal called NOGRIP (Non-Obtrusive Gesture Recognition for Intuitive Performance). One PhD student, Jeroen Lichtenauer, had already begun working in this area under the guidance of Emile and of Marcel Hendriks, who became a full professor in the area of machine learning recently, and the plan was to create four, later three, PhD positions. With a background in speech recognition and a specific interest in a combination of language and technology I seemed to fit well enough and I became the second PhD student, under Huib's guidance. He suggested Ans Koenderink van Doorn as my mentor.

NOGRIP was about the development of gesture recognition technology and the application of this technology. Realistic applications help uncover the requirements for applicable technology. One envisioned application was an electronic learning environment for deaf children to practice sign language vocabulary. This application was the main goal of a parallel project called ELo, led by the *NSDSK* (Nederlandse Stichting voor het Dove en Slechthorende Kind) in Amsterdam for which Delft would supply the technology and consultancy. Gerard Spaai had just started as research director at the NSDSK and taken over the lead in the project which had been initiated by Hanneke de Ridder, who was director of the NSDSK and wife to Huib. This ELo project, in which Delft had a share, received a grant from the VSB fund, and even though further external funding was not yet realized, the ICT Delft Research Centre financed the start of the work (and I have to thank them, specifically Laura Zondervan, for continued financing of my PhD position). ELo became the application driving the development of the technology and to some extent it also inspired our research questions with regards to the perception of signs. Of all types of gestures, sign language signs appeared to be viable candidates for automatic recognition given that they were part of a lexicon of semiotic signs with well defined meanings. That is at least how we imagined sign language to be at the time. But our interest has never been confined to sign language, because the technology was and still is believed to be applicable to the automatic recognition of other types of gestures as well. This should explain why sign language signs are primarily treated as cases of gestures in this dissertation. It is also one of the reasons why I studied the perception of signs for about five years without feeling the absolute necessity of mastering that sign language myself. My interest in sign language is part of my broader interest in gesture.

I would like to take this opportunity to thank all those NGT users whose expectations regarding the progress of technology we sometimes had to lower for their patience and understanding. Many signers have also acted as informants and participated in our experiments or helped in the production of movies of signs for which I am very grateful. Several people at different organizations were kind enough to help me reach such signers: Jacqueline van Dalen (NSDSK and daycare centre "Wip Wapper" in Amersfoort), Mariëlle Elzenaar (NSDSK, Hogeschool Utrecht), Beppie van den Bogaerde (Hogeschool Utrecht), Connie Fortgens (Royal Auris Foundation, Polanoschool Rotterdam), Trude Schermer (Nederlands Gebarencentrum), Johan Ros (Radboud University), Inge Zwitserlood en Daan Hermans (Viataal) and many others who recommended me to friends or family. Mariëlle Elzenaar also performed the signing (and the fidgeting) in most of the material used in the experiments and I fondly remember our recording sessions.

My time as a PhD student would not have been the same without Ans Koenderink-van Doorn, my 'academic mother'. If I was perhaps sometimes a headstrong thirty-something youngster, now is the time to apologize. She taught me many things, not just about perception research but also about the academic world and about art. Our long conversations always started with some aspect of gesture but often wandered off in different directions. She would let me ramble on, maybe join in with her vast experience, but remind me in the end of the work at hand. Concerning the bread and butter of science, which is writing about experimental data, Ans was the most dedicated coach one could wish for. All my writings, even the feeble ones, quickly came back with serious comments, and my writing style has undergone major changes as a result. Huib de Ridder was never far away from the work I was discussing with Ans, but kept a close watch, especially near the final stages of a piece of work. He always kept an open mind about our results, often encouraging me to analyze them in another way to gain further insight. Whenever he would find the time to read them, my writings would come back full of suggestions and comments. To be guided so intensely by these two 'concerned teachers' must be any PhD student's privilege.

The Delft University of Technology was not a great store of knowledge about gestures and sign language when I started in 2004, although Caroline Hummels provided helpful entry points into the literature on gesture recognition. Many people at other research facilities have been a great help in finding my way in the scientific fields of gesture and sign language. First among these are Marianne Güllberg and Asli Özyürek who, at the time, led the Nijmegen Gesture Centre at the Max Planck Institute for Psycholinguistics, and Amanda Brown and Roel Willems who introduced me there. The series of Gesture lectures organized at the MPI have been a great source of knowledge, contacts and inspiration throughout the years. At several occasions I was also invited to present my work in progress there for which I was very grateful. I always greatly enjoyed the discussions surrounding such lectures and for me it was a chance to meet many of the wonderful minds working on Gesture in various disciplines. One of these is Adam Kendon who has been very supportive of my research and my writing. Adam Kendon and a second reviewer made valuable suggestions that noticeably improved the paper that is included here as chapter 2 and he also guided other writings. Another great store of knowledge was also situated in Nijmegen: the NGT group at the Radboud University around Onno

Crasborn, with Els van der Kooij, Inge Zwitserlood, and Johan Ros. They gently introduced me to their work on phonological categories and phonetic implementation while also serving as champions of the richness of natural signed languages. As part of the ELo project, Connie Fortgens (Auris) and Mariëlle Elzenaar (NSDSK) were also always nearby to deal with questions regarding the signs we were working with.

Jeroen Lichtenauer and Gineke ten Holt were my fellow PhD students in Delft and I have greatly enjoyed working together with them in our research. I believe that together with the people at the NSDSK and Auris we were part of a very interesting multidisciplinary project. Gerard Spaai led this group of specialists, and succeeded in giving room to everyone, whilst not forgetting the end goal of the project: to test the feasibility of ELo as an educational aid. And even though there were sometimes difficulties communicating across disciplines I feel all people involved showed a lot of respect for each other, and for me this was in itself a very educational experience. Jeroen Lichtenauer took his time to explain to me often difficult concepts and details about his algorithms, while I tried to convince him of the necessity of solving certain 'usability' problems (such as the need to 'put your hands on the table' before and after a sign). With time I developed a better understanding of the algorithms involved in automatic gesture recognition and I am grateful to Jeroen, Gineke, Emile and Marcel for the learning experience. Gineke ten Holt also coded movement phases and performed as actor in part of the material. Jeroen Lichtenauer arranged and performed the tests with automatic sign recognition algorithms reported in chapter 5. Elmar Wennes and Michel Esveld were responsible for the ELo software design and development, for which I provided usability consultancy, and they always remained cheerful whenever it was time to get the prototyping job done (maybe once or twice Elmar got a little grumpy while playing 'Wizard-of-Oz' during the ELo field trials).

To all the girls who shared my room: Life and lunch at IO would not have been the same without you. Elif and Jenneke always had a sympathetic word or a song to brighten up the day, and we shared a lot of 'PhD worries'. Els introduced me to everyone when I first started, most importantly the 10 o'clock coffee club with Arnold, Arend, Theo, Henk, and Kees and the others. Arend Harteveld†, who passed away so unexpectedly, was a great support: He created both the software that was used to record the experiments and the software that was used to code the movement phases of the material used in chapters 2 and 4.

I really enjoyed the comradeship of my fellow PhD students and assorted IO regulars at the second and third floor and the ID StudioLab. Gaël was often in for yet another cup of coffee (bringing chocolates if I was lucky) but that was really just an excuse for us to continue arguing about whatever had come up during lunch. Jasper, with his enthusiasm and his uselog, inspired me to create my own weblog 'A Nice Gesture'. Students Johan van Gastel and Deniz Unsal conducted the experiments reported in chapter 3. Marijke, Elif, and Armagan made Aula-dinners bearable with their comradeship if we stayed on, working until security would kindly remind us to pack it in. Elif also designed the cover and was a great support during the final week of pre-printing stress. Piet Westendorp†, despite all his charm, was unsuccessful at making a dinner at Luniz bearable, but, more importantly, he did teach me, in his own way, some important lessons about the academic world and my position in particular. Piet also passed away unexpectedly and his presence is sorely missed.

My current employer TNO, in the person of Myra van Esch-Bussemakers, has shown considerable patience and understanding in the final months of preparing this dissertation, which overlapped my 'day job', and I am grateful for it.

Closer to home I am indebted to Edwin for the good times and for reminding me of my project schedule, and for kicking my lazy butt during fitness now and then. I have seen many of my friends less than I would have wished, as I have been somewhat preoccupied: mea culpa. And of course I could not have done this without the loving support of parents, family and, above all, my wife Corine (I promise I'll never do it again, dear). To our children, Leonie, Rik, and Sabine, I dedicate this book because science is for the future and the future, my dear little angels, belongs to you.