Talking hands

Reference in speech, gesture, and sign

Marieke Hoetjes

Talking hands. Reference in speech, gesture, and sign

Marieke Hoetjes
PhD Thesis
Tilburg University, 2015

TiCC PhD series No. 40

Financial support was received from The Netherlands Organization for Scientific Research (NWO) for the Vici-project "Bridging the gap between psycholinguistics and computational linguistics: The case of referring expressions".

ISBN: 978-94-6203-867-7
Print: CPI Wöhrmann print service
Cover design: Marlous Bervoets

Talking hands
Reference in speech, gesture, and sign

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan Tilburg University
op gezag van de rector magnificus,
prof.dr. E.H.L. Aarts,
in het openbaar te verdedigen ten overstaan van
een door het college voor promoties aangewezen commissie
in de aula van de Universiteit
op woensdag 7 oktober 2015 om 14.15 uur
door

Marieke Wilhelmina Hoetjes

geboren op 20 oktober 1980 te Slijk-Ewijk

Promotores

Prof. Dr. Emiel Krahmer

Prof. Dr. Marc Swerts

Promotiecommissie

Dr. Sonia Frota

Dr. Albert Gatt

Prof. Dr. David House

Prof. Dr. –Ing. Stefan Kopp

Prof. Dr. Alfons Maes

Contents

1

Introduction

I remember watching the news on television when I was about six or seven years old, and, not being able to understand the politicians' talk about complicated things, wondering when the time would come when I would at least be able to understand what they were saying with their hands. I honestly thought that one day I would be able to 'read' their hands, just as I was able to listen to, and, to some extent, understand their speech. At a young age, I was already aware that communication does not only consist of auditory aspects, but that the movements that people make with their hands as they are talking also play a role. Fast forward some twenty-odd years and there I was, listening to, but especially watching, Al Gore give a speech on the occasion of receiving an honorary doctorate at Tilburg University[1]. As with many politicians, he was a passionate speaker, making good use of his hands, and finally I was able to understand what he was saying, not just in speech, but also in gesture. His hands and arms were not making random movements, but the gestures he produced were nicely aligned with the content of his messages. Al Gore also varied his gesture production; some of his gestures occurred more often or were larger than others, and this did not seem to be random either. He also seemed to produce these gestures not just for himself, but especially for the audience; his gesture production made his speech fascinating to watch and listen to. In short, he was letting his hands do a lot of the talking. This thesis concerns some of the things that Al Gore's speech exemplified: variance in gesture production, and the effect of this variance on gesture perception.

Unfortunately for my chances to be recognized as a six-year-old genius, but luckily for science, I was not the first to assume that gesture is somehow relevant, and related, to speech. David McNeill, in his seminal book Hand and Mind (McNeill, 1992), mentions how interest in gesture dates back at least two millennia. However, it wasn't until around the time when I first wondered about politicians' "language in their hands" (Mol, 2011) that gesture studies as a field of research emerged, most notably with work by Adam Kendon (1980, 1986, 2004) and David McNeill (1985, 1992). The field of gesture studies has been flourishing ever since, especially in the last few decades, with people studying different types of gestures, the relationship between gesture and speech, the role of gesture in communication, the use of gesture in (second) language

[1] Part of Al Gore's speech can be found online: https://www.youtube.com/watch?v=r1gNfJiFj-s

acquisition and so forth. This thesis aims to contribute to this field by studying how people use gesture to refer to objects.

We produce referring expressions whenever we describe objects in our everyday surroundings. These referring expressions often do not only consist of speech, but also of gesture. We know that there can be variation in referring expressions in speech, but not much is known yet about possible variation in referring expressions in gesture. Can variation in gesture production be related to variation in speech production? And if there is variation in gesture production, how does this impact the listener? In this thesis, we aim to find out the answer to these questions by focusing on the production of repeated references.

This thesis consists of four independent studies, and although there are clear links between the chapters, each chapter can in principle be read on its own. The purpose of this introductory chapter is to provide some background information about gesture and gesture-speech models, and to give an overview of the studies reported in this thesis, including some detail about relevant methodological considerations.

Gesture

Most people know what a gesture is when they see one. However, an exact definition of a gesture is slightly more difficult to give. David McNeill describes gestures as "the movements of the hands and arms that we see when people talk" (1992, p. 1) and states that they are "symbols of action, movement, and space […]" (ibid.). Adam Kendon describes gesture as "visible action when it is used as an utterance or part of an utterance" (2004, p. 7). These definitions include different types of gestures, but they exclude movements that are not related to speech, such as self adaptors (like scratching one's nose, or touching one's hair).

There are several ways in which gestures can be further categorized. One way of categorizing different types of gestures is by using Kendon's continuum, proposed by McNeill (1992):

Gesticulation → Language-like gestures → Pantomimes → Emblems → Sign languages.

On the left of this continuum there is gesticulation, which consists of spontaneous idiosyncratic movements of the hands, such as when a speaker produces a gesture representing the act of bending while saying 'and he *bends it way back*' (McNeill, 1992, p. 12, gesture produced during the italicized speech). Moving to the right, there are language-like gestures, which are gesticulations that are grammatically integrated in speech (McNeill, 2006), for example when a gesture is produced instead of the verb 'throw' in the sentence "and she [...] it down there". Pantomimes are gesticulations that communicate a meaning or even an entire story without the need of any speech. Emblems are culturally specific gesticulations with a fixed form and meaning (Wagner, Malisz, & Kopp, 2014) that can be produced without speech, such as the Dutch emblem for '*lekker*' ('tastes good'), produced by waving one's (left) hand next to one's (left) ear. On the rightmost side of the continuum there are sign languages, which are languages used by deaf communities. As one moves from the left of the continuum to the right, the presence of speech with gesticulation becomes less obligatory, while the gesticulations themselves start to have more linguistic properties (and thus also become more standardised and conventionalised) (McNeill, 1992, 2006). In this thesis we focus on the gesticulations from the left side of Kendon's continuum, which, following general practice, we will henceforth call gestures for short. Gestures, although they are closely related to speech (to be discussed in more detail below), are not conventionalised like other aspects of the linguistic system, and are made up on the spot by a speaker without adhering to linguistic rules (McNeill, 1992). In addition, in this thesis we report on one study on sign language (which will be introduced below).

Although the gestures that we are interested in in this thesis are spontaneous idiosyncratic movements of the hands and arms that at first sight may seem fairly random, it has been found that these gestures are in fact structured in a certain way. The complete gestural movement, that is, all movement between initial and final rest position of the hands and/or arms, consists of several phases (Kendon, 1980, 2004; McNeill, 1992). Together these phases form a gesture phrase. A gesture phrase consists of an optional preparation phase, during which the limb(s) move to the position in which the obligatory stroke phase takes place. The stroke may be followed by a retraction phase, during which the limb(s) return to rest position. A stroke can also be directly preceded or followed by a hold phase, which consists of a temporary hold from movement (McNeill, 2006). The stroke phase is the essential and meaningful phase of a gesture: during the stroke phase, most effort is used, and the stroke phase is the phase in which the (semantic) meaning of the gesture is most clearly expressed (McNeill, 1992, 2006).

Apart from the fact that these gestures are internally structured in similar ways, they can also be grouped with regard to their type. Several groupings of gesture types have been proposed (e.g. Ekman & Friesen, 1969; McNeill, 1992), generally distinguishing between gestures that are semantically related to speech ('imagistic' or 'representational' gestures), and gestures that are not. An often used classification was developed by McNeill (1992), who defined several gesture types (which are often interpreted as mutually exclusive gesture categories, although according to McNeill, 2005; 2006, they should be seen more as dimensions, meaning that one gesture can contain aspects of several gesture types). The four main types of gestures according to McNeill (1992) are iconic, metaphoric, beat and deictic gestures. These different gestures have different semantic functions in the discourse. We will shortly describe each type below.

Iconic gestures are imagistic gestures that have "a close formal relationship to the semantic content of speech" (McNeill, 1992:12). Iconic gestures represent a concrete event or object. For example, Al Gore, in his speech at Tilburg University, produced two iconic gestures when he mentioned someone who was wearing "a straw hat with the price tag still hanging on the hat"; first producing a gesture around his own head that indicated the shape of the hat, followed by an iconic gesture indicating the location of the price tag (see figure 1.1). The role of iconic gestures is often to illustrate or clarify an (aspect of an) object (McNeill, 1992), as is the case in the example in figure 1.1, where a specific aspect of the price tag hanging from the hat (its location) was presented in gesture (and, in this particular case, not in speech).

Metaphoric gestures are imagistic gestures like iconic gestures, but differ from iconic gestures in that they do not represent something concrete, but rather something abstract (McNeill, 1992). An example of a metaphoric gesture was produced by Al Gore in his speech at Tilburg University when he mentioned "the integration of research and learning" while producing a sweeping gesture from left to right during the word 'integration'. This gesture (as is the case for most metaphoric gestures, McNeill, 1985) showed an image of an abstract concept and thereby served to make something abstract (in this case the concept of integration) more concrete.

Figure 1.1. Example of an iconic gesture, produced by Al Gore in November 2010 at Tilburg University. Still shows hand positioned at the end of the stroke phase, arrow indicates path and movement of the hand during the stroke phase.

Beat gestures and deictic gestures are considered to be non-imagistic (Kendon, 2004; McNeill, 1992). Beat gestures (also called 'baton', Ekman & Friesen, 1969) are gestures in which (part of) the hand moves up and down in a simple movement according to the rhythm of speech. Beat gestures do not have a clear semantic relationship with speech but they are often used for indicating which part of an utterance is considered particularly important or relevant (Krahmer & Swerts, 2007), and thus mainly serve a pragmatic purpose, comparable to how pitch accents emphasize certain words or phrases (e.g., Gussenhoven, 2004). An example of a beat gesture was, again, produced by Al Gore in his speech at Tilburg University, when he produced one beat gesture for each research field as he mentioned “economics, law and ethics”.

The last type of gestures as defined by McNeill (1992) are deictic gestures. Deictic gestures are pointing gestures, which can refer to both abstract and concrete objects and can be used whenever someone wants to locate something. A concrete example is when someone says ‘that one’, while pointing to a specific object. Deictic gestures are generally produced with the arm(s) and hand(s), but other parts of the body may also be used, such as the head, or, in some cultures, the lips (Enfield, 2001).

In addition to the types of gesture proposed by McNeill (1992), interactive gestures are also often distinguished. Interactive gestures (Bavelas, Chovil, Lawrie, & Wade, 1992) are pragmatic gestures that help to maintain the flow of conversation. These refer especially to gestures that are typically used when a speaker has word finding difficulties, or when a speaker wants to keep the turn even though she[2] may not be speaking at the time. These gestures do not have a semantic meaning, but they serve a pragmatic role.

Gesture and speech

In the last few decades (dating back to at least Kendon, 1972; McNeill, 1985), it has become generally accepted that there is a close relationship between speech and gesture. Firstly, gesture and speech are arguably related on a semantic and temporal level. This can be seen, for instance, in the gesture stroke synchronising with co-expressive speech (McNeill, 1992, 2006). Also, gestures are produced by all speakers, even congenitally blind speakers who have never seen someone gesture (Iverson & Goldin-Meadow, 1998), suggesting that gesture is an inherent part of speech production. Moreover, gesture and speech develop together in children (see Gullberg, De Bot, & Volterra, 2008, for an overview) and may break down together in disfluency, for example in cases of stuttering (Mayberry & Jaques, 2000) and in patients with aphasia (Mol, Krahmer, & van de Sandt-Koenderman, 2013).

Although there is general agreement that speech and gesture are closely related, the exact details of this relationship between speech and gesture are not so clear, and many studies on gesture have focused on gaining more knowledge about what the relationship between speech and gesture actually looks like. Over the years, several speech-gesture hypotheses and models have been proposed that each show (sometimes subtle) differences in the way in which they consider speech and gesture to be related. A rough distinction can be made between on the one hand hypotheses that assume that gesture facilitates speech, meaning that they consider gesture to be secondary to speech, and on the other hand models that assume that speech and gesture are more equal partners of the same process (Kendon, 2007). Although the goal of this thesis is not to take a

[2] Following common practice, throughout this thesis, 'she' will be used to indicate the speaker, and 'he' will be used to indicate the addressee.

particular stand with regard to these models, a short overview of the existing speech-gesture hypotheses and models can nevertheless serve as useful background knowledge.

Most (but not all) models reflect the close link between speech and gesture to the extent that they are based on Levelt's (1989) 'blueprint for the speaker', a framework of speech production, consisting of three autonomous consecutive processing components, or stages: the conceptualizer, the formulator and the articulator. In the conceptualizer the speaker decides what she wants to say, which results in a "preverbal message". Then, during the formulator stage, the preverbal message is used as input with which the words of the utterance are planned, using lexical retrieval and grammatical encoding, and resulting in a surface form. In the final articulator stage the surface form is phonologically encoded and articulated, resulting in auditory speech. Together, the three stages form the entire speech production process, from the conception of a message up to the production of actual speech. The speech-gesture models based on this blueprint mainly differ with regard to where, and to what extent, the speech and gesture streams interact during the production process.

There are two influential hypotheses about why people gesture that assume, in different ways, that gesture is auxiliary to speech: the Lexical Retrieval Hypothesis (Krauss, Chen, & Gottesman, 2000; Krauss & Hadar, 1999), and the Information Packaging Hypothesis (Alibali, Kita, & Young, 2000; Kita, 2000). The Lexical Retrieval Hypothesis, partly inspired by work by Dobrogaev (1929) and Butterworth and Beattie (1978), proposes that gesture production facilitates lexical retrieval (hence its name). According to this hypothesis, producing a gesture (during speech) will help in the retrieval and generation of the phonological form of an utterance. This means that gesture does not play a role in the speech production process until fairly late, when the surface form of the utterance has to be produced (during the formulation stage). This is in contrast with the Information Packaging Hypothesis (Alibali, et al., 2000; Kita, 2000) which also proposes that gesture plays a facilitative role, but does so at a different, earlier, point during speech production. In the Information Packaging Hypothesis, the idea is that gesture helps in the selection and ordering of imagistic thought for expression in speech. This means that in the Information Packaging Hypothesis, gesture already plays a role during the conceptual planning of an utterance, and facilitates formulation.

Several models have been proposed that suggest that gesture and speech are equal partners of the same production process. Although they all propose that gesture and speech are integral parts of an utterance, they differ in where in the speech production process speech and gesture are related, and to what extent gestures are intended communicatively. Firstly, the aforementioned Lexical Retrieval Hypothesis (Krauss & Hadar, 1999) led to the Process model (Krauss, et al., 2000) which states that speech and gesture production are two independent processes that are related in working memory but do not interact until the formulator stage when the gesture can help retrieve the word. Secondly, the Sketch Model by de Ruiter (2000, 2007) states that a communicative intention underlies the production of a deliberate "coherent multimodal message" (de Ruiter, 2007, p. 25). In this model, there is a communicative intention, which is planned in the conceptualization stage, and followed by two separate but parallel formulation stages; one for speech and one for gesture. Part of the planned information to be communicated may be given via speech, and part via gesture. The idea that there might be a trade-off between information given in speech and information given in gesture was further developed in the trade-off hypothesis (de Ruiter, Bangerter, & Dings, 2012), which claims that when it becomes more difficult to produce speech (for whatever reason), it becomes more likely that a gesture will be produced, to "take over some of the communicative load" (de Ruiter, et al., 2012, p. 233).

Another speech-gesture model in which speech and gesture are considered equal partners is the Interface Model by Kita and Özyürek (2003). In this model, speech and gesture production are two independent processes that collaborate and interact with each other. According to this model, there is an online interplay between imagistic and linguistic thinking during the conceptualization stage. This means that both the underlying imagery that needs to be represented and the (structure of the) language that is spoken are important for gesture production. This model also proposes that the structure of the language can influence the gestures that are produced (Kita & Özyürek, 2003).

Finally, there are some theories not based on Levelt's blueprint of the speaker. McNeill (1992) and McNeill and Duncan (2000) proposed the Growth Point Theory, which assumes that a speaker has an "idea unit", from which an utterance is derived, both in speech and in gesture. In their view, speech and gesture are completely intertwined and cannot be considered separately. The Growth Point Theory proposes that, due to this tight connection between speech and gesture, gesture can provide a window straight into thought and this also means that gestures might be expressed involuntarily.

Hostetter and Alibali (2008, 2010) proposed the Gesture as Simulated Action framework, which is similar to the Growth Point Theory in that they consider speech and gesture to be two inseparable aspects of the same system. According to this framework, gestures are simulated actions in the speaker's mind, or in other words, "gestures emerge from the perceptual and motor simulations that underlie embodied language and mental imagery" (Hostetter & Alibali, 2008, p. 502). The assumption is that language and imagery cause mental simulations, which in turn can cause motor activations. Whether or not the motor activations are executed (meaning that an actual gesture is produced) depends on a specific threshold, the level of which may differ between speakers and situations.

Gesture and reference production

The overall aim of the work presented in this thesis is to understand more about the way in which speech and gesture are related. This is done by studying reference production. Reference production is one of the core aspects of human communication. Referring expressions occur in many situations in daily life, whenever a particular person or object is being described or discussed. Children learn from an early age to use referring expressions in speech (Matthews, Butcher, Lieven, & Tomasello, 2012) and the first type of gesture that children produce, a deictic gesture (Liszkowski, 2005), can be considered a gestural referring expression (the pointing gesture indicating "I want *that*"). Following an initial, more exploratory study on the relation between speech and gesture, asking whether any changes will occur in speech when speakers do not have the gestural modality at their disposal (chapter 2), we study how participants repeatedly refer to relatively complex, but concrete, objects, using referring expressions. These referring expressions might range from, for example, "the large yellow object shaped a bit like a vase", to a much shorter referring expression such as "the vase".

In a repeated reference, the information in the reference is not new anymore but given, as compared to an initial reference. This givenness may be the reason why, when the same object is repeatedly referred to, this repeated referring expression is usually not produced in exactly the same way as the initial referring expression. Repeated references are generally reduced, at least with regard to the number of words (Brennan & Clark, 1996; Clark & Wilkes-Gibbs, 1986), and often also acoustically (Aylett & Turk, 2004; Bard, et al., 2000; Fowler, 1988) (discussed in more detail in chapters 3, 4 and 5). Some studies have also shown that repeated references are accompanied by fewer gestures (e.g. de Ruiter, et al., 2012; Levy & McNeill, 1992), but much remains unknown about the production of gestures in repeated references. Also, nothing is known about sign production in sign language (discussed below) during repeated references. In this thesis we therefore propose the following specific research questions (discussed in more detail below): Do gestures change (and if so, how) in repeated references during successful communication (chapter 3)? Do signs in sign language change (and if so, how) in repeated references during successful communication (chapter 4)? Do gestures change (and if so, how) in repeated references that are produced when communication is unsuccessful (chapter 5)?

Methodology

Before describing the studies of this thesis in some more detail, there are several recurring methodological aspects that we will briefly introduce. Firstly, in this thesis we use both production and perception studies. In each empirical chapter of this thesis we report on a production study and on one or more perception studies. The assumption here is that by studying both production and perception we can separate what the speaker does (production, in speech, gesture or sign) from what an addressee actually picks up (perception). The production experiments take the form of instruction giving (chapter 2) or of picture description tasks (chapters 3, 4 and 5), while in the perception experiments participants are asked to judge either sound fragments, or aspects of the form or interpretation of a gesture. In general, the production experiments can give us information about what a speaker does, but cannot tell us to what extent this behaviour is (also) relevant for the addressee. The perception studies might help in this regard. More detail about the reasons for conducting both production and perception studies can be found in the four empirical chapters (chapters 2, 3, 4, 5).

Secondly, in chapters 3, 4 and 5 of this thesis we address our research questions by mainly focusing on two aspects (of either gesture or sign): frequency, and form. Especially in gesture research, gesture frequency, referring to the number of gestures, is an often-used variable (e.g. de Ruiter, et al., 2012; Galati & Brennan, 2014; Holler, Tutton, & Wilkin, 2011). However, there are several ways in which the number of gestures can be counted. These range from fairly general measures such as "how many gestures occur in my dataset" to more precise but still differing measures such as "how many gestures occur per word in my dataset" and "how many gestures occur per semantic attribute[3] in my dataset". In this thesis, several measures of gesture frequency are used and compared. The different measures of gesture frequency can be interpreted in different ways, and can inform us about different aspects of the relationship between speech and gesture. These different measures are also informative for the way in which one thinks that speech and gesture are related. This will be discussed in more detail in chapter 3.

Form (of either gesture or sign) is an important aspect to study, since it can provide us with information, for example about the relation between speech and gesture, which we might miss if we were only to study frequency. After all, it may be the case that the same number of gestures is produced but that, depending on the (linguistic) context, these gestures differ in *how* they are produced. However, depending on the research question, there are many different aspects of gesture form that can be relevant and can be analysed, and consequently, methods across studies often differ. For example, one could annotate aspects of gesture form such as gesture size (e.g. Galati & Brennan, 2014), gesture position (e.g. Gullberg, 2006), gesture duration (e.g. Krauss, 1998), gesture precision (e.g. Gerwing & Bavelas, 2004), the gesture's mode of representation (e.g. Müller, 1998), the use of gesture space (Holler, et al., 2011) and so forth. Methods of analysis often differ across studies, even when the same aspect of gesture form, such as gesture precision, is analysed (cf. Galati & Brennan, 2014; Gerwing & Bavelas, 2004) . These differences between studies mean that it can be difficult to relate results from different studies to each other. We will discuss this in more detail in chapter 3. In this thesis, the two chapters that study gesture form (chapter 3 and 5) use the same methodology, allowing a direct comparison between these two studies. This is discussed at the end of chapter 5.

[3] A semantic attribute is a characteristic of an object, which may be described in several words. For example, the phrase "The man with the beard" consists of 5 words, and 2 attributes ('gender', described as "the man" being the first attribute, and 'facial hair', described as "with the beard" being the second attribute).

Thirdly, in three of the empirical chapters (chapters 2, 3, and 5), we include visibility as a factor in the design of our studies, by placing a large opaque screen between some of the speakers and addressees. Following previous gesture studies (see Bavelas & Healing, 2013, for a selected overview and discussion), this was done to study whether and when changes in gesture production are more speaker- or more addressee-oriented. The idea is that although we cannot distinguish to what extent gestures are produced for the speaker or for the addressee in face to face communication, when communication is visually restricted, we can separate (aspects of) gestures that are produced for the speaker from (aspects of) gestures that are produced for the addressee. As Alibali et al. (2001, p. 169) state: "if speakers produce gestures in order to aid listeners' comprehension, they should produce fewer gestures when their listeners are unable to see those gestures". Importantly, visibility can affect some types of gesture but not others (e.g. de Ruiter, et al., 2012), and can not only affect gesture rate, but also gesture form (see e.g. Bavelas, Gerwing, Sutton, & Prevost, 2008). The assumption is that the (aspects of the) gestures that are produced when there is no mutual visibility are produced for the speaker and serve cognitive needs (Kita, 2000; Krauss, 1998). Of course, there are alternative explanations, for example that any gestures that are still produced when the speaker and the addressee cannot see each other are produced out of habit (Cohen & Harrison, 1973). The role of visibility in our studies is discussed in more detail in each relevant chapter, especially in chapters 3 and 5.

Fourthly, in all empirical chapters in this thesis, we consistently focus on the individual as our 'unit of analysis' (Bavelas & Healing, 2013). This means that although we report on experiments in which two people (one speaker and one addressee) took part at the same time, we focus our analyses on the speaker only. In all studies an addressee was present so that there was a practical goal to the experiment, e.g., the speaker produced descriptions so that the addressee could determine which object was described. Also, we chose to include an addressee because mere addressee presence can have a positive effect on gesture production. However, there was little interaction between speaker and addressee, and in some studies (chapters 3 and 5) extended interaction between speaker and addressee was explicitly discouraged. This was done so that data from different stimuli and different participants was as comparable as possible, as the implicit assumption was that the addressee could cause error variance (see Bavelas & Healing, 2013, for discussion). This issue is further touched upon in the discussions of chapters 3 and 5.

Current studies

Having introduced some of the methodological aspects of this thesis, we can now introduce the research questions and the empirical chapters in some more detail.

In chapter 2, we report on our first, explorative, study, entitled 'Does our speech change when we cannot gesture?' The question here is, as the title suggests, what happens in speech when speakers do not have the gestural modality at their disposal. More specifically, this study sets out to investigate the claim by Dobrogaev (1929) that speech becomes less fluent and more monotonous when speakers cannot gesture. Finding out whether speech changes when people cannot gesture can inform us about how close the relationship between speech and gesture is, since a very close relationship would suggest that if one changes, the other is likely to change also. In this chapter we study our research question by conducting a production experiment in which speakers have to instruct addressees how to tie a tie. Participants are prevented from gesturing during half of the experiment and instructions have to be given repeatedly, and, for some participants, without visibility of the addressee. In an additional perception experiment we study whether naïve listeners can hear whether someone gestures or not.

The results of chapter 2 led to a set of studies which are reported in chapters 3, 4 and 5. In chapter 3, entitled 'Reduction in gesture during the production of repeated references', the question is whether gestures change (and if so, how) in repeated references. In this study we report on a production experiment in which participants had to repeatedly describe complex objects, and on two perception experiments in which participants had to judge or interpret gestures given in these (repeated) descriptions. In the production experiment of this study, one group of participants did not have visibility of their addressee. We focus on possible reduction in several measures of gesture frequency and of gesture form.

In our third and fourth study (chapters 4 and 5) we study whether the effects that were found in chapter 3 can be generalised to other contexts. First, we study sign language, a context in which the visual modality is the main modality used in communication. This is in contrast with using the visual modality for gesture production during speech, when, although gesture and speech work together in creating a message, the visual modality is not (always) essential for communication. Second, we study a context of miscommunication, to see whether reduction in repeated references also occurs in that setting. We discuss both studies in some more detail below.

In our third study, entitled ‘Do repeated references result in sign reduction?’ the question is whether signs in sign language change (and if so, how,) in repeated references. In this chapter we study whether our findings on speech and gesture from chapter 3 can be generalized to signs produced in sign language. We studied speakers of Sign language of the Netherlands. Sign language of the Netherlands (NGT) is one of many sign languages of the world used by deaf communities. Sign languages are fully fledged languages, with their own linguistic structures, including morphological patterns, phonological rules, etc. (Liddell, 2003). Signs in sign languages can be defined and distinguished by three basic aspects (Stokoe, 1960): location, hand shape, and movement. A change in one of these aspects can change the entire meaning of a sign. Signs in sign language are distinct from gestures that are used in spoken languages. For example, in sign languages, as in other languages, smaller parts are combined to create larger wholes (McNeill, 1992). Several aspects or grammatical features can be combined to form signs and sentences in sign language, just as (parts or forms of) words can be combined in a spoken language to create a specific meaning (a linguistic property). This is not the case for gestures, where smaller gestures cannot be combined to create a larger gesture or a specific meaning. Another major difference between signs and gestures is that in languages (so also in sign languages), units have a standard form and meaning. Speakers of NGT, if they want to produce a certain meaning, have to produce a certain sign in a certain way if they want to be understood by other speakers of NGT. Gestures however, do not follow any standards of form or meaning, and are created on the fly. These differences between signs and gestures are also indicated in the earlier mentioned Kendon’s continuum (McNeill, 1992), showing that gestures do not have linguistic properties, but sign languages do. An important side note to make, however, is that speakers of sign language are not restricted to one side of Kendon’s continuum, as they may also produce emblems, pantomimes, and gestural elements (Liddell, 2003) in their communication.

In chapter 4 we study whether signs in sign language change (and if so, how,) when information is repeated by conducting identically designed experiments as in chapter 3, but with the main difference that here our participants are speakers of NGT (and that naturally there is no visibility condition, as this would make communication very complicated for deaf speakers). In this study we, again, conduct both a production and a perception experiment, in which we study whether there is reduction in repeated references in sign language. We focus on aspects of sign frequency and sign form, and on how these can be adapted in such a way that speakers use language efficiently.

In our fourth and final study, 'On what happens in gesture when communication is unsuccessful' (chapter 5), we ask whether gestures change (and if so, how,) in repeated references that are produced when communication is unsuccessful. In this study we changed the context compared to the studies reported in chapter 3 and 4 in such a way that speakers, again, had to give repeated descriptions, but not because stimuli simply happened to reoccur, as was the case in chapters 3 and 4, but because a previous description was not considered to be sufficient or correct by the addressee. In other words, in this study we study whether repetition always affects speech and gesture in the same way or whether it matters in what exact discourse context this repetition takes place. The idea is that producing reduced repeated references (as found in chapters 3 and 4) is not beneficial for the communicative situation when previous references have not been considered adequate. Also in this study, we report on both a production and a perception experiment. We focus on gesture frequency and on gesture form and on what these can tell us about speakers' effort in producing repeated references.

Final remarks

We end this introductory chapter with some final remarks.

The four empirical studies presented in this thesis (chapters 2, 3, 4 and 5) have all been published in peer-reviewed scientific journals. The author of this thesis was the main researcher in all empirical studies. The chapters are self contained texts, and all have their own abstract, introduction and discussion section. Due to the fact that the studies are self contained, some textual overlap between the chapters and between the chapters and this introduction chapter was unavoidable. The final chapter of this thesis contains a general discussion and conclusion. The studies reported in this thesis have been conducted in a timeframe of several years. Naturally, this means that some changes in insight about theory, but also about methodology, have occurred. In addition, due to differing requests from reviewers and journal editors, some (minor) differences in annotation and analysis as well as in phrasing and presentation of the results may occur.

In all studies reported in this thesis, we investigate different aspects of 'talking hands'. The metaphor in the title of this thesis can be applied to each study in a different way. In the first study we see what happens when the hands cannot do the talking, whereas in the second study we see what happens when they can. In the third study we look at what happens when the hands have to do all the talking and in the fourth and final study we see what happens when the hands are talking but not heard.

2

Does our speech change when we cannot gesture?

Abstract

Do people speak differently when they cannot use their hands? Previous studies have suggested that speech becomes less fluent and more monotonous when speakers cannot gesture, but the evidence for this claim remains inconclusive. The present study attempts to find support for this claim in a production experiment in which speakers had to give addressees instructions on how to tie a tie; participants had to perform half of this task while sitting on their hands. Other factors that influence the ease of communication, such as mutual visibility and previous experience, were also taken into account. No evidence was found to support the claim that the inability to gesture affects speech fluency or monotony. An additional perception task showed that people were also not able to hear whether someone gestures or not.

This chapter is based on:

Hoetjes, M., Krahmer, E. & Swerts, M. (2014). Does our speech change when we cannot gesture? *Speech Communication*, 57, 257-267.

Introduction

Human communication is often studied as a unimodal phenomenon. However, when we look at a pair of speakers we can quickly see that human communication generally consists of more than the mere exchange of spoken words. Many people have noted this and have been studying the multimodal aspects of communication such as gesture (e.g., Kendon, 2004; McNeill, 1992). Studying multimodal aspects of communication is not a recent thing, with Dobrogaev stating back in the 1920s that human speech consists of three inseparable elements, namely sound, facial expressions and gestures. According to Dobrogaev it is unnatural to completely leave out or suppress one of these three aspects and doing so will always affect the other two aspects of speech (Chown, 2008). However, by suppressing one of these inseparable elements, we can find out more about the relationship between all multimodal elements of communication, such as speech and gesture. In fact, Dobrogaev studied the effect of not being able to gesture on speech (Dobrogaev, 1929) by restraining people's movements and seeing whether any changes in speech occurred. He found that speakers' vocabulary size and fluency decreases when people cannot gesture. This study is often cited by gesture researchers, for example by Kendon (1980), Krahmer and Swerts (2007), McClave (1998), Morsella and Krauss (2005) and Rauscher, Krauss and Chen (1996), but unfortunately it is very difficult to track down, it is not available in English and therefore its exact details are unclear. Other studies, however, have since done similar things, with people looking at the effect of (not being able to) gesture on language production and on acoustics.

Influence of (not being able to) gesture on language production

There have been several studies looking at the effect of not being able to gesture on speech, with different findings. For example, Hostetter, Alibali and Kita (2007) asked participants to complete several motor tasks, with half of the participants being unable to gesture. They found some small effects of the inability to gesture, in particular that speakers use different, less rich, verbs and are more likely to begin their speech with "and" when they cannot use their hands compared to when they can move their hands while speaking. In a study on gesture prohibition in children, it was found that words could be retrieved more easily and more tip of the tongue states could be resolved when the children were able to gesture (Pine, Bird, & Kirk, 2007). Work by Beattie and

Coughlan (1999) however, found that the ability to gesture did not help resolve tip of the tongue states.

There have also been some studies on gesture prohibition that focused on spatial language. It has been found that speakers are more likely to use spatial language when they can gesture compared to when they cannot gesture (Emmorey & Casey, 2001). Graham and Heywood (1975), on the other hand, found that when speakers are unable to gesture, they use more phrases to describe spatial relations. This increase in use of spatial phrases might be a compensation for not being able to use gesture (de Ruiter, 2006).

According to the Lexical Retrieval Hypothesis, producing a gesture facilitates formulating speech (Alibali, et al., 2000; Krauss, 1998; Krauss & Hadar, 1999; Rauscher, et al., 1996), and not being able to gesture has been shown to increase disfluencies (Finlayson, Forrest, Lickley, & Mackenzie Beck, 2003). In a study by Rauscher, Krauss and Chen (1996) it was found that when speakers cannot gesture, spatial speech content becomes less fluent and speakers use more (nonjuncture) filled pauses. However, a study by Rimé, Schiaratura, Hupet and Ghysselinckx (1984) found no effect of being unable to gesture on the number of filled pauses.

Overall, there seems to be some evidence that not being able to gesture has an effect on spatial language production (as one would expect considering that gestures are prevalent in spatial language, e.g. Rauscher, et al., 1996), but other findings remain inconclusive and are sometimes difficult to interpret.

*Influence of (not being able to) gesture on acoustic*s

Apart from his claims on vocabulary size and fluency, the study by Dobrogaev (1929) is often associated with the finding that people's speech becomes more monotonous when they are immobilized. This has, as far as we know, never been replicated, but several other studies have looked at some acoustic aspects of the direct influence of gestures on speech. For example, it has been found that producing a facial gesture such as an eyebrow movement often co-occurs with a rise in pitch (F0) (Cavé, et al., 1996) and that manual gestural movement also often co-occurs with pitch movement (Flecha-García, 2010; McClave, 1998), also described in the so-called "metaphor of up and down" (Bolinger, 1983). Bernardis and Gentilucci (2006) found a similar result, namely that producing a gesture enhances the voice spectrum, or, more specifically, that producing a gesture at the same time and with the same meaning as a specific word (such as the Italian word 'ciao' accompanied by a waving gesture) leads to an increase in the word's second formant (F2). Also on an acoustic level, Krahmer and Swerts (2007) found that producing a beat gesture has an influence on the duration and on the higher formants (F2 and F3) of the co-occurring speech. In a perception study, Krahmer and Swerts (2004) found that listeners also prefer it when gestures (in this case eyebrow gestures) and pitch accents co-occur. The above mentioned studies suggest that there is also a relationship between gesture and speech on an acoustic level. However, we are not aware of any studies that looked at the effect of not being able to gesture on acoustics in general and on pitch range specifically.

Other factors influencing gesture production

In the present study we want to look at the effect of not being able to gesture on several aspects of speech production. It has been assumed, for example in the above mentioned Lexical Retrieval Hypothesis, that there is a link between gestures and cognitive load. Arguably, not being able to gesture can be seen as an instance of an increased cognitive load for the speaker. We can then hypothesise that not being able to gesture affects speech even more in communicatively difficult situations where speakers also have to deal with an additional increased cognitive load, because of the context or because of the topic. An increased cognitive load due to context could occur when people cannot see each other when they interact. An increased cognitive load due to topic could occur when people have to do a task for the first time, compared to a decreased cognitive load when speakers have become more experienced in that task. We aim to take both these aspects of cognitive load into account in order to compare and relate the cognitively and communicatively difficult situation when people have to sit on their hands to other communicatively difficult situations, namely when there is no mutual visibility and during tasks with differing complexity, in this case when participants are more or less experienced (due to the number of attempts).

In fact, both mutual visibility and topic complexity have been shown to influence gesture production. Previous studies (Alibali, et al., 2001; Bavelas, et al., 2008; Emmorey & Casey, 2001; Gullberg, 2006; Mol, Krahmer, Maes, & Swerts, 2009) have found that speakers still gesture when they cannot see their addressee, although the nature of the gestures changes, with gestures becoming fewer and smaller (see chapter 3 for further discussion). Also, a study by Clark and Krych (2004) found that mutual visibility leads to more gesture production and helps speakers do a task more quickly.

Several studies suggest that there can be an influence of topic complexity on the production of gestures. It has been argued that gestures facilitate lexical access (Krauss & Hadar, 1999; Rauscher, et al., 1996) and are thus at least sometimes produced for the speaker herself. More complex tasks and a larger cognitive load will thus lead to more gestures to help the speaker. On the other hand, research has also suggested that gestures can be produced for the addressee and thus serve a communicative purpose (Alibali, et al., 2001; Özyürek, 2002). In this case, more complex tasks and a larger cognitive load will also lead to more gesture production by the speaker, but with the purpose to help the addressee understand the message.

Summary of previous research

Previous research, in short, has acknowledged that there might be a direct influence of gestures on language production and acoustic aspects of speech and that mutual visibility and topic complexity may play a role, but many of these studies have had some drawbacks. Unfortunately, the details of Dobrogaev’s (1929) intriguing paper cannot be recovered, and other studies either found very small effects of being unable to gesture on speech (e.g. Hostetter, et al., 2007), only focused on one particular aspect of speech (e.g. Emmorey & Casey, 2001) or used an artificial setting (e.g. Krahmer & Swerts, 2007). This means that many aspects of the direct influence of gestures on speech remain unknown.

Current study

In the present study, the goal is to answer the research question whether speech changes when people cannot gesture, which we address using a new experimental paradigm in which participants instruct others on how to tie a tie knot. The previous claims as discussed above are tested by comparing speech in an unconstrained condition in which subjects are free to move their hands compared to a control condition in which they have to sit on their hands. Two other aspects of cognitive load, mutual visibility and topic complexity (expressed in the number of attempts), are also taken into account.

We conduct a production experiment and a perception experiment. The production experiment takes place in the form of a tie-knotting instructional task, which combines natural speech with a setting in which it can be expected that speakers will gesture. The task enables the manipulation of the ability to gesture, mutual visibility and the number of attempts. We will look at the number of gestures people produce, the time people need to instruct, the number of words they use, the speech rate, the number of filled pauses used, and the acoustics of their speech, all across conditions with or without the ability to gesture, with or without mutual visibility and with varying number of attempts.

We expect that not being able to gesture will make the task more difficult for the participants, and that this will become apparent in the dependent variables mentioned above. Following previous research (e.g. Alibali et al., 2001; Bavelas et al., 2008; Emmorey & Casey, 2001; Gullberg, 2006; Pine et al., 2007), we expect that the number of gestures produced by the director is influenced by a communicatively difficult situation (due to lack of ability to gesture or lack of mutual visibility), naturally with fewer gestures being produced when there is no ability to gesture, but also with fewer gestures being produced when the director and the matcher cannot see each other. We also expect that directors' speech will change, with instructions taking longer, measured either in time or in number of words, and speech rate becoming lower, when the communicative situation is more difficult than it normally is, foremost because of the inability to gesture, but also because of lack of mutual visibility, or because of the number of attempts (where the first attempt is considered to be more complex than the second or third attempt and the second attempt is considered to be more complex than the third attempt). Since we assume that the number of filled pauses indicates the level of processing difficulty and that they can also be seen as a measure of fluency, we expect that a more difficult communicative situation leads to more processing difficulty and more filled pauses. Considering previous findings on acoustics and gesture, (Bernardis & Gentilucci, 2006; Krahmer & Swerts, 2007), we assume that speech will be more monotonous when speakers cannot gesture, and that this will be apparent by a smaller pitch range and a lower intensity when people are unable to gesture.

In addition to the production experiment we conduct a perception experiment, where participants are presented with pairs of sound fragments from the production experiment and are asked to choose in which sound fragment the speaker was gesturing. The perception task on the selected audio recordings is conducted to see whether people can *hear* when somebody is gesturing.

Considering previous research, we expect that sound fragments where the speaker could not gesture will be different from sound fragments where the speaker could gesture and the expectation is that participants will be able to hear this difference.

Production experiment

Participants

Thirty eight pairs of native speakers of Dutch participated in the experiment (25 male participants, 51 female participants), half of them as instruction givers (“directors”), half as instruction followers (“matchers”). Participants took part in random pairs (these could be male, female, or mixed pairs). The participants were first year university students (*M* = 20 years old, range 17-32 years old). Participants took part in the experiment as partial fulfilment of course credits.

Stimuli

Directors watched video clips on a laptop, containing instructions on how to tie two different (but roughly equally complicated) types of tie knot. To control for topic complexity, each clip with one type of tie knot instruction was presented and had to be instructed three times (hence described as the within subjects factor ‘number of attempts’) before the other video clip was presented three times. This was done because the assumption was that instructing a tie knot for the first time causes a larger cognitive load than instructing it for the third time (as things tend to get easier with practice). Each video clip, containing instructions for a different tie knot, was cut into six fragments. Each fragment contained a short (maximally 10 or 15 seconds) instructional step for the knotting of a tie. The video clips contained the upper body of a person who slowly knotted a tie without speaking or using facial expressions. Each fragment was accompanied by a small number of key phrases, such as ‘...wide...under...thin...’, ‘tight’ or ‘...through...loop...’. The key phrases were printed in Dutch and presented above the video clips. These key phrases were added to make the task a little bit easier for the participants, and to make sure that instructions from different directors were comparable. A still from one of the clips’ fragments can be seen in figure 2.1.

Figure 2.1. Still of the beginning of a fragment of one of the stimulus clips, in this case accompanied by the phrases ‘behind’ and ‘up’.

Procedure

The participants entered the lab in pairs and were randomly allocated the role of director or matcher. The two participants sat down in seats that were positioned opposite each other. The seat of the director did not contain any armrests. Participants were asked to sign a consent form, were given instructions about the experiment on paper and the possibility to ask for clarifications, after which the experiment would start.

Directors then watched all six video fragments of one tie knot on the laptop and gave instructions to the matcher how to tie an actual tie that the matcher was holding after watching each fragment. The directors were only allowed to watch each video fragment once and the matcher could not see the screen of the laptop. This procedure was repeated three times for the same tie knot, after which the fragments for the other tie knot were shown three times. Matchers thus had to tie the same tie knot on themselves three times followed by the other type of tie knot which also had to be tied three times. The order in which directors were presented with the video clips of the two different tie knots was counterbalanced over participants. Half of the directors had to sit on their hands for the first half of the experiment, whereas the other half of the directors had to sit on their hands during the second half of the experiment. This means that all directors conducted half of the task, instructing one of the two tie knots, while sitting on their hands. Getting directors to sit on their hands was achieved simply by asking directors to sit on their hands at the beginning or halfway through the experiment. If directors were asked to sit on their hands at the beginning of the experiment they were told they were free to move their hands halfway through the experiment. No information was given about why sitting on their hands was necessary. For half of all participant pairs, an opaque screen was placed in between the director and the matcher so as to manipulate (lack of) mutual visibility. Examples of the experimental setup can be seen in figure 2.2.

Figure 2.2. Examples of experimental setup. In both images, the director is visible on the right; the matcher is on the left (only knees visible). On the right hand side the setup with opaque screen between director and matcher is shown.

The experimenter was in the lab during the experiment and, for the entire duration of the experiment, controlled the laptop on which the video fragments were shown. This was due to the fact that the directors were unable to control the laptop while they were sitting on their hands. The experimenter, using a remote control, switched to the next video fragment when it was clear that the director had said all there was to say and the matcher had understood the instructions and tied (part of) the tie knot accordingly. The proceedings of the experiment were videotaped (both audio and video). The director was filmed from the left side, as in figure 2.2. The audio recorder was placed on the table, to the right of the director, as can be seen in figure 2.2. After the experiment, participants filled out a short questionnaire, asking, among other things, about their experience with tie knotting (nobody had any significant experience, and participants found both tie knots equally difficult to instruct) and whether they knew the person they had just done the experiment with (most people, across conditions, did). Finally, the participants were debriefed about the experiment. The entire experiment took about 30 minutes.

Design

The experiment had a mixed design (2x2x3), with one between subjects factor, mutual visibility (levels: screen, no screen) and two within subject factors, ability to gesture (levels: able, unable) and number of attempts (levels: 1st, 2nd, 3rd attempt). Half of the participant pairs had a screen between them for the entire duration of the experiment and the other half were able to see each other during the experiment (mutual visibility). All directors had to sit on their hands (ability to gesture) either during the first half of the experiment or during the second half of the experiment (this order was counterbalanced). The ability to gesture was designed as a within-subject factor because previous gesture research has found that there may be large individual differences in gesture production (e.g. Chu & Kita, 2007). All directors had to instruct the two different tie knots three times (number of attempts). The order in which the tie knots were presented was counterbalanced. This design means that each director would instruct one tie knot three times while sitting on his/her hands and the other tie knot three times while being able to gesture.

Data analysis

Video and audio data from the director was recorded. The speech from the video data was transcribed orthographically and the gestures produced during all first attempts were annotated using a multimodal annotation programme, ELAN (Wittenburg, Brugman, Russel, Klassmann, & Sloetjes, 2006). The audio data was used for the acoustic analyses and for the perception experiment. We conducted analyses for several dependent measures.

Firstly we looked at the number of gestures that were produced by the director. The gesture analysis was based on a subset of the data. For one third of all the data (each director's first attempt for each tie knot) we selected all the gestures that were produced. All speech-accompanying hand gestures were counted, leaving out possible head and shoulder gestures and all gestures that were not related to speech (e.g. self-grooming gestures). A gesture was identified as such following Kendon's (1980) definition of a gesture phrase, where a gesture consists of at least a stroke. For the number of gestures, the obvious assumption is that people will gesture less when they are prevented from doing so. The question is, however, to what extent the gesture production is also influenced by one of the other aspects of cognitive load, mutual visibility.

Secondly, we analysed the directors' speech, in duration in seconds, in number of words and in speech rate. The assumption is that these aspects of speech serve as a measure for speech fluency. We measured speech duration in time (in seconds) between the start of one video clip instruction and the start of the following video clip instruction. For the speech duration in number of words, all of the directors' instructions were transcribed orthographically. The transcriptions were divided per video clip instruction, leading to 36 transcriptions (2 tie knots x 3 attempts x 6 fragments) per participant. The mean number of words for each of these instructions was counted, including filled pauses (e.g., 'uhm') and comments about the experiment itself (e.g., 'can I see the clip again?'). Speech rate was defined as the number of words that were produced per second. The main question here is whether the inability to gesture makes it more difficult for directors to instruct the matcher to the extent that the instructions differ in length, in number of words, or in speech rate.

The use of filled pauses in the director's speech was also analysed. On the basis of previous literature (e.g., Rauscher, et al., 1996) we assume that filled pauses are a measure of speech fluency, with less fluent speech containing more filled pauses than more fluent speech. From the transcribed directors' instructions we counted the number of filled pauses (i.e. the Dutch "uh" and "uhm") across conditions. We divided this number by the number of words used to get a rate of filled pauses. This was done in order to factor out any effects due to a change in the number of words used.

For the acoustic analyses, a subset of the audio data was used which was also used for the perception experiment (as described below). The sound pair recordings were analysed using Praat software (Boersma & Weenink, 2010). The minimum and maximum pitch, the mean pitch and pitch range and the mean intensity of each sound fragment were analysed. These aspects were taken into account because previous research (e.g., Dobrogaev, 1929) has suggested that speech becomes more monotonous when speakers cannot gesture. For the acoustic analyses we looked at whether there was an effect of the ability to gesture, and did not take mutual visibility or the number of attempts into account.

For the subset of the data for the gesture analyses (the first attempt at describing each tie knot), we analysed whether there was an effect of ability to gesture, or an effect of mutual visibility on the number of gestures that were produced. For the speech analyses (time, number of words, speech rate and filled pauses) we analysed whether there was an effect of ability to gesture, an effect of mutual visibility or an effect of number of attempts. For the subset of the data for the acoustic analyses we analysed whether there was an effect of the ability to gesture. Unless noted otherwise, all tests for significance were conducted with repeated measures ANOVA. We conducted Bonferroni post hoc tests where applicable. All significant main effects and interactions will be discussed.

Results

Table 2.1 (see below) shows an overview of the results of the production experiment. All the dependent variables are shown as a function of the ability to gesture. Below we discuss each of the variables in more detail.

Number of gestures We found an unsurprising main effect of ability to gesture on the mean number of gestures produced by the director (F (1, 36) = 26.8, $p < .001$, $\eta_p^2 = .427$), showing that the experimental manipulation worked. There was no effect of mutual visibility on the number of gestures (see table 2.2). Noteworthy however (as can be seen in more detail in table 2.2), is the fact that directors do still gesture sometimes when they have to sit on their hands ("slips of the hand") and that directors still gesture frequently when there is a screen between themselves and the matcher. Furthermore, the large standard deviations in table 2.2 show that there are large individual differences with regard to the number of gestures that participants produce.

Speech Duration in Time The mean speech duration of all fragments was 31 seconds (SD = 13.7). There was no effect of ability to gesture on speech duration in time (see table 2.1), nor was there an effect of mutual visibility. There was, however, a significant effect of the number of attempts, F (2, 72) = 23.38, $p < .001$, $\eta_p^2 = .394$ (see table 2.1), with people getting quicker in instructing a tie knot when they have done so before. Bonferroni post hoc tests showed that all three attempts differed significantly from each other, $p < .05$.

Speech Duration in Words No effects of ability to gesture or mutual visibility on the number of words produced by the director were found. However, there was a significant effect of the number of attempts (for the means, see table 2.1), $F(2, 72) = 9.06$, $p < .001$, $\eta_p^2 = .201$. Bonferroni post hoc analysis showed that significantly fewer words were used in the third attempt than in the first attempt ($p < .001$). This shows the same picture as for the speech duration in time, in that people need fewer words in instructing a tie knot when they have done so before.

Speech Rate The mean speech rate for all fragments was 1.3 words per second ($SD = .42$). There were no main effects of the ability to gesture, of the number of attempts or of mutual visibility on the speech rate (see table 2.1). There were also no interaction effects.

Filled pauses No main effects of ability to gesture or mutual visibility on the rate of filled pauses produced by the director were found. However, there was a significant effect of the number of attempts. Significantly fewer filled pauses (for the means, see table 2.1) were used to instruct each following attempt, $F(2, 72) = 19.76$, $p < .05$, $\eta_p^2 = .354$, showing that the rate of filled pauses decreases once people have instructed a tie knot before (all three attempts differed significantly from each other, $p < .05$). There was also an interaction effect between the ability to gesture and the number of attempts on the rate of filled pauses, $F(2, 72) = 3.27$, $p = .044$. For the first attempt the inability to gesture led to a decrease in the rate of filled pauses, whereas for the second and third attempt the inability to gesture led to an increase in the rate of filled pauses (see table 2.1).

Acoustic analyses We found no significant effect of the ability to gesture on any of the dependent acoustic measures (for the means, see table 2.1). Pitch range was not affected by the inability to gesture, which means that speech did not become more monotonous when people could not gesture compared to when they could (and did) gesture.

Table 2.1. Overview of the number of gestures, duration, number of words, speech rate and number of filled pauses, for the first, second and third attempt; and acoustic measurements (maximum, minimum and mean pitch, pitch range (Hz) and intensity), as a function of ability to gesture.

	Able to gesture (*SD*)	Not able to gesture (*SD*)	Mean total (*SD*)
Gestures*	12.68 (*13.9*)	.66 (*2.3*)	6.67 (*8.1*)
Duration attempt 1	36.2 *(16.1)*	35.1 *(11.9)*	35.6 *(14.0)*
Duration attempt 2	29.8 (*12.6*)	30.4 (*14.6*)	30.1 (*13.6*)
Duration attempt 3	25.4 (*11.6*)	29.0 (*15.6*)	27.2 (*13.6*)
Duration all attempts	30.5 *(13.4)*	31.5 *(14.0)*	31.0 *(13.7)*
Words attempt 1	46.3 (*28.2*)	46.8 (*24.5*)	46.5 (*26.3*)
Words attempt 2	41.2 (*24.9*)	43.4 (*30.6*)	42.3 (*27.7*)
Words attempt 3	34.3 (*19.2*)	40.3 (*26.9*)	37.3 (*23.0*)
Words all attempts	40.6 *(24.1)*	43.5 *(27.3)*	42.0 *(25.7)*
Speech rate attempt 1	1.2 *(.35)*	1.3 *(.43)*	1.3 *(.39)*
Speech rate attempt 2	1.3 *(.42)*	1.3 *(.45)*	1.3 *(.43)*
Speech rate attempt 3	1.3 *(.44)*	1.4 *(.47)*	1.3 *(.45)*
Speech rate all attempts	1.3 *(.40)*	1.3 *(.45)*	1.3 *(.42)*
Filled pauses attempt 1	.034 (*.017*)	.030 (*.018*)	.032 (*.017*)
Filled pauses attempt 2	.022 (*.021*)	.029 (*.020*)	.025 (*.020*)
Filled pauses attempt 3	.019 (*.019*)	.021 (*.018*)	.020 (*.018*)
Filled pauses all attempt	.025 *(.019)*	.027 *(.019)*	.026 *(.019)*
Max Pitch (Hz)	248.5 *(83)*	251.65 *(93.5)*	250 *(88.2)*
Min Pitch (Hz)	136.5 *(47)*	138.75 *(60)*	137.62 *(53.5)*
Mean Pitch (Hz)	192.5 *(65)*	195.2 *(76.7)*	193.85 *(70.8)*
Mean Pitch Range (Hz)	112 *(77)*	112.9 *(67)*	112.45 *(72)*
Mean Intensity (dB)	65.40 *(5.9)*	65.95 *(6.2)*	65.67 *(6.0)*

For all dependent variables, $\alpha = .05$. No significant effect of ability to gesture on any of the dependent variables, except *: $F(1, 36) = 26.8$, $p < .001$.

Table 2.2. Mean number of gestures as a function of ability to gesture and mutual visibility.

	Screen (SD)	No Screen (SD)	Mean total (SD)
Able to gesture	10.53 (13.18)	14.84 (14.65)	12.68 (13.90)
Unable to gesture	1.05 (3.22)	.26 (.56)	.66 (1.89)
Mean total	5.79 (8.20)	7.55 (7.60)	6.67 (7.9)

Perception experiment

To see whether a possible change in acoustics due to the inability to gesture can be perceived by listeners, we conducted a perception test on a selection of the data from the production experiment.

Participants
Twenty participants (9 male, 11 female, age range 24-65 years old), who did not take part in the instructional director matcher task, took part in the perception experiment (without receiving any form of compensation).

Stimuli
Twenty pairs of sound fragments from the audio recordings of the production experiment were selected, in order to perceptually compare speech accompanied by gesture to speech without gesture. The sound fragments were presented in pairs and were selected on the basis of their similarity in the type and number of words that the directors used. Each pair of sound fragments consisted of two recordings of the same director instructing a matcher, using very similar or exactly the same words in both recordings. The pairs of recordings consisted of one audio fragment produced when the director was unable to use his or her hands (see example 1) and one audio fragment produced when the director was able to gesture and actually produced at least one gesture (see example 2, where an iconic gesture was produced during the bracketed phrase).

(1) "Nou je pakt hem vast" – *Well, you hold it.*
(2) "Oh je [pakt hem] weer hetzelfde vast".- *Oh you [hold it] again in the same way.*

Our selection criteria for sound fragments to be included in the perception experiment meant that all sound pairs that met our criteria (namely sound pairs with similar wording and of similar length, of which one sound fragment was produced when the director was unable to gesture, and of which the sound fragment that was produced when the director could gesture actually contained at least one gesture) were included in the perception experiment.

The order in which the fragments were presented was counterbalanced over the experiment. This means that for some sound pair fragments the first sound fragment that the participants heard was the one in which the speaker could not gesture, whereas for other sound pair fragments the second sound fragment was the one in which the speaker could not gesture.

Procedure

The twenty participants listened to the twenty pairs of sound recordings and were asked to decide for each pair in which one the director was gesturing. The participants' instructions did not mention whether they should focus on a specific aspect of speech and the participants were only allowed to listen to each fragment once, forcing them to base their decision on initial impressions.

Design and analysis

The relatively small number of sound pair fragments meant that we only took into account whether the speaker was able to gesture or not. We did not take mutual visibility or number of attempts into account. For each pair of sound fragments, a participant received a point if the answer given was correct, that is, if the participant picked the sound fragment where the speaker produced a gesture. We tested for significance by using a t-test on the mean scores.

Results

We found no effect of the ability to gesture on the number of correct answers (M = 10.95 out of 20 correct) in the perception test. Participants were unable to hear in which fragment the director was gesturing and scored at chance level, $t_{(19)} = 1.84$, *n.s.*

Discussion

In this first, exploratory, study of this thesis, the primary goal was to see whether we can observe a direct effect of producing gestures on speech. This was inspired by the often cited study by Dobrogaev (1929) where participants were immobilised while speaking, with the alleged consequence that their speech became less fluent and more monotonous. Unfortunately, even though this study is often cited, its details cannot be recovered. In any case, Dobrogaev's observations were anecdotal and not based on controlled experimental data. Therefore, the present study was unable to use Dobrogaev's exact methodology and had to come up with its own experimental setup.

The setup that was used had several advantages. Firstly, the setting in which participants were able to gesture and could see their addressee was fairly natural (in comparison with, for example Krahmer and Swerts, 2007), with participants being free to talk as they wished. Secondly, the overall setting allowed us to take several aspects of gesture and speech production into account. We could create control conditions in which there was no ability to gesture, in which there was no mutual visibility and in which participants performed tasks of differing difficulty. The design ensured that even though the overall setting was fairly natural, the proceedings of the experiment were still relatively controlled and this meant that speech from the participants in different conditions was comparable. Furthermore, the experiment was set up in such a way as to make it as likely as possible that participants would (want to) gesture. The nature of the task was likely to elicit gestures since it is hard to conduct a motor task such as instructing someone to tie a tie knot without using your hands. In addition, the director was seated on an armless chair, making it more likely that he or she would gesture. Also, the experiment was set up with two participants since the attendance of an (active) addressee has been shown to lead to more gesture production (Bavelas, et al., 2008). In short, effort was taken to ensure that the task would elicit many gestures and the setup was such that a range of communicative situations could be taken into account, from free moving face to face interaction to restrained movement without mutual visibility.

Considering this setup and results from previous studies, the hypothesis was that there would be significant differences between speech with and speech without gestures, giving us an insight into the direct influence of gestures on speech. The results, however, showed no significant main effects of the ability to gesture on almost all the dependent measures we took into account. The only main effect we did find, of the ability to gesture on the number of gestures, was unsurprising and merely served as a manipulation check. We found no main effect of the ability to gesture on the duration of the instructions, on the number of words used, the speech rate, or on the number of filled pauses used. The acoustic analyses also did not show any significant differences between the cases when the director was prevented from gesturing and the cases when the director gestured, and participants of the perception test were unable to hear a difference between fragments of the directors' speech with and without gestures.

As was noted in the introduction of this chapter, not being able to gesture can be seen as a complicated communicative setting, arguably comparable to other communicatively difficult settings such as when there is no mutual visibility or during a complicated task, such as describing a complex tie for the first (rather than the third) time. Interestingly, we did find that the number of attempts resulted in some significant differences, which are in line with what was found in earlier studies (e.g. Clark & Wilkes-Gibbs, 1986). For speech duration we saw that it was getting shorter for each consecutive attempt. The same applies to the number of words that directors used to instruct the video clips, where we found that the number of words was getting smaller with each consecutive attempt. There was also an effect of the number of attempts on the rate of filled pauses, with fewer filled pauses being used with each consecutive attempt.

Although we did not find main effects of the ability to gesture on our dependent measures (except the manipulation check of the number of gestures), we did find a significant interaction between the ability to gesture and the number of attempts on the rate of filled pauses. When participants were unable to gesture in the first attempt their speech had a lower rate of filled pauses than when they were able to gesture. In the second and third attempt however, the inability to gesture led to an increase in the rate of filled pauses.

Taking into account the general focus of this study and our experimental setup in which it was expected that people would feel a strong need to gesture, also when they were not able to, it was a surprise to see that there were no main effects of the ability to gesture on any of the relevant measures taken into account. The interaction effect between the ability to gesture and the number of attempts on the use of filled pauses was a surprising result. Apparently being unable to gesture caused the initial instructions to become more fluent, but the second and third instructions to become less fluent. It was expected that the inability to gesture would cause the instructions to be less fluent overall, not just in the second and third attempt. It should be kept in mind, however, that when we look at the descriptives, as given in table 2.1, we see that the decrease in rate of filled pauses when participants are unable to gesture in the first attempt is only .004 (filled pauses per word), which suggests that (albeit significant) this effect should be interpreted with care.

There are two ways in which we can look at the fact that, overall, we did not find an effect of the ability to gesture. It could be the case that there really is no difference between speech with and speech without gesture in the data from this study or it could be the case that some differences exist but that we have not found them yet. Starting with the latter option, it might be that there are differences that we have not looked at so far. Previous studies have found effects of gestures (or the enforced lack of them) on speech but these effects have been fairly small and detailed (for example only related to spatial language). It is conceivable that this also applies to the current data set. However, the focus of this study was on speech fluency and monotony. The variables that we took into account are all variables that can be considered to be related to speech fluency and monotony. We did not find any main effects on these variables. Therefore, we do not consider it very likely that large differences with regard to speech fluency and monotony exist in this dataset that we have not analysed yet.

A question might be whether the fact that we treated filled pauses as "words" may have artificially increased our measure of speech rate, thereby concealing possible existing rate differences between experimental conditions. In general, including filled pauses does indeed increase speech rate, but this did not bias our results in any way. This would only be a problem if the relative contribution of filled pauses to speech differs across conditions, which was not the case.

If it is the case that there really are no differences in fluency and monotony in this data set between speech with and speech without gestures, we have to consider why this would be so. Are gestures simply not as influential on speech as has been previously assumed or are there other reasons which might have caused the lack of an effect? It might be that the task was not as difficult as was assumed with participants not feeling the need to use gestures as much as was anticipated. This would mean that since participants were not likely to gesture anyway, the inability to gesture did not cause any speech problems for the participants. However, the fairly large mean number of gestures that were produced (as shown in tables 2.1 and 2.2) show that this is unlikely. Moreover, during debriefing, participants often mentioned that they found the task very difficult.

It might be the case that there was no effect of the inability to gesture because, although participants were prevented from using their hands for part of the experiment, this did not stop them from gesturing. We found that asking people to sit on their hands did not stop them completely from moving around. Minor movements, such as movements of the finger tips or muscle tensions could still have occurred, as well as gestures produced by other parts of the body, such as foot, head and shoulder gestures. These have presently not been taken into account. Also, it can be argued that even when people do not produce a physical gesture or movement, this does not necessarily mean that they did not intend to produce a gesture. In other words, a lack of effect could also be due to an intended, but not realised motor command. This would mean that speech and gesture are so closely related that it is not possible to completely separate the two, not even by refraining people from using their hands.

Given these uncertainties, it is difficult to say what the impact of this study is on models of speech-gesture production. Most models proposed in the literature rest upon the assumption that speech and gesture are closely related (e.g. Kendon, 1980, 2004; McNeill, 1992 inter alia), but how the two are related exactly is still a matter of some debate. Consider, for instance, the models proposed by Kita and Özyürek (2003), Krauss, Chen, and Chawla (1996), and de Ruiter (2000), which are all based on the blueprint of the speaker proposed by Levelt (1989). These models all propose the addition of a new gesture stream, which shares its point of origin with the speech production module but is otherwise separated. The models differ primarily in where the two streams (speech and gesture) part. Krauss and colleagues, for example, argue that the separation happens before conceptualization, while both de Ruiter, and Kita and Özyürek argue that it takes place in the conceptualizer. McNeill and Duncan (2000) take a different perspective and argue that speech and gesture are not separate streams, but are produced jointly, based on what they call "growth points". Thus, even though these researchers agree that speech and manual gestures are closely related, they disagree on how tight this relation is (see the introductory chapter of this thesis for further discussion).

Different explanations of our results could potentially have different implications for speech-gesture models. If the lack of an effect of ability to gesture on speech production is caused by the fact that speakers cannot really be prohibited from gesturing (meaning that participants were still gesturing in some way, or had an intention to do so, even when their hands were restrained), this would provide evidence for the claim that speech and gesture are very closely related indeed. If, on the other hand, the lack of an effect was caused by the fact that it does not matter for speech production whether speakers gesture or not, this would suggest that, at least as far as fluency/monotony of speech is concerned, speech and gesture are not so closely related. If we are to assume, somewhat simplifying, that speech properties such as fluency or monotony are largely determined by the later phases of speech production (such as the articulator, in Levelt's terms), our findings would still be consistent with models arguing for a separation between speech and gesture streams before or in the conceptualizer.

However, before definitive conclusions about this can be drawn, more research is needed. Various lines of future research naturally suggest themselves, both related to the gestures that were studied and the task. In the present study, all speech-accompanying gestures were taken into account, and no distinction between different types of gestures was made. Krahmer and Swerts (2007) found evidence, in a rather controlled setting, of the impact of gestures on speech production, as discussed in the introduction of this chapter. However, they only looked at beat gestures. Beats can be characterized as short and quick flicks of the hand, that often serve the purpose of emphasizing a word or phrase (McNeill, 1992), and in this sense they are comparable to the role that pitch accents play in Germanic languages and perhaps they are also linked closer to speech than other kinds of gestures. In fact, Krahmer and Swerts (2007) explicitly argue that different kinds of gestures might be integrated differently in models of speech-gesture production. It is conceivable, for instance, that beat gestures do, but other kinds of gestures do not, directly influence speech production. The work by Bernardis and Gentilucci (2006) also suggests a close link between gesture and speech, for a different type of gestures, (i.e. conventionalised greeting gestures), but more work on a wider range of gestures is clearly needed. In particular, the impact of different kinds of gestures on speech production should be studied in more detail in future research.

In a somewhat similar vein, the task that was used in this study could have been of influence as well. Previous research has suggested that gestures are particularly useful in spatial and motor descriptions (e.g. Hostetter & Alibali, 2010; Hostetter, et al., 2007). With this in mind, we opted for a production experiment in which participants had to describe concrete tie-knotting actions to an addressee. With this task we expected that participants would feel a strong need to gesture, which indeed turned out to be the case (exemplified by the fact that many gestures were produced when participants were able to do so and by the fact that some participants had 'slips of the hands', i.e. gestured, even when they were supposed to be sitting on their hands). However, it might be the case that a different task might have yielded different results. What, for instance, if speakers were asked to describe something more abstract or what if the task would be more difficult (perhaps resulting in more tip-of-the-tongue states)? In general, it is conceivable that different tasks cause speakers to produce different kinds of gestures, which in turn might differently influence speech production as well.

In conclusion, the strength of the experimental design with its fairly natural setting has led to a large data set, of which many aspects can be studied. The measures that have been analysed presently did not show any main effects of the ability to gesture on speech and the (lack of) results may be only applicable to the domain of instructing motor tasks (which tie knotting can be argued to be an example of). However, we have been able to show that topic complexity, in this case in the form of the number of attempts that directors had at giving instructions, influences many aspects of speech. We showed that directors used less time, fewer words and fewer filled pauses for each consecutive attempt. This is in line with previous findings on repeated references, for example by Clark and Wilkes-Gibbs (1986). In the present study directors did not produce references per se but longer stretches of speech when instructing the matcher how to tie a tie. In the following chapters of this thesis however, we will continue this line of research by focusing specifically on repeated references.

Acknowledgements

We would like to thank Bas Roset and Nick Wood for statistical and technical support and help in creating the stimuli, Joost Driessen for help in transcribing the data, Martijn Goudbeek for statistical support and Katya Chown for providing background information on Dobrogaev. Parts of this paper were presented at the Tabu dag 2009 in Groningen, at the Gesture Centre at the Max Planck Institute for Psycholinguistics, at the 2009 AVSP conference, at LabPhon 2010 and at ISGS 2010. We would like to thank the audiences for their suggestions and comments. Finally, thanks to the anonymous reviewers for their useful and constructive comments.

3

Reduction in gesture during the production of repeated references

Abstract

In dialogue, repeated references contain fewer words (which are also acoustically reduced) and fewer gestures than initial ones. In this chapter, we describe three experiments studying to what extent gesture reduction is comparable to other forms of linguistic reduction. Since previous studies showed conflicting findings for gesture rate, we systematically compare two measures of gesture rate: gesture rate per word and per semantic attribute (Experiment I). In addition, we ask whether repetition impacts the form of gestures, by manual annotation of a number of features (Experiment I), by studying gradient differences using a judgment test (Experiment II), and by investigating how effective initial and repeated gestures are at communicating information (Experiment III). The results revealed no reduction in terms of gesture rate per word, but a U-shaped reduction pattern for gesture rate per attribute. Gesture annotation showed no reliable effects of repetition on gesture form, yet participants judged gestures from repeated references as less precise than those from initial ones. Despite this gradient reduction, gestures from initial and repeated references were equally successful in communicating information. Besides effects of repetition, we found systematic effects of visibility on gesture production, with more, longer, larger and more communicative gestures when participants could see each other. We discuss the implications of our findings for gesture research and for models of speech and gesture production.

This chapter is based on:

Hoetjes, M., Koolen, R., Goudbeek, M., Krahmer, E., Swerts, M. (2015). Reduction in gesture during the production of repeated references. *Journal of Memory and Language*. 79-80, 1-17.

Introduction

When we communicate, we continuously refer to objects and persons in our vicinity. Typically, the same target is referred to multiple times during an exchange, and speakers may use both speech and gesture when doing this. For example, a speaker who wants to point out a particular building for her addressee can produce an initial description such as "the brown building at the back of the university campus shaped like this", accompanied by two hand gestures indicating first the location and then depicting the shape of the building. Later in the interaction, when she refers to the same building again, a typical description might be "the building shaped like this", produced in tandem with only the shape gesture.

A substantial body of literature has shown that, as the preceding example suggests, repeated references consist of fewer words (e.g., Brennan & Clark, 1996; Clark & Wilkes-Gibbs, 1986). In addition, we know from various studies that repeated references can be reduced acoustically as well, in such a way that, for example, the second realisation of the word "building" in our example may be less intelligible (when heard in isolation) than the initial one (e.g., Aylett & Turk, 2004; Bard, et al., 2000; Fowler, 1988). Finally, and most importantly for the current study, a number of studies have shown that repeated references are also accompanied by fewer gestures (e.g., de Ruiter, et al., 2012; Holler & Stevens, 2007; Holler, et al., 2011; Levy & McNeill, 1992).

Most of the earlier studies on gesture reduction focused on numeric, quantitative reduction, and while they agree that repeated references contain fewer gestures per description than initial ones, a closer look reveals a mixed pattern of results. To study the relative contribution of gesture and speech to repeated references, researchers generally focus on *gesture rate*. Reconsider our example: the initial description combines 13 words with 2 gestures, and thus has a gesture rate per word of .15 (= 2/13). The repeated reference consists of 5 words and 1 gesture, suggesting that in this case the gesture rate has actually increased to .2. Indeed, some studies (e.g., Holler, et al., 2011) found a general increase in gesture rate per word, while others did not (de Ruiter, et al., 2012), or found a reduction in gesture rate (Galati & Brennan, 2014; Jacobs & Garnham, 2007).

An alternative is to look at gesture rate as a function of the semantic attributes in a referring expression. In our initial example, four attributes of the target were included (colour, type, location, shape), and combined with two gestures, yielding a gesture rate per attribute of .5. The repeated example with one gesture mentions two attributes (type, shape), and thus has a gesture rate per attribute of .5 as well. This highlights the importance of how gesture rates can be conceptualised, indicating that different metrics may yield different results.

In view of the mixed results of earlier studies, and given the importance of comparing different metrics for gesture rate, we will systematically compare these two in the current study, asking (1) whether repeated references lead to reduction in gesture per words, (2) whether repeated references lead to reduction in gesture per attribute, and (3) whether we can observe any differences in how these gesture rates develop with repetition.

In addition, we investigate whether the gestures produced in repeated references themselves are different in form from comparable initial gestures. It could be, for instance, that the initial shape gesture in our running example is produced with two hands, depicting the shape precisely and multiple times, while the repeated reference is accompanied by a single one-handed gesture only vaguely suggesting the shape of the target building. Alternatively, it might be that repeated gestures are similar to initial ones in general form, but differ in more gradient ways, much like repeated articulations of the same word ("building") tend to be less clearly articulated. This question has received little attention in the literature, although some studies have looked at some qualitative aspects and generally find evidence for reduction in form (e.g., Galati & Brennan, 2014; Gerwing & Bavelas, 2004; Holler & Stevens, 2007). However, these studies tend to vary with respect to the measures that are used, resulting in an incomplete understanding of how repetition influences how speakers realize their gestures qualitatively. We systematically compare the gestures produced during initial and repeated references, asking (1) whether repeated gestures differ in general form from initial ones and (2) whether there are perceivable gradient differences between initial and repeated gestures. In addition, (3) taking the analogy with repeated realisation of words seriously, we predict that repeated gestures are less "intelligible" when presented without context than initial ones; a prediction which has not been tested before.

By combining quantitative and qualitative analyses, as we do in this paper, we hope to reconcile the conflicting earlier results on gesture rate in repeated references, and further our understanding of the relative contribution of gesture and speech in repeated references, which also has implications for psycholinguistic models of speech and gesture production.

Background

Reduction in speech

Roughly speaking, we can divide previous research on reduction in spoken repeated references into studies that look at reduction at the acoustic level, and studies that look at reduction at the lexical level.

The idea that certain predictable words are reduced acoustically has a long history. Lieberman (1963) compared productions of the word "nine" in a context where it was not predictable ("The word you are about to hear is nine.") with those in a context where it was fully predictable, at least for a native speaker of English ("A stitch in time saves nine", meaning that it is better to do something now than wait until later). Lieberman (1963) found that in the unpredictable context, the word "nine" was longer, had a higher pitch peak (F0) and was rated as more intelligible when taken out of context.

One way in which words can become more predictable is by producing them repeatedly. In particular, realisations of words that represent new information in a discourse tend to be articulated differently (e.g. longer duration, higher pitch) than realisations of the same words occurring later in the discourse, where they express given information (Aylett & Turk, 2004; Bard, et al., 2000; Brown, 1983; Fowler & Housum, 1987; Kaland, Krahmer, & Swerts, 2014; Lam & Watson, 2010). As in the "nine" example of Lieberman (1963), the references to given information are generally less intelligible when presented in isolation than the references to new information (e.g., Bard, et al., 2000; Fowler, 1988; Fowler & Housum, 1987).

Bard, et al. (2000), for example, tested whether speakers adjust the reduction in their references to what the listener does or does not know. Bard and colleagues studied this using the Map Task paradigm (Anderson, et al., 1991), in which pairs of speakers communicated about a route on a schematic map with labelled landmarks (like a rope bridge or a banana tree). By manipulating the maps, the knowledge of speakers and listeners was manipulated independently. Words introducing landmarks to two successive listeners were less intelligible when they were repeated, whether they were new for the second listener or not (Experiment 1). In addition, repeated references became less intelligible, also when the listener expressed that he could not see the landmark (Experiment 2). This suggests that speakers reduce repeated references, irrespective of the needs of the listener. Bard, et al. (2000) suggest that this pattern of results can be explained by assuming a two-component language production model, consisting of a fast component, which depends on the speaker's knowledge, and a slow, optional component drawing inferences about what the listener knows (but see e.g., Galati & Brennan, 2010; Galati & Brennan, 2014 for a different take on this issue).

Lexical reduction in repeated references has been documented in a seminal paper by Clark and Wilkes-Gibbs (1986), in which pairs of participants engaged in a director-matcher task. In this task, one participant (the director) is instructed to describe an array of humanoid tangram figures, in such a way that another participant (the matcher) can rearrange the figures in front of him such that they match the described ordering. Crucially, this task is repeated six times, so that each tangram figure is discussed multiple times, during different trials. In a typical example, a director might describe a figure in trial 1 as "a person who's ice skating, except they're sticking two arms out in front," while in trial 6 the same figure is referred to simply as "the ice skater" (Clark & Wilkes-Gibbs, 1986, p. 12). This general finding has been replicated many times, and is often explained in terms of an emerging common ground between interlocutors (Clark & Brennan, 1991), where common ground can informally be understood as the information that is shared by interlocutors (or which they assume to share). In this view, common ground makes it possible to reduce repeated references, because speakers can rely on common ground in subsequent references. By repeatedly referring to a target, interlocutors quickly agree on how to refer to an object, and in doing so establish these as common ground. The emergence of a "conceptual pact" (Brennan & Clark, 1996) such as "the ice skater" is a good illustration of this; over time, interlocutors form a shared conceptualization of a particular target, which allows them to refer to it in a more efficient way (using fewer words).

This short overview illustrates that reduction - both acoustically and lexically- in speech has been well established. In recent years, reduction in gesture has been studied as well, and we turn to these studies next.

Reduction in gesture

Speech-accompanying, or co-speech gestures (henceforth called gestures) can be defined as the (usually manual) symbolic movements that people make while they speak (Kendon, 2004; McNeill, 1992). As the phrase co-speech gestures suggests, these movements are closely related to the speech they accompany. Indeed, it has long been suggested that gesture and speech are tightly connected at the semantic level (Kendon, 1972, 1980, 2000, 2004; McNeill, 1985, 1992; McNeill & Duncan, 2000), and many studies found quantitative support for this claim (e.g., Kita & Özyürek, 2003; Krahmer & Swerts, 2007; So, Kita, & Goldin-Meadow, 2009). For example, So, et al. (2009) found, in a scene description experiment, that speakers could use gesture locations to identify referents in discourse, but that they tended to do this only when the referent was also identified in the accompanying speech. The authors interpret this as an illustration of gesture going "hand-in-hand" (So, et al., 2009, p. 123) with speech. Similar ideas have been expressed by, among others, Bavelas, et al. (2008) and Clark (1996). Clark, for instance, argued that gestures, much like intonation, are an integral part of the communicative signal, suggesting that it would be "difficult to produce the speech without the gesture" (Clark, 1996, p. 179).

Based on considerations such as these, a reduction in speech might be accompanied by a reduction in gesture, and this is indeed what has been claimed. Levy and McNeill (1992), for instance, conducted an analysis of four narratives describing a commercial film and noted that speakers were more likely to gesture in their initial references to people than in later references to the same people in the same scenes. In addition, the authors suggested that new information should not only be accompanied by more gestures, but also by more complex ones than given information.

Various studies have followed up on these initial observations, looking at both quantitative and qualitative analyses of gesture, but the pattern of results is "complex" (Holler, et al., 2011, p. 3), with various "conflicting findings" (de Ruiter, et al., 2012, p. 235), partly because studies rely on different methods, ranging from collecting narrations to referential communication tasks, and consider a range of differing dependent variables.

Gerwing and Bavelas' study (2004) was the first test of the idea that gestures referring to given information are "sloppier" (p. 176) than those referring to new information, just like words referring to given information are produced with a sloppier articulation. The authors tested this by having participants play with a number of toys, including a finger cuff (also known as a Chinese finger trap, which 'traps' one's index fingers at both ends of a small cylinder), and afterwards asked them to explain, without the toys being present, to two other participants what they did with these toys. One of the listeners in this triad had played with the same toys, the other one with different ones, and the speaker was aware of this. Gerwing and Bavelas (2004) concentrated on the gestures that speakers used in their initial identification of the finger cuff, and found that when speakers described it to the participant who had also played with this toy, their gestures were more "elliptical" (Gerwing & Bavelas, 2004, p. 170), compared to the gestures made when describing the toy to a person who had not played with it before (i.e., no common ground), in which case the associated gestures were more elaborate and complex. This was established by having two independent analysts judge which of the two dialogues in each triad contained gestures that conveyed "more information, were more complex, or were more precise" (p. 168) and revealing that the two judges reliably selected the no common ground dialogues as the ones having more informative gestures. A qualitative analysis of a number of gestures confirmed that gesture parts depicting new information were larger and more precise (Gerwing & Bavelas, 2004, p. 182).

Holler and Stevens (2007) obtained similar results in a referential communication task. They asked participants to locate targets in *Where's Wally?* pictures, and observed that when speakers referred to the size of an object in one of these pictures to an addressee for whom this information was new (unknowing recipients), they generally represented it only in gesture or in gesture and speech. By contrast, when the size information was shared knowledge, speakers mainly realised this information in speech only. In addition, Holler and Stevens (2007) had two independent judges score the perceived size of gestures on a 7-point Likert-scale, and found that size scores for gestures produced to knowing recipients were lower than those for unknowing ones.

Similarly, Jacobs and Garnham (2007) asked speakers to retell a comic strip story multiple times, either to the same listener or a different one. They found that repeated narration to the same listener resulted in a decreased gesture rate, but this did not occur when retelling to different addressees, for whom the story was new. Galati and Brennan (2014), using a similar design, found that speakers who retold a story to an old addressee (i.e., one who had heard the story before) gestured less frequently than when they retold it to a new addressee. In addition, Galati and Brennan (2014) showed that the gestures in retellings to old addressees were smaller and less precise than in those retold to new addressees.

However, other studies have yielded results that are only partly compatible with this. Holler and Wilkin (2009), for example, had speakers narrate stories to an addressee, where some narrative scenes were part of the common ground, because speaker and addressee had watched them together. Using a semantic feature account, the authors found that utterances, taking into account information from speech and gesture, expressed less semantic content when there was common ground between speaker and listener. However, they also reported the “paradoxical result” (Galati & Brennan, 2014, p. 449) that speakers gestured at a higher rate (per 100 words) in the common ground condition, suggesting that gestures are relatively more communicatively important when there is common ground. Holler, Tutton and Wilkin (2011) similarly found that gesture rate increased with accumulating common ground, when objects were repeatedly referred to.

To further complicate the picture, de Ruiter, Bangerter and Dings (2012) found no evidence for an increase in gesture rate in repeated references, but also little or no evidence for a decrease in gesture rate. De Ruiter, et al. (2012) explicitly contrasted the aforementioned hand-in-hand hypothesis (So, et al., 2009) with an alternative, which they call the trade-off hypothesis (based on obervations in, among others, Bangerter, 2004; de Ruiter, 2006; Melinger & Levelt, 2004; Van der Sluis & Krahmer, 2007). This hypothesis suggests that when speaking gets harder, speakers will rely more on gestures (and vice versa, although this second part was not tested by de Ruiter, et al., 2012). This leads to the prediction that during the production of repeated references (which, as argued above, are easier to produce than initial ones), speakers will rely less on gestures, which should lead to a decrease in gesture rate. De Ruiter and colleagues studied this using an adaptation of the tangram matching task, inspired by Bangerter (2004), in which directors could identify targets to matchers from a mutually visible array of tangram figures on a wall poster. Since the trade-off between speech and gesture may depend on the type of gesture, the authors coded deictic (pointing) gestures as well as iconic gestures, illustrating a feature of the target (for instance its shape). The authors studied the gesture rate per 100 words, and, in general, found little support for the trade-off hypothesis (with one exception: the gesture rate for pointing gestures decreased when speakers produced repeated references, see de Ruiter, et al., 2012, p. 244).

To sum up: some studies find evidence that gesture rate decreases when information is shared or repeated (e.g., Galati & Brennan, 2014; Jacobs & Garnham, 2007), some find that it increases (Holler, et al., 2011; Holler & Wilkin, 2009), and others find that it stays the same (de Ruiter, et al., 2012). However, as illustrated in our opening example, and also noted by others (Galati & Brennan, 2014; Holler & Wilkin, 2009), it is not only the number of words speakers use, but also the semantics of their utterances that are relevant. Galati and Brennan (2014, p. 444) even suggest that gesture rates per words can be misleading and that rates per "unit of semantic content" should be considered as well.

Considering the qualitative aspects of gestures referring to given information: there is indeed some evidence that these are reduced in comparison to gestures referring to initial information, but so far only a limited number of studies have looked into this, all using different measures, ranging from, for instance, an analysis of which dialogue contains more informative and precise gestures (Gerwing & Bavelas, 2004), to coding of size information in gesture as judged on a 7-point scale (Holler & Stevens, 2007), the location of the gesture in gesture space (Holler, et al., 2011), and the distance between hands in two-handed gestures or displacement of the hand in one-handed gestures, both on a 7-point scale (Galati & Brennan, 2014).

This paper aims to further our understanding of gesture production when referring to new or given information, by systematically comparing gesture rates per word and per semantic attribute, and by looking in detail at the qualitative aspects of the produced gestures, by manual annotation, but also using judgment studies of gestural precision and intelligibility.

Visibility and gesture

Following many previous studies (see Bavelas & Healing, 2013, for discussion), we include visibility as an additional variable in our design, in such a way that one group of participants will be able to see each other (mutual visibility), while the other group is prevented from doing so using a screen (no visibility).

Traditionally, gesture researchers have used visibility-designs to get a better understanding of the extent to which speakers produce gestures for their addressees[4]. For example, Alibali, et al. (2001, p. 169) write "if speakers produce gestures in order to aid listeners' comprehension, they should produce fewer gestures when their listeners are unable to see those gestures." Indeed, various studies have found that the gesture rate (per word) decreases when participants are not able to see each other, although speakers do still produce gestures when the listener cannot see them (e.g., Alibali, et al., 2001; Bavelas, et al., 2008). It has also been found that the decrease in gesture rate in part depends on the *kind* of gestures under consideration; the rate with which speakers produce beat gestures, for example, is roughly the same with and without visibility, while deictics and (obligatory) iconics (i.e., iconic gestures needed for understanding) are more frequent with mutual visibility (Alibali, et al., 2001; Bavelas, et al., 2008; de Ruiter, et al., 2012).

These results raise an obvious question: why do speakers still produce some gestures in the no-visibility condition? This is unexpected when one assumes that speakers produce gestures for the benefit of their addressees. Various explanations have been offered, including the suggestion that these gestures may serve cognitive needs of the speaker (Alibali, et al., 2001; Kita, 2000; Krauss, 1998; Melinger & Kita, 2007). But alternative interpretations have also been defended: speakers may produce gestures that are not visible for the addressee out of habit (Cohen & Harrison, 1973) or for an imagined audience (Fridlund, 1994).

Clearly, these are complicated issues, but one consensus that seems to be emerging is that different gestures can have multiple functions (e.g., Alibali, et al., 2001), with some gestures being more speaker- and others more addressee-oriented. Perhaps more important for the current study is that, besides gesture rate, visibility may also influence the qualitative form of the gesture (e.g., Bavelas, et al., 2008; Gullberg, 2006). Bavelas and colleagues, for example, found that speakers describing an 18th century dress with a distinctive shape used larger gestures in a mutual visibility condition (as if placing the

[4] It is worth noting, incidentally, that visibility designs have also been used in studies where gesture is not the main focus of attention, such as Clark and Krych (2004) and the aforementioned study by Bard et al. (2000, p. 6), who had participants separated by a "flimsy barrier" in one of their experiments (see also Anderson, Bard, Sotillo, Newlands, & Doherty-Sneddon, 1997).

dress around their own body, Bavelas, et al., 2008, pp. 509-510) as opposed to speakers describing the same dress via telephone (in which case the gestures were more likely to be on the same scale as the dress on the picture). We include visibility in our design to study whether and when gesture reduction, both in terms of gesture rate and in terms of gesture form, is more speaker- or more addressee-oriented.

The present studies

To further our understanding of gesture production when speakers refer to new or given information, we conduct a series of production and judgment experiments. In Experiment I we collect data from speakers who refer repeatedly to the same target. For this, we rely on a director-matching referential communication task. Referential communication tasks do not require speakers to tell a narrative, and hence references need not be embedded in a larger structure where different factors (such as relative importance to the overall narrative) may conceivably influence the realisation of referring expressions (see de Ruiter, et al., 2012; Holler & Stevens, 2007 for similar arguments). We opt for abstract, hard to describe figures with different shapes ("Greebles", Gauthier & Tarr, 1997), which are expected to result in spontaneous descriptions containing both verbal and gestural references to these shapes, both in initial and in repeated descriptions. Both initial and repeated references to the same target are fully transcribed and analysed in terms of the semantic attributes used by speakers. All gestures produced by speakers during these references are analysed as well, allowing us to study both the number of gestures per 100 words and the number of gestures per semantic attribute. Since we are primarily interested in how speakers attenuate their descriptions as a function of repetition, we focus on the individual speaker and not on interactive aspects in our analyses (cf. Bavelas & Healing, 2013).

Besides the quantitative analyses, in which we compare gesture rate per 100 words and per semantic attribute as a function of repetition, we also study how the gestures themselves differ between initial and repeated references. When speakers repeatedly express the same shape in gesture, can we observe qualitative differences between these gestures? Based on the literature, we approach this question from two perspectives, and using two different methods. On the one hand, relevant gestures of initial and repeated descriptions are manually annotated and compared. Based on earlier work (Galati & Brennan, 2014; Holler & Stevens, 2007), we expect gestures produced during repeated references to a target to be smaller. In addition, we ask whether other systematic "discrete" differences can be observed, where we expect gestures produced during repeated references to be shorter in duration, more often produced with one hand, and containing less repetitive movements during the stroke. On the other hand, a conceivable alternative is that the gestures do not change in this discrete manner, but instead differ in a more gradient way, in line with, for instance, Gerwing and Bavelas (2004). This possibility is tested using a judgment test (Experiment II), in which naïve participants are asked to say which of two gestures, one taken from an initial and one from a repeated reference, contains "more information, is more complex, or more precise" (as in Gerwing & Bavelas, 2004, but then applied at the level of gesture rather than dialogue). Finally, if gestures from repeated descriptions are indeed sloppier, analogously to the way in which repeated words are articulated in a sloppier way (as suggested by Gerwing & Bavelas, 2004), we would expect that these are less intelligible/communicative as well. We test this in Experiment III, where participants are shown video clips with either a gesture from an initial or from a repeated reference to a Greeble object, and are asked to indicate which from a pair of Greebles is the one the speaker is gesturing about. Our findings have implications for current psycholinguistic models of speech and gesture production, which we describe in the General conclusion and discussion section of this chapter.

Experiment I: Production of repeated references

Participants

In total, 162 speakers of Dutch took part in the experiment. In the visibility condition there were 106 participants, all undergraduate students (31 male, 75 female, age range 18-29 years old, $M = 21$ years and 7 months), who took part in pairs as partial fulfilment of course credits. From these pairs, data from 5 pairs was left out because there were technical problems, leading to a data set consisting of data from 48 pairs of participants (48 directors and 48 matchers). In the no-visibility condition there were 56 participants, all undergraduate students (21 male, 35 female, age range 17-30 years old, $M = 20$ years and 7 months). From these pairs, data from one pair was left out because the participants had not understood the procedure of the experiment, leading to a data set consisting of data from 27 pairs of participants (27 directors and 27 matchers). In both conditions, participants were randomly assigned the role of director or matcher.

Stimuli

The stimulus materials consisted of pictures of Greebles[5], which are hard to describe, small yellow objects, initially designed so as to share abstract characteristics with human faces. These Greebles vary in terms of their main body shapes ("Samar", "Galli", "Radok", "Tasio"), their gender ("Plok", "Glip"), the different types of protrusions that they have ("Boges", "Quiff", "Dunth") and in terms of the shapes and sizes of these protrusions (see figure 3.1 for an example Greeble, and see Gauthier & Tarr, 1997, for a more detailed description of the Greebles and their properties).

Since directors would naturally be unfamiliar with the specialized vocabulary developed to describe Greebles ("Tasio", "Glip", etc.), they were expected to describe in detail the shapes and protrusions in both their initial and repeated descriptions, for which both speech and gesture would be helpful. In this way, we could collect sequences of shape descriptions, both in word and gesture, for initial and repeated descriptions. In order to make the Greebles look less like animate figures (which might possibly cause participants to rely less on the shape information in their descriptions), they were turned upside down compared to the way in which they were presented in Gauthier and Tarr (1997).

Two picture grids containing 16 Greebles were created. Each picture grid was used for 15 trials, which made a total of 30 trials. The order in which the directors were presented with the two picture grids was counterbalanced over participants. In each trial, there was one target object (marked by a red square), which was surrounded by 15 distractor objects. An example of a picture grid can be seen in figure 3.2.

[5] Images courtesy of Michael J. Tarr, Center for the Neural Basis of Cognition and Department of Psychology, Carnegie Mellon University. URL: http://www.tarrlab.org/

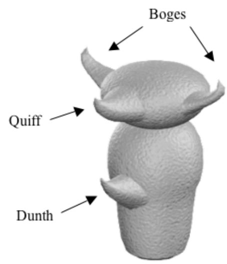

Figure 3.1. Example Greeble, in this case with the main body shape "Tasio" and of the gender "Glip" (names in figure refer to specific types of protrusions).

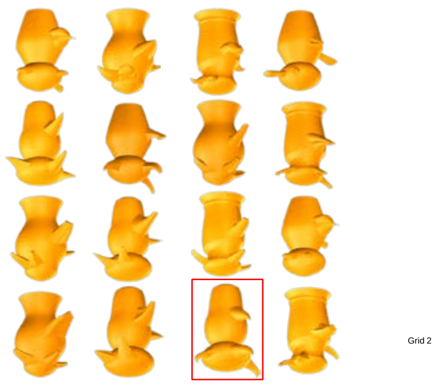

Figure 3.2. Example of one of the picture grids presented to the director. The object with the square surrounding it is the target object of that particular trial.

The crucial manipulation in the task was that several Greebles had to be described repeatedly. In each of the picture grids, two Greebles had to be described twice, and two Greebles had to be described three times; five Greebles were referred to only once. Repeated references to the same object always had a reference to another object in between and were never the first or the last trial of the picture grid. We analysed all descriptions of the Greebles that had to be described three times (i.e. a total of twelve trials per participant; 2 grids x 2 target Greebles x 3 descriptions).

Procedure

The experiment was performed in a lab, where the director and the matcher were seated at a table opposite each other (see figure 3.3 for an example of the setup).

Figure 3.3. Setup of experiment I, in visibility condition, matcher sits on the left and director sits on the right.

The procedure for both visibility conditions was identical, apart from the fact that in the no-visibility condition, there was a large opaque screen between participants, obscuring the view of their entire body (figure 3.3 shows the visibility condition). Both participants were filmed during the experiment, with slightly different camera positions, depending on the visibility condition: in the visibility condition, one camera was positioned behind the matcher (filming the director) and another camera was positioned to the side of the director (filming the entire setup, as in figure 3.3). In the no-visibility condition, both cameras were situated at the side of the screen, one filming the director and one filming the matcher.

The participants were given written instructions and had the opportunity to first ask questions, after which the experiment started. The director was presented with the trials on a computer screen (which was positioned to her side, as in figure 3.3), and was asked to provide a description of the target in such a way that it could be distinguished from the 15 distractor objects. The matcher had a box filled with 16 stacks of cards (one small stack for each Greeble) in front of him, which were not visible to the director (regardless of the visibility condition the participants were in). The cards in the matcher's box showed the same objects as on the director's screen, but these objects were ordered differently for the director and the matcher. Directors were made aware of this, and it was explained during the instruction phase that visual location on the screen could thus not be used, since the matcher saw the figures in a different order in front of him. The instructions stressed that directors were free to describe their target in any way they wanted, but the use of gesture was not explicitly mentioned. The instructions did mention that it was possible that some targets occurred multiple times.

Based on the director's target description, the matcher had to pick the corresponding card from the box in front of him. Once the matcher had found the card that he thought was being described, the experimenter advanced the director to the next trial. Matchers were instructed not to interrupt the director or ask any questions, but for each new object first wait for the director to finish his description, after which they could indicate that they had found the described object. This instruction was inspired by similar instructions in, among others, Alibali, et al. (2001) and Mol, et al. (2009). By instructing our participants in this way, we could collect initial and repeated descriptions in situations that were as comparable as possible, to ensure that any effects could be attributed to our manipulations, and not to possible differences in verbal interaction (see Holler & Wilkin, 2009, p. 273 for a similar argument). After 15 trials, the director was shown the second picture grid containing 16 new objects, and the matcher was presented with a new box filled with stacks of cards of these objects.

Data analysis

Speech annotation For the speech analysis we analysed the duration and the number of words for each reference (this served as a manipulation check, and to compute the number of gestures per 100 words). The duration was based on the moment at which the matcher indicated that the correct object card had been found. This moment was the end point of one reference, and the beginning of another reference (a new trial was shown to the director as soon as the matcher had found the correct object). To analyse the number of words, all speech within a reference was transcribed orthographically.

Repetitions, hesitations, false starts and corrections were all transcribed and counted as words[6].

From the transcribed speech data we annotated the number of attributes per reference, so that we could compute the number of gestures per attribute. The number of attributes is a measure of the references' semantic content. When constructing the trials, we made sure that all targets could be distinguished by means of 4 attributes. We designed an annotation scheme containing 45 attributes that speakers could potentially use when describing a Greeble. This scheme was based on the basic characteristics of Greebles (main body shape, gender, protrusions) and was expanded with attributes describing all other properties that they can possibly have (mainly concerning the protrusions' shapes, locations and sizes). An example of a participant's description of a Greeble and its annotated attributes can be seen below. The annotation shows the ID of each attribute, the name of each attribute, followed by the value of this attribute and the part of the reference (in Dutch) that the attribute consists of. A combination of an attribute and a value is referred to as a property of the target.

Example of a participant's description of a Greeble (in Dutch and English literal translation), followed by the accompanying, systematic, attribute annotations.

> "Eh dit is weer die klassieke vaasvorm met die taille, eh er zit aan de rechterkant echt een hele brede eh uitsteeksel"
> "*Uh, this is again that classic vase shape with that waist, uh, there is on the right side really a very wide uh protrusion*"
> <ATTRIBUTE ID="a1" NAME="family" VALUE="galli">die klassieke vaasvorm met die taille</ATTRIBUTE>
> <ATTRIBUTE ID="a2" NAME="DunthLocation" VALUE="right">aan de rechterkant</ATTRIBUTE>
> <ATTRIBUTE ID="a3" NAME="DunthWidth" VALUE="wide">hele brede</ATTRIBUTE>
> <ATTRIBUTE ID="a4" NAME= "Protrusion"VALUE= "dunth">uitsteeksel </ATTRIBUTE>

[6] Contractions were counted as one word, however, there was only one type of contraction in the data (namely, the Dutch 'zo'n', '*such a*').

Gesture annotation For the gesture analysis we used the multimodal annotation programme ELAN (Wittenburg, et al., 2006). To analyse the quantity of gestures, all gestures occurring during the critical trials (12 per director) were identified and selected. For the qualitative analyses we annotated a subset of these gestures in detail. To make the analyses for first, second and third references as comparable as possible, we selected for each reference the first gesture that a speaker produced when describing the shape of the target object. For these gestures, only the stroke (i.e. the most effortful and meaningful part of the gesture, see Kendon, 1980, 2004; McNeill, 1992) was analysed in detail, without sound. The onset of the stroke was determined by the first video frame in which the most effortful movement started, and the offset of the stroke was determined by the first video frame in which the stroke phase turned into a post-stroke hold phase, or a retraction phase. When a director produced a reference without a gesture, this was treated as a missing value in our analyses on gesture form.

For the gestures that were annotated in detail, we determined the type of gesture, differentiating between iconic, deictic and beat gestures (following McNeill, 1992). Iconic gestures were considered as such when a gesture depicted a particular feature of the target object, such as its main shape or the shape of one of the protrusions. Deictic gestures were pointing gestures, generally used to indicate a specific location of one of the object's protrusions. Beat gestures consisted of a simple rhythmic movement without a semantic relation to the speech it accompanied. We found that overwhelmingly, iconic gestures were used, see table 3.1. Therefore, the different types of gestures were taken together in all qualitative gesture analyses, as described below.

Table 3.1. Distribution of iconic, deictic and beat gestures, over initial, second and third references. For each director, only the first gesture that was produced when describing the shape of the target object was included in this analysis.

Repetition	Iconic	Deictic	Beat
1	214	6	3
2	194	6	4
3	178	5	4

We took the following aspects of gesture form into account:

- Gesture duration: the duration of the stroke (as defined above), in seconds.
- Gesture size: indicating whether the gesture was produced with a finger (1), the hand (2), the forearm (3) or the entire arm (4). If a gesture involved movement of, say, hands and forearm, we noted down the highest score (3).
- Number of hands: indicating whether the gesture was produced with one or with two hands.
- Number of repeated strokes: a stroke was considered repeated when (near) identical strokes followed each other without a retraction phase in between.

The assumption was that gestures associated with initial references would have a longer duration, a larger size, were more likely to be produced with two hands and to repeat the stroke. To assess the reliability of the coding, a subset of 23 gestures (produced by 23 participants) was coded by a second independent annotator, who was blind to the experimental conditions. There was agreement on 83% of cases for gesture size, on all cases for the number of hands, and on 91% of cases for the number of repeated strokes.

Statistical analyses The experiment consisted of a 3 x 2 design, with factors Repetition (levels: initial, second, third) and Visibility (levels: no screen, screen). The statistical procedure consisted of two repeated measures ANOVAs, one by participants (F_1) and one by items (F_2). On the basis of these, *minF'* was computed (Clark, 1973), to see whether the results could be generalised over participants and items simultaneously, while keeping the experiment wise error rate low (Barr, Levy, Scheepers, & Tily, 2013, p. 268). We used Mauchly's test for sphericity to test for homogeneity of variance. When this test was significant we applied a Greenhouse-Geisser correction on the degrees of freedom, but for the purpose of readability we report the uncorrected degrees of freedom for these cases. Bonferroni corrections were used for post hoc multiple comparisons. We only report when analyses show significant results.

Results

Manipulation check As expected based on previous literature, reference duration and the number of words used were lower in repeated references and were unaffected by (a lack of) mutual visibility, while the number of gestures decreased in repeated references and when there was no mutual visibility.

Figure 3.4 provides an overview of the mean reference duration across all conditions. The reference duration decreased in repeated references, $F_1(2,144) = 53.160$, $p < .001$, $\eta_p^2 = .425$; $F_2(2,9) = 9.992$, $p = .005$, $\eta_p^2 = .689$; $minF'(2,13) = 8.411$, $p = .005$. Post-hoc tests showed that all three references differed significantly from each other (all $p < .05$). Figure 3.5 provides an overview of the mean number of words across all conditions. The number of words decreased in repeated references, $F_1(2,144) = 46.497$, $p < .001$, $\eta_p^2 = .392$; $F_2(2,9) = 20.348$, $p < .001$, $\eta_p^2 = .819$; $minF'(2,18) = 14.153$, $p < .001$. Post-hoc tests showed that all three references differed significantly from each other (all $p < .05$). Figure 3.6 provides an overview of the mean number of gestures across all conditions. The number of gestures decreased in repeated references, $F_1(2,144) = 13.102$, $p < .001$, $\eta_p^2 = .154$; $F_2(2,9) = 7.089$, $p = .014$, $\eta_p^2 = .612$; $minF'(2,21) = 4.600$, $p = .022$. Post-hoc tests showed that initial references differed from both second and third references (both $p < .05$), whereas second and third references did not differ ($p = .51$). There was also an effect of visibility, with fewer gestures being produced when participants could not see each other, $F_1(1,72) = 10.361$, $p = .002$, $\eta_p^2 = .126$; $F_2(1,9) = 176.878$, $p < .001$, $\eta_p^2 = .952$; $minF'(1,79) = 9.787$, $p = .002$.

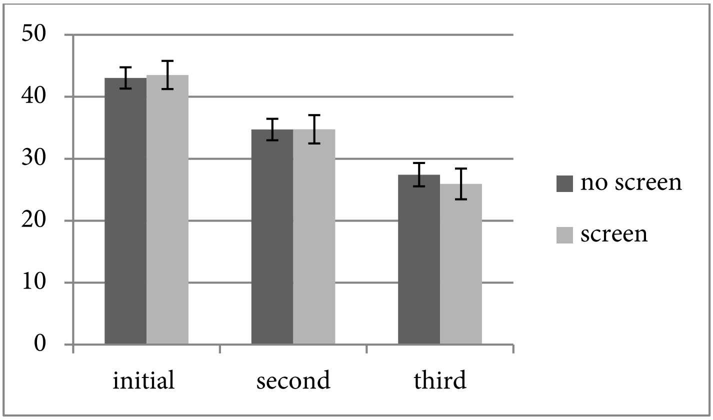

Figure 3.4. Mean duration (in seconds) for each reference, in both visibility conditions. Error bars represent standard errors.

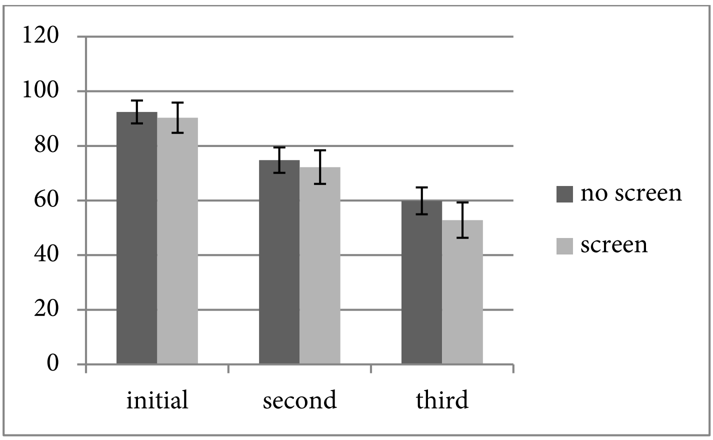

Figure 3.5. Mean number of words for each reference, in both visibility conditions. Error bars represent standard errors.

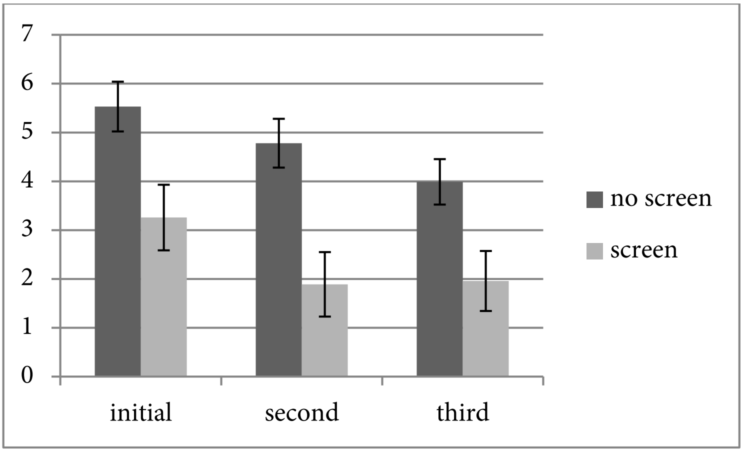

Figure 3.6. Mean number of gestures for each reference, in both visibility conditions. Error bars represent standard errors.

Gesture rate Turning to the two measures of gesture *rate*, we firstly found that there was no significant effect of repetition on the number of gestures per 100 words (see table 3.2), indicating that the decrease in the number of words and the number of gestures, as reported in the manipulation check, is proportionally the same, i.e., number of words and number of gestures decrease to the same extent (as in de Ruiter, et al., 2012). However, for the number of gestures per attribute, we did find an effect of repetition[7] (see table 3.2). The number of gestures per attribute was lower in second references as compared to initial references, and higher in third references as compared to second references, $F_1(2,144) = 21.577$, $p < .001$, $\eta_p^2 = .231$; $F_2(2,9) = 16.346$, $p = .001$, $\eta_p^2 = .784$; $minF'(2,27) = 9.300$, $p < .001$. Post-hoc tests showed that second references differed from both initial and third references (both $p < .05$), whereas initial and third references did not differ significantly from each other ($p = .67$).

[7] We also conducted analyses on the number of attributes per reference, and found that initial references ($M = 11.09$, $SE = 0.31$) contained fewer attributes than second references ($M = 15.37$, $SE = 0.63$), which in turn contained more attributes than third references ($M = 8.82$, $SE = 0.34$), $F_1(2,144) = 93.467$, $p < .001$, $\eta_p^2 = .565$; $F_2(2,9) = 15.084$, $p = .001$, $\eta_p^2 = .770$, $minF'(2,12) = 12.98$, $p = .001$. Post-hoc tests (with Bonferroni correction) showed that all three references differed significantly from each other (all $p < .05$).

Table 3.2. Mean values, standard errors and confidence intervals of the two types of gesture rate: number of gestures per 100 words, and number of gestures per attribute, in initial, second and third references.

Gesture rate	Repetition	Mean (SE)	95% Confidence Interval	
			Lower Bound	Upper Bound
Gestures/100 words	1	4.928 (0.472)	3.986	5.870
Gestures/100 words	2	4.421 (0.467)	3.491	5.351
Gestures/100 words	3	6.046 (1.102)	3.849	8.242
Gestures/attribute	1	.430 (0.040)	.350	.510
Gestures/attribute	2	.227 (0.025)	.178	.276
Gestures/attribute	3	.385 (0.047)	.292	.479

For both measures of gesture rate we found an effect of visibility (see table 3.3). When there was no mutual visibility, fewer gestures per 100 words were produced than when there was mutual visibility, $F_1(1,72) = 17.787$, $p < .001$, $\eta_p^2 = .198$; $F_2(1,9) = 36.065$, $p < .001$, $\eta_p^2 = .800$; $minF'(1,54) = 11.912$, $p = .001$, and likewise fewer gestures per attribute were produced, $F_1(1,72) = 24.974$, $p < .001$, $\eta_p^2 = .258$; $F_2(1,9) = 133.359$, $p < .001$, $\eta_p^2 = .937$; $minF'(1,79) = 21.030$, $p < .001$.

Table 3.3. Mean values, standard errors and confidence intervals of the two types of gesture rate: number of gestures per 100 words, and number of gestures per attribute, in conditions of visibility (no screen) and no-visibility (screen).

Gesture rate	Visibility	Mean (SE)	95% Confidence Interval	
			Lower Bound	Upper Bound
Gestures/100 words	no screen	7.587 (0.703)	6.185	8.990
Gestures/100 words	screen	2.676 (0.928)	.825	4.526
Gestures/attribute	no screen	.515 (0.041)	.434	.596
Gestures/attribute	screen	.180 (0.053)	.073	.286

Finally, as is illustrated in figure 3.7, for the number of gestures per attribute there was a significant interaction between repetition and visibility, $F_1(2,144) = 8.348$, $p = .001$, $\eta_p^2 = .104$; $F_2(2,9) = 6.951$, $p = .015$, $\eta_p^2 = .607$; $minF'(2,29) = 3.793$, $p = .034$, which shows that the effect of repetition, with fewer gestures per attribute in second references, followed by more gestures per attribute in third references, is especially prevalent in the visibility condition.

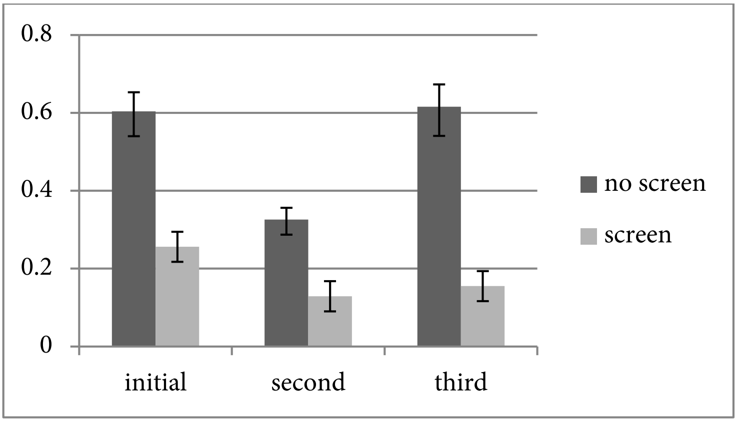

Figure 3.7. Mean number of gestures per attribute for each reference, in both visibility conditions. Error bars represent standard errors.

Gesture form In addition to the gesture rate measures, we analysed several qualitative aspects of the gestures. Table 3.4 shows the mean values and standard errors for these variables in all three references.

The statistical analyses showed that, although the decrease in gesture duration for repeated references was significant in F_1 and F_2, it was not significant in *minF'*, $F_1(2,166) = 3.781$, $p = .026$, $\eta_p^2 = .061$; $F_2(2,9) = 4.577$, $p = .043$, $\eta_p^2 = .504$; $minF'(2,41) = 2.070$, $p = .139$. For gesture size, number of hands and number of repeated strokes, there was a comparable numerical effect, with second references obtaining somewhat lower scores than initial ones, and third references lower still, but these differences were not statistically reliable.

Table 3.4. Overview of mean results (M and SE) for gesture duration (in seconds), gesture size (range 1-4), number of hands (range 1-2, with e.g. 1.70 indicating 70% two-handed gestures) and number of repeated strokes, in initial, second and third references.

	Initial (SE)	Second (SE)	Third (SE)
Gesture duration	1.11 (0.07)	0.92 (0.06)	0.99 (0.07)
Gesture size	3.27 (0.06)	3.17 (0.06)	3.14 (0.07)
Number of hands	1.70 (0.04)	1.67 (0.05)	1.58 (0.06)
Number of repeated strokes	0.16 (0.05)	0.18 (0.04)	0.15 (0.04)

There was no interaction between repetition and visibility for any of these variables, but there was an effect of visibility on gesture duration and gesture size (see table 3.5). Gestures were shorter in duration when there was no mutual visibility, $F_1(1,58) = 6.084$, $p = .017$, $\eta_p^2 = .085$; $F_2(1,9) = 36.161$, $p < .001$, $\eta_p^2 = .801$; $minF'(1,67) = 5.208$, $p = .026$. Gestures produced without mutual visibility were also smaller than with mutual visibility (see table 3.5), $F_1(1,58) = 78.052$, $p < .001$, $\eta_p^2 = .574$; $F_2(1,9) = 154.267$, $p < .001$, $\eta_p^2 = .945$; $minF'(1,50) = 51.828$, $p < .001$.

Table 3.5. Mean values, standard errors and confidence intervals of gesture duration (in seconds), and gesture size (range 1-4), in conditions of visibility (no screen) and no-visibility (screen).

	Visibility	Mean (SE)	95% Confidence Interval	
			Lower Bound	Upper Bound
Duration	no screen	1.147 (.055)	1.036	1.258
Duration	screen	0.873 (.096)	0.681	1.066
Size	no screen	3.654 (.052)	3.549	3.758
Size	screen	2.731 (.090)	2.551	2.912

Summarising the main findings of Experiment I, we found that for gesture rate there was no effect of repetition on the number of gestures per 100 words, but that there was an effect of repetition on the number of gestures per attribute: these were lower in second references than in initial ones, and then increased in third references back to the level of initial references. Lack of visibility caused both gesture rates to be lowered. For gesture form we found no significant effects of repetition, although we did find effects of visibility on gesture duration and gesture size.

Experiment II: Precision judgment of repeated references

In this judgment test participants judged gesture precision, looking at pairs of gestures taken from initial and repeated (third) references, as produced in Experiment I, to see whether there might be more gradient differences in gesture.

Participants

In total, 39 Dutch undergraduates (14 male, 25 female, age range 18-29 years old, M = 20 years 8 months) took part. Twenty participants took part in the visibility condition, and 19 participants in the no-visibility condition, all as partial fulfilment of course credits. The participants had no previous knowledge of and had not taken part in Experiment I.

Stimuli

For the visibility condition, 66 pairs of video clips were selected from the visibility condition of Experiment I. For the no-visibility condition, 31 pairs of video clips were selected from the no-visibility condition of Experiment I. The pairs of video clips contained minimal pairs of gestures with one gesture in each video clip, produced by the same director, illustrating the main shape of the same object. One video clip showed a gesture produced in an initial description of an object, the other video clip showed a gesture produced during a third description of the same object. The order in which the initial and third gestures were presented in the pairs of video clips was counterbalanced over trials. In each trial, a picture of the target object that was described during gesture production was positioned above the video clips (see figure 3.8), and the participants were told that the gestures were produced when describing this particular picture.

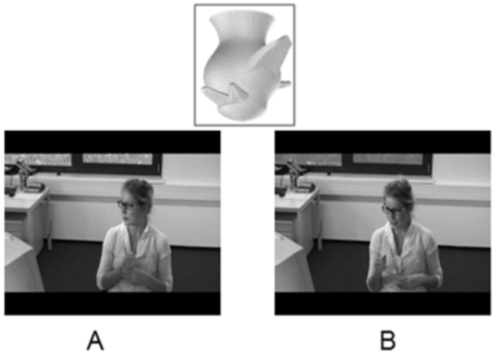

Figure 3.8. Example still of stimulus item in precision judgment experiment.

Procedure

The participants were presented with the pairs of video clips. For each pair of video clips, they had to decide in which video clip they thought the gesture was “the most precise”. It was explained to participants that a gesture “is more precise, for example when it provides more information about the shape of the object or when it is more complex” (English translation of Dutch instruction). Experiment II was a forced choice test, and although repeated viewing of the video clips was possible, participants were asked to go with their first intuition, and repeated viewing hardly occurred. The judgment test took about 20 minutes and was administered without sound.

Data analysis

In each trial, one point was given when an initial gesture was chosen to be the most precise and no points (0) were given when a repeated gesture was chosen to be the most precise. We conducted a binomial test to check for significance (i.e. whether the distribution between 0 and 1 was equal, or not). We looked at the overall number of times that an initial gesture was chosen to be the most precise, as well as at both visibility conditions separately.

Results

A binomial test showed that, overall, initial gestures were chosen significantly more often (in 1085, or 57%, of 1909 cases) than repeated gestures, $p < .001$. This was the case for both visibility conditions; in 765, or 58%, of cases in the visibility condition ($p < .001$), and in 320, or 54%, of cases in the no visibility condition ($p = .039$), the initial gesture was chosen to be the most precise[8]. These results show that participants consider gestures from initial references to be the most precise, regardless of whether these gestures were produced in contexts of mutual visibility or not.

Experiment III: Gesture Interpretation

Finally, in Experiment III, we ask whether repeated gestures, when presented without context, are less ‘intelligible’ than initial gestures. Previous studies on speech (e.g., Bard, et al., 2000) found that words taken from repeated references, when presented without context, were less intelligible. The question is whether a similar process occurs for gesture. To answer this question a final experiment was set up where participants had to watch a selection of gestures taken from Experiment II, and choose which Greeble object was the target associated with the gesture they were shown.

The hypotheses were, firstly, that it is more difficult to choose the correct object when the gesture was produced in a repeated reference (and hence participants will make more incorrect choices), compared to when the gesture was produced in an initial reference, and, secondly, that it is more difficult to choose the correct object when the gesture was produced in a context without mutual visibility.

Participants

Participants were 35 Dutch university students (6 male, 29 female, age range 18-30 years old, $M = 21$ years old,) who took part in the experiment as partial fulfilment of course credits. The participants had not taken part in either Experiment I or Experiment II.

---

[8] In a previous version of the precision judgment experiment the participants were not shown a picture of the target object and were not given additional information about what they should consider to be precise; the effect we found was essentially the same.

Stimuli

The experiment was set up in a 2 x 2 design, with the within subject factors visibility (levels: no screen, screen) and repetition (levels: initial, third). Eighty gestures were semi-randomly selected from the precision judgment experiment, so that they were evenly distributed over the two factors; gestures from contexts with (40 gestures) and without (40 gestures) mutual visibility between the director and the matcher, half of which in turn were taken from contexts of initial and half from third references. To control for individual variation between the directors' gestures, sets of gestures of the same director producing a gesture about the same object (as in the minimal pairs of video clips in Experiment II) were selected. The video clips were ordered semi-randomly, in such a way that video clips showing the same director gesturing about the same object were never presented one after the other. To control for possible learning effects, two reverse stimulus orders were used.

Procedure

The experiment consisted of 80 slides, with one video clip of one gesture on each slide. For each slide, there was a separate piece of paper with two Greeble objects on it, picture A and picture B. The task for the participants was to choose for each video clip whether the gesture in the video clip was produced in a description of object A or in a description of object B. The participants noted down their answers on an answer sheet. One of the two objects that the participants could choose from was always the object that was being described (i.e. the correct answer), and the alternative object was always a Greeble object with a main body shape different from the correct answer. The order of the correct answers (A or B) was counterbalanced over the trials in the experiment. The experiment was preceded by two practice trials to get the participants used to the short video clips.

Participants were given written instructions and the possibility to ask questions. The slide presentation was opened and participants were allowed to go through the slides and the booklet of Greeble object pictures by themselves. The video clips started playing as soon as a new slide was opened and participants were allowed to watch each video clip only once. For each video clip the participants had to choose A or B from the accompanying page in the booklet of object pictures. Participants were encouraged to go with their first intuition, also in cases where they found the task difficult. The experiment took about 20 minutes and was administered without sound.

Data analysis

Each correct answer given by each participant received one point. To test whether participants were better able to pick the correct object, depending on whether a gesture was produced in an initial or repeated gesture which was produced with or without mutual visibility, we conducted chi-square analyses.

Results

In table 3.6 the total scores for all four conditions are shown. Results from the chi-square test of goodness-of-fit showed that there was an equal distribution for initial and repeated gestures, $\chi^2(1) = 1.755$, $p = .185$. There was, however, not an equal distribution for mutual visibility, $\chi^2(1) = 74.360$, $p < .001$. Thus participants were better at selecting the correct object based on a gesture taken from a description in which the director and the matcher could see each other than when the gesture was taken from a description in which the director and the matcher could not see each other, but whether the gesture was taken from an initial or a repeated description had no effect. A chi-square test of independence was conducted to examine the relation between repetition and visibility, and we found no significant relation between the two, $\chi^2(1) = .262$, $p = .609$.

Table 3.6. Scores for number of correct trials, across conditions, in Experiment III.

	No visibility	Mutual visibility	Total
Initial gesture	376	578	954
Third gesture	364	533	897
Total	740	1111	1851

General conclusion and discussion

In this chapter, we studied how speakers gesture during initial and repeated references to hard to describe objects, i.e., Greebles. To this end, we used an adaptation of the director-matcher, referential communication paradigm (e.g., Clark & Wilkes-Gibbs, 1986; de Ruiter, et al., 2012; Holler & Stevens, 2007; Krauss & Weinheimer, 1966), combined with a visibility manipulation such that some participant pairs could see each other (mutual visibility), while others could not. Our findings extend earlier research by providing arguably the largest (in terms of participants) and most comprehensive (in terms of different analyses) study on gesture in repeated references to date.

Earlier research has shown that repeated references in successful communication are different from initial ones, in the sense that they contain fewer words (e.g., Clark & Wilkes-Gibbs, 1986), that these words can be reduced acoustically (e.g., Bard, et al., 2000), and that repetition causes speakers to gesture less (e.g., Levy & McNeill, 1992). Our findings in Experiment I were in line with this, showing that our paradigm worked as intended. Our main foci of attention in the present study were the influence of repetition on two different types of gesture rate (with respect to words and semantic content) and on gesture form.

Repetition and gesture rate

In view of earlier, inconsistent findings, we systematically compared reduction in gesture rate per word with reduction in gesture rate per attribute. We found a small numeric increase comparing the first and the last reference for the gesture rate per word (consistent with the pattern observed by Holler, et al., 2011). However, in our data this difference was not statistically reliable, similar to the findings of de Ruiter, et al. (2012). The similar reduction in repeated references in words and in gestures (causing gesture rate per word to stay the same) thus offers evidence for the "hand-in-hand" hypothesis (So, et al., 2009).

When looking at the gesture rate per attribute, a more nuanced picture emerges. Comparing the first and third reference to a target revealed no differences in gesture rate per attribute, which again appears to be in line with the hand-in-hand hypothesis. However, the second reference is associated with a reduced gesture rate, as compared to the preceding and following one. This drop in gesture rate per attribute is caused by an increase of the number of attributes that are included in the second description, which is not mirrored by an increase in the number of gestures (nor the number of words for that matter). We conjecture that this U-shaped pattern is related to the nature of the task. Describing Greebles is hard — speakers have not been confronted with these objects before, and they do not have a vocabulary ready when they start the director-matching task. This might explain the relatively high gesture per attribute rate during the initial descriptions, and could be interpreted as evidence for the trade-off hypothesis (when speaking gets harder, speakers gesture more, de Ruiter, et al., 2012). However, during the experiment speakers gradually learn which attributes are useful when describing a particular Greeble, and how to convey these efficiently in words and gesture (cf. the reduction in the numbers of words and gestures, which is fully consistent with earlier studies). During the third and final description, speakers use fewer attributes, presumably because they have learned which set of attributes is most helpful in distinguishing the target Greeble from the others, causing a relative increase of gesture per semantic attributes. Interestingly, this pattern is most clearly observed in the mutual visibility condition, which we discuss in more detail below.

Taken together, these results show that it is important to look at both the gesture rate per word and the gesture rate per attribute, since these can reveal subtly different effects. However, it also raises an important question: when should researchers rely on gesture rate per word and when on gesture rate per semantic attribute?

Gesture rate: per word or per attribute?

If there were a one-to-one correspondence between words and attributes, it should not matter how gesture rates are computed. However, although words and attributes are obviously related, it is easily seen that they do not necessarily stand in a one-to-one relationship. On the one hand, some attributes require more words to be realized in a referring expression than others. In general, it can be assumed, for instance, that premodifiers (i.e., adjectives occurring before the head noun) consist of fewer words than postmodifiers (such as prepositional phrases or relative clauses), and whether an attribute is expressed as a pre- or a postmodifier is more or less coincidental and may differ from one language to another (see e.g., Goudbeek & Krahmer, 2012, for discussion). In addition, utterances may include hedges ("I think") and fillers ("uh"), which do not have a direct counterpart in the semantic representation of the description; it is conceivable that such non-attribute related words occur more often in initial than in repeated references, which might complicate reduction patterns. In a somewhat similar vein, it is often assumed that gestures encode meanings in a globally and non-compositional fashion, with one gesture expressing various meanings (e.g., Galati & Brennan, 2014; Hostetter & Alibali, 2008; McNeill, 1992). Hostetter and Alibali

(2008, p. 501), for example, discuss the English example "She climbed up the ladder" produced with a single gesture consisting of wiggling fingers moving upwards horizontally, thereby combining various meaning components. It is interesting to observe that the possibilities of gesture to express multiple meanings simultaneously may differ with task and domain; in the Greeble dataset we tend to find that a single gesture expresses a single semantic attribute. For all of these reasons, the relation between meanings on the one hand, and words and gestures on the other, is not straightforward. By only computing gesture rate per word, one risks missing important information (such as the U-shaped pattern in gestures per attribute that we observed).

As we have seen, with some notable exceptions, most gesture researchers only compute gesture rate per word, presumably, at least to some extent, because it is easier and less time consuming. Defining a semantic representation for a task can be complicated, in particular when the task is relatively open ended. An advantage of Greebles, and one of the reasons why we opted for using them in this study, is that their body shapes and protrusions differ in predictable ways, which facilitated the development of a semantic representation. Our data collection is thus "semantically transparent" (in the terminology of Van Deemter, Gatt, van der Sluis, & Power, 2012) in the sense that we know the semantic attribute-values of the target Greebles as well as of all distractors, thus enabling semantic annotation of speech and the subsequent computation of gesture rates per attribute.

In general, if time and resources allow, if a clear semantic representation for the task can be defined, and if in said task the relation between attributes and words is not one-to-one (which might especially be the case in complex domains), researchers are advised to report both gesture rate per word and per attribute. In addition, as we shall discuss below, this distinction also has implications for models of speech and gesture production. Finally, it may be worth noticing that observations such as the above (the nature of the task; how meanings are expressed in words; how much information can be conveyed by a single gesture) may also partly explain why earlier research revealed conflicting results when looking at gesture rates per word, as described in the introduction of this chapter.

Repetition and gesture form

Besides gesture rate, we also studied whether the gestures produced during repeated references are different in their realization from comparable gestures produced during initial references, asking whether there are discrete differences in form and/or whether differences are more gradient in nature, with repeated gestures appearing less “precise” than initial ones.

For this purpose, in Experiment I we compared gestures expressing the same property of a Greeble (its general form or body shape). When looking at gesture form, we found that gestures during initial references numerically lasted somewhat longer than gestures produced during repeated references. However, these findings, while significant in F_1 and F_2, were not significant in the *minF'* analyses, and hence cannot be considered statistically reliable. We did find clear effects of visibility, with gestures that could be seen by the addressee lasting significantly longer and being bigger than ones that were not visible.

We also asked, in two different ways, whether there were gradient differences between initial and repeated gestures. One judgment study (Experiment II) presented participants with minimal pairs of gestures, taken from an initial and a repeated reference, and asked which of the two was more “precise” for a particular Greeble object. The results of this judgment study revealed that initial gestures were indeed perceived as being more precise than repeated ones. These findings are consistent with the observations of Gerwing and Bavelas (2004), although it is important to note that their findings were obtained by two annotators comparing larger stretches of dialogue. Another study (Experiment III) presented participants with a video clip of one gesture (taken from an initial or a repeated reference, produced with or without a screen), and they were asked which of two Greeble objects was the one the speaker was talking about. The results showed that gestures which were produced when the speaker knew that these would not be seen (in the no-visibility condition) were, as expected, less ‘intelligible’ than gestures taken from contexts of mutual visibility. However, participants did not perform better on this task when viewing gestures from initial descriptions as compared to when viewing gestures from repeated descriptions.

In general, when looking at repetition and gesture form a clear and consistent picture emerges. Gestures produced during initial descriptions are judged to be more precise than those produced during repeated descriptions, even though the manual coding does not reveal reliable differences. This suggests that the reduction is gradient, and that the form of the gesture (e.g., whether it is produced with one or two hands) generally does not change between initial and repeated references. Moreover, even though they are reduced in precision, we found that gestures in repeated references are still effective at communicating information; when participants are asked to decide which target object is being referred to based on just one gesture (a hard task!), they can do this roughly equally well when the gesture was produced during an initial or a repeated reference. The resulting picture is conceptually very similar to the way words are articulated when referring to initial or new compared to repeated or given information (e.g., Bard, et al., 2000). However, visibility is an important factor in all these analyses.

On the effects of visibility

In general, we found clear effects of visibility. A reduction due to lack of mutual visibility was found for the overall number of gestures, as well as for both measures of gesture rate. Lack of mutual visibility also had an effect on general gesture form, with speakers in that case producing smaller gestures that were also shorter in duration (Experiment I). We also found that gestures produced when there was no mutual visibility were less intelligible (Experiment III). It is interesting to observe that while gesture and speech in our data seem to go hand in hand when considering the effects of repetition (at least when considering gesture rate per word), this does not appear to be the case when considering the effects of visibility. Lack of visibility impacts gesture but not speech; participants produce substantially fewer gestures when separated by a screen, but the same amount of speech with and without visibility.

Earlier gesture studies using a visibility design have led to sometimes conflicting results (see e.g., Alibali, et al., 2001; Bavelas & Healing, 2013, for discussion). Interestingly, Alibali, et al. (2001, p. 184) when discussing conflicting effects of visibility on gesture rate per word, observe that "[a]mong visibility studies, those that have demonstrated effects of visibility on gesture production (e.g., Cohen, 1977; Cohen & Harrison, 1973; Krauss, Dushay, Chen, & Rauscher, 1995) used tasks with high spatial content (giving directions, describing abstract figures), which may have elicited primarily representational gestures". This suggestion nicely ties in with our findings obtained with the Greeble objects, which are both highly spatial and abstract. Bavelas and Healing (2013) argue that in a number of earlier visibility studies -including Alibali, et al. (2001) and Mol, et al. (2009) - the visibility manipulation may have confounded with addressee responsiveness. Since we based this part of our design on the aforementioned studies, this criticism may be applied to our study as well (although it is interesting to observe that Alibali, et al., 2001, p. 182, discuss and discard this possible alternative explanation of their results). In any case, this issue certainly warrants further study.

Importantly, Bavelas and Healing (2013, p. 79) stress that gesture rate is not the best way to assess visibility effects, and write: "A closer look at how speakers use their gestures reveals that visibility affects many aspects of gestures including the kinds of gestures, their size, location, and relationship to words. All of these differences seem to be done for the addressee's benefit." Our results on gesture form are perfectly in line with this. This suggests that many of the gestures produced by speakers in the mutual visibility condition were indeed designed with the addressee in mind, which has implications for models of speech and gesture production.

Implications for models of speech and gesture production

Over the years, various models of speech and gesture production have been proposed, including Krauss, Chen and Gottesman's (2000) Process model, Kita and Özyürek's (2003) Interface model, de Ruiter's (2000) Sketch model, and McNeill and Duncan's (2000) Growth Point theory (see e.g., Chu & Hagoort, 2014; Hostetter & Alibali, 2008; Wagner, et al., 2014, for recent comparisons and discussion). These models all seek to describe how speakers produce multimodal utterances and are concerned with issues such as the timing and integration of gesture and speech, and the role that gestures play in communication. Our present findings are relevant for both of these issues.

Many of the aforementioned models take Levelt's (1989) 'blueprint for the speaker' as their starting point. In this blueprint, speech production is assumed to be a modular process involving three main, consecutive stages. A speaker first has to decide what she wants to say, a decision made in the conceptualizer stage, and resulting in a semantic "preverbal message". Notice, importantly, that this is the stage in which speakers in our Experiment I decide which attributes of the target Greeble to include in their referring expression, based on how helpful they are in distinguishing the target from the other Greebles (cf., Gatt, Krahmer, van Deemter, & van Gompel, 2014; Olson, 1970). In a second stage, known as the formulator and involving lexical retrieval and grammatical encoding, the words of the actual utterance are planned, based on the preverbal message. Finally, in the third stage, the utterance plan is phonologically encoded and articulated, resulting in overt, auditory speech. Models of gesture production typically involve two stages: a Motor Planning stage, sometimes referred to as the Gesture Planner or the Action Generator, during which the motor instructions are produced, and a Motor Execution stage, during which these programs are executed, resulting in overt, visible gestures (Chu & Hagoort, 2014; Wagner, et al., 2014).

The main difference between the various extensions of Levelt's (1989) model concerns the exact points of interaction between the speech and gesture production processes. All agree that there is early interaction, with a joint origin for speech and gesture, either in working memory or in the conceptualizer stage. However, some models assume that after this initial interaction, the two processes develop independently —or "ballistically", in terms of Levelt, Richardson and La Heij (1985). This is true, for instance, for the Sketch model (de Ruiter, 2000), while both the Process model (Krauss, et al., 2000) and the Interface model (Kita & Özyürek, 2003) assume that there is further interaction during later stages of the production process. Krauss and colleagues, for example, argue for interaction between the Motor system and the formulator, to account for their observation that the production of gestures may facilitate lexical retrieval. McNeill's Growth Point theory makes the strongest claim concerning interaction, by arguing that speech and gesture are two inseparable parts of a single process (rather than two interacting processes), jointly arising from a single idea (growth point).

Our data do not allow us to draw conclusions about the underlying representations from which gestures arise, but it seems plausible that visual inspection of the Greebles allows speakers to select distinguishing visual features of the target to be expressed (say the "Dunth" protrusion in figure 3.1), comparable to how the Sketch Generator in de Ruiter's (2000) model accounts for this. At this early stage, the "Dunth" attribute becomes part of the pre-verbal message, and since our participants do not have words for Greeble "Dunths", they may express the spatial properties of this shape in gesture, combined with, say, a phrase such as "a protrusion shaped like this".

Interestingly, the two different gesture rates we reported in this study can be seen to operate at two different levels in models of speech and gesture production - the gesture rate per attribute relates to the early interaction of speech and gesture at the pre-verbal level of conceptualisation, while the gesture rate per word is more directly related to later interactions, at the level of the formulator, where words arise. Given that all models assume early interactions between speech and gesture, our gesture per attribute findings do not clearly differentiate between the models. However, the gesture rate per word findings, generally suggesting that speech and gesture go "hand-in-hand", are arguably more difficult to explain for a "ballistic" model, than for an interactive model assuming that the production of speech and gesture also interact at the later stages of speech production, such as McNeill and Duncan's (2000) Growth Point theory. Also our suggestion that with repetition the qualitative reduction in gesture production is comparable to the acoustic reduction in speech production is consistent with this perspective.

A second, partly related issue concerns the question whether gestures communicate information, and whether they were intended as such by the speaker. Models of gesture and speech production are usually not explicit about whether gestures communicate information to an addressee, which is perhaps not surprising given that they are models of the *speaker* (a limitation also discussed by, among others, Mol, Krahmer, Maes, & Swerts, 2012). Still, our findings clearly show that addressees may obtain information from gesture, since in Experiment III we found that participants could determine which of two Greebles was being described, at least based on certain, single gestures. Presumably, this is because these gestures tended to be not redundant with the accompanying speech, but really added information to it, for instance, about the precise form of the target Greeble described by the speaker (see e.g., Singer & Goldin-Meadow, 2005 for comparable observations in a very different setting, namely child learners). Moreover, the finding that in Experiment III participants performed better when seeing gestures produced in the mutual visibility condition strongly suggests that these gestures were intended by the speaker to be communicative, as we observed above. This is in contrast with the Process model (Krauss, et al., 2000), which assumes that gestures are not part of the speaker's communicative intention, but rather have a facilitative function for the speaker herself (since they may help with lexical retrieval).

Future research

There are several avenues of research that could be addressed in future work. Firstly, future work could study whether a similar reduction process occurs for different types of gestures, such as deictic and beat gestures. In the present study the gestures that were produced were almost exclusively iconic gestures. Arguably, this was due to the affordances of the stimuli, which were fairly abstract and spatial, and caused participants to produce many iconic gestures. Although at least one study has been conducted taking into account the use of deictic gestures in repeated references (de Ruiter, et al., 2012), this study focused on gesture rate, and we do not yet know whether there may be any effects on the form of deictic gestures. A task explicitly designed to elicit deictic gestures (or beat gestures, for that matter) in repeated references could address this point, and clarify whether we can generalise the results obtained in the present study to other types of gesture.

Secondly, in the present study we purposely used the individual speaker as the "unit of analysis" (Bavelas & Healing, 2013, p. 65), since we were primarily interested in speakers' possible reduction of references as a function of repetition, and we did not want other factors (apart from visibility) to play a role. Therefore, in this study there was little interaction between the director and the matcher. It is an interesting question whether the effects that we found can also be observed when there is more free interaction between the director and the matcher. It is conceivable that in the case of more free interaction between participants stronger reduction effects are observed, since explicit feedback from the matcher could help in creating truly shared conceptual pacts (Brennan & Clark, 1996), which often consist of reduced references as they develop.

Thirdly, although the task of describing the abstract Greeble stimuli may seem quite difficult, and speakers did not have a vocabulary ready (as mentioned above), they were still successful in conducting the task, and hardly any errors were made by the matchers. Future work could study whether a reduction process as found in the present study can be generalised to all kinds of repeated references, or whether only successful repeated references, such as in the present study, are reduced in this manner. It can be argued that the repeated references in our study were reduced, not simply because they were repeated, but (also) because their successfulness indicated that reduction was possible without causing communication problems. However, what if producing a reference does not lead to successful object identification? If the addressee indicates that a particular reference description is not sufficient, what will happen in subsequent references? Will reduction still occur, or will the speaker not reduce a repeated reference, but ‘expand’ it in some manner? Assuming that speakers want to be communicatively efficient (Aylett & Turk, 2004), reducing a repeated reference when previous references were unsuccessful, and thereby providing less information to the addressee, might not be the best strategy, and speakers may choose to do otherwise.

A fourth possible avenue for further research is also related to whether the findings from the present study can be generalised to other contexts in which repeated references occur. As discussed in the introduction, previous studies found reduction in repeated references in speech, and other previous studies, as well as the present study, found reduction in repeated references in gesture. A question is whether a similar reduction process occurs in sign language. Sign language is interesting to study in this respect, since in sign language the visual domain is the dominant domain that is available for interaction, unlike in spoken reference production. Will repeated references in sign language be reduced in a manner comparable to speech, to gesture, or in a sign-specific manner? A first study in this direction (Hoetjes, Krahmer, & Swerts, 2014a, see chapter 4 of this thesis) suggests that repeated references in sign language are in fact reduced in a similar way to speech and gesture.

Acknowledgments

We would like to thank Elsa Jonkers, Kristel Bartels, Joost Driessen, and Manon Yassa for practical help in collecting the data, Nick Wood and Bas Roset for technical support, Judith Holler and Jette Viethen for helpful comments and Martin Pickering, Jan de Ruiter and two anonymous reviewers for comments on a previous version of this paper. Earlier versions of this study were presented at the Optimal communication colloquium at Radboud University Nijmegen (2011), and at the 33rd Annual Conference of the Cognitive Science Society, Boston (2011).

4

Do repeated references result in sign reduction?

Abstract

Previous research on speech and gesture has found that repeated references are often linguistically reduced in terms of, for example, the number of words and the acoustic realization of these words, compared to initial references. The present study looks at the production of repeated references by 14 signers of Sign Language of the Netherlands (NGT). Participants had to describe figures to an addressee, who had to pick the correct figure from a large group of figures. Several figures had to be described several times. The question was whether there would be reduction in the repeated references. We found systematic effects of repetition in that repeated references were shorter, contained fewer signs, and shorter signs than initial references. Moreover, in order to measure sign precision, a perception test was used where participants had to judge, in a forced choice task, which sign they considered to be the most precise, looking at 40 pairs of video clips with signs produced in either initial or repeated references to the same object by the same signer. We found that non-signing participants (but not signing participants) consider signs produced during repeated references to be less precise than the signs produced during initial references. Taking together these results suggest that a similar reduction process in repeated references occurs in NGT as has been found previously for speech and gesture.

This chapter is based on:

Hoetjes, M., Krahmer, E. & Swerts, M. (2014) Do repeated references result in sign reduction? *Sign Language & Linguistics*. *17*(1), 56-81.

Introduction

Variability is ubiquitous in speech production, with words never pronounced in exactly the same way more than once. For example, someone might first pronounce the phrase 'of course' slowly and precisely, followed by an instance where it is pronounced quickly, less precise and more like 'fcourse' (Ernestus & Warner, 2011). This example of language variability shows that speech can be reduced (in this case by shortening and merging words). While various studies have looked at reduction in speech, reduction in signs remains largely unexplored, partly because it was unclear how sign reduction can be measured. The present study addresses this point.

Little experimental research has been done on reduction in sign language. Tyrone and Mauk (2010), as a notable exception, looked at sign lowering (in American Sign Language), which can be seen as an instance of reduction. Sign lowering, according to Tyrone and Mauk, occurs when "a sign [is] being produced in a lower location than in the citation form" (Tyrone & Mauk, 2010:317). In their study they found that several phonetic factors, such as production rate, influence the exact location of the produced sign. The question is whether there may be other factors causing signs to be reduced as well, and whether there are, apart from location, other ways in which signs can be reduced. Previous research on speech (e.g. Bard, et al., 2000; Fowler, 1988; Fowler & Housum, 1987) has shown that when speakers produce a repeated reference, that this repeated reference is often shorter and uttered less clearly, and thus becomes less intelligible for the listener than an initial reference. In other words, the repeated reference is reduced compared to the initial reference. In the present study, we will combine these two strands of research (on sign language and on speech) by looking at reduction in signs in repeated references. Repeated references are a suitable domain to study sign reduction since they are a naturally occurring phenomenon, produced whenever the same object is described more than once, but can also be elicited in a controlled manner, especially in an experimental setting. The present study will look at repeated references in signs produced by speakers of Sign Language of the Netherlands (NGT). Speakers of NGT produced repeated references in an experimental setup, allowing measurement of several aspects of signs that may be reduced, such as sign duration and sign precision.

Reduction in spoken repeated references

In conversation, people often produce referring expressions to describe objects in the world around them, for example when describing a building that they recently visited. The production of repeated references occurs when people refer to the same object more than once in the conversation. In the example case of a building being described, one can imagine that in an initial description many details of the building are mentioned, such as its exact location, orientation, size and colour. In a repeated description in the same conversation many of these details may be left out because the conversational partner already knows which building is being discussed. This means that descriptions of the same object can range from "the tall brown building at the back of the university campus" to "the building". This process of reduction in repeated references has been described in detail in chapter 3 of this thesis.

Previous research has found that in speech, repeated references are often reduced in at least two ways (Aylett & Turk, 2004; Bard, et al., 2000; Brennan & Clark, 1996; Clark & Wilkes-Gibbs, 1986; Fowler, 1988; Fowler & Housum, 1987; Galati & Brennan, 2010; Lam & Watson, 2010). Firstly, repeated references to the same target object usually contain fewer words than initial references (Clark & Wilkes-Gibbs, 1986; Galati & Brennan, 2010), as can be seen in the example above where in the repeated reference the information on the building's size, colour and location is omitted. Brennan and Clark (1996) argue that this is due to the fact that people establish so-called "conceptual pacts" as more common ground is established over the course of the conversation.

Secondly, repeated references often contain repeated words, and we know from earlier studies that these are often reduced acoustically (Aylett & Turk, 2004; Bard, et al., 2000; Fowler, 1988; Fowler & Housum, 1987; Lam & Watson, 2010). This acoustic reduction may be due to the fact that repeated references can be claimed to be (partly) redundant, since (some of) the words have already been mentioned before. Several decades ago, Lieberman (1963) claimed that redundant words are shorter and perceived as less intelligible when taken out of context and presented to listeners. Fowler and colleagues (Fowler, 1988; Fowler & Housum, 1987) found that repeated, redundant, words are indeed shortened. Samuel and Troicki (1998) also showed that redundant speech is articulated less clearly, and, more recently, Aylett and Turk (2004), in their work on the smooth signal redundancy hypothesis, found an inverse relationship between redundancy and duration, with more redundant speech having a shorter duration. Research by Lam and Watson (2010) provided additional evidence that repeated references have reduced prominence and are also reduced in duration, compared to initial references (see also Bell, Brenier, Gregory, Girand, & Jurafsky, 2009). Words from repeated references, when taken out of context and presented to a listener, have also been found to be less understandable for the listener because their pronunciation is less clear in repeated references than in initial references (Bard, et al., 2000; Galati & Brennan, 2010). This reduction in repeated references that has been found in research on speech can also be related to previous work on the influence of discourse status on the form of referring expressions (Gundel, Hedberg, & Zacharski, 1993), in that repeated references are more likely to be realized in the form of more attenuated expressions (e.g., a pronoun instead of a description).

It is thought that reduction in referring expressions may be due to speakers' efficiency, in production and planning processes (Arnold, 2008; Arnold, Kahn, & Pancani, 2012; Bard, et al., 2000; Bard & Aylett, 2005; Ferreira, 2008), and in communicative strategies (e.g. Aylett & Turk, 2004; Fenk-Oczlon, 2001; Lieberman, 1963; Lindblom, 1990; Zipf, 1936). The use of communicative strategies, with speakers as efficient language users, has been demonstrated in a range of studies (for an overview, see Jaeger & Tily, 2011). Back in 1936, Zipf proposed his Principle of Least Effort, which states that language users prefer to take the least effort necessary to get a message across. Shannon's noisy channel model (1948) can also be related to this reduction process, where, given the context, the more probable a word is, the more likely it is to be reduced in its linguistic form. Lindblom (1990), in his theory of hyper- and hypo-articulation, claims that speakers adapt to the listener's needs, meaning that redundant speech is reduced as long as 'sufficient discriminability' remains. As mentioned above, more recent work on acoustics, by Aylett and Turk (2004), among others, has found that predictable words are indeed reduced, at least with regard to duration. Jaeger (2010) proposed the hypothesis of Uniform Information Density (UID), which states that "speakers prefer utterances that distribute information uniformly across the signal (information density)" (Jaeger, 2010:25)[9]. What this means

[9] One of the reviewers of the article on which this chapter is based suggested that the idea behind the UID hypothesis may be related to another principle of 'cognitive economy', namely Menzerath's law (Altmann, 1980), which roughly states that "the bigger the whole, the smaller the

is that elements of an utterance with a relatively high information value, for example due to the fact that the element is new or specifically important in the conversation, are lengthened. Likewise, elements with relatively lower information value, for example because the element contains old information and/or is not that important for successful communication, are shortened. This way, the amount of information that is transmitted in the utterance becomes more uniform and optimal for speaker and addressee. These ideas of language efficiency can also be considered to be in line with Grice's (1975) Maxim of Quantity, which states that speakers make their "contribution as informative as required (for the current purpose of the exchange)" and proposes to speakers to "not make your contribution more informative than is required".

It can be argued that the reduction in repeated references that previous studies have found is due to the abovementioned processes: when speakers produce repeated references, they fully reproduce those (auditory) aspects of the referring expression that contain important or new information and are necessary for quick target identification. The less informative and more predictable aspects of the referring expression may be omitted, leading to reduced references.

Reduction in visual repeated references: gesture and sign language

The idea that predictable linguistic material is reduced has been applied to several aspects of speech communication, such as syntax (Jaeger, 2010) and phonetics (Bard, et al., 2000). Taking into account that communication not only involves verbal aspects such as syntax and phonetics, but can also contain or consist of visual aspects such as gestures (Kendon, 2004; McNeill, 1992) or signs (Stokoe, 1960), we may wonder whether a reduction process such as described above also occurs for the visual domain.

Relevant previous research on gesture (discussed in more detail in chapter 3) has looked at the effect of common ground (Gerwing & Bavelas, 2004; Holler & Wilkin, 2009) and repeated references (de Ruiter, et al., 2012; Hoetjes, Koolen, Goudbeek, Krahmer, & Swerts, 2011; 2015, see chapter 3 of this thesis) on gesture production, albeit with somewhat inconclusive results. Gerwing and Bavelas (2004) found that gestures that were produced when there was common ground were less complex, less

---

parts". An example is that "words composed of a high number of syllables tend to be composed of a "relatively" low number of phonemes" (Fenk & Fenk-Oczlon, 1993, p. 11).

informative and less precise than gestures produced when there was no common ground. Holler and Wilkin (2009) found that utterances contained "less semantic information when common ground exists", while gestures appeared to "carry a greater communicational weight due to a higher gesture rate" (Holler & Wilkin 2009:285). When we look at repeated references, de Ruiter et al. (2012), when testing their trade-off hypothesis, found that repetition did not affect gesture rate (in number of gestures per 100 words). In chapter 3 of this thesis we found that both speech and gesture were affected in repeated references. Speech was reduced with regard to semantics, number of words and overall duration of the referring expression; gestures were reduced with regard to their absolute number, but did not increase in their rate (in number of gestures per 100 words).

What most of the studies looking at gesture have in common is that they take two modalities into account, both speech and gesture. Looking at both modalities is inherent to co-speech gestures since gestures are closely integrated with speech, both on a semantic, temporal and a pragmatic level (Kendon, 2004; McNeill, 1992). The question for these types of research is often what the exact relationship between speech and gesture is and what the role of each modality is in the discourse (Kelly, Manning, & Rodak, 2008; Krahmer & Swerts, 2007). When we consider sign language, research naturally tends to focus mainly on one modality, the visual modality. There has been a range of research on phonological and phonetic aspects of sign language (Brentari, 1998; Crasborn, 2001; Johnson & Liddell, 2010; Liddell & Johnson, 1989; Sandler, 1989; Sandler & Lillo-Martin, 2006; Schembri, et al., 2009; Tyrone & Mauk, 2010; van der Hulst, 1993), starting with Stokoe's seminal work from 1960, proposing that signs in sign languages consist of three main parameters (handshape, location and movement, see Stokoe, 1960). Few studies have looked at sign language from the perspective of efficient language use, although some studies discussed the efficient use of the different modalities when producing sign language as compared to speech (Gee & Goodhart, 1988; Klima & Bellugi, 1979; Leuninger, Hohenberger, Waleschkowski, Menges, & Happ, 2004). More particularly, these studies suggest that due to their differences in modality, speech, on the one hand, tends to consist of many small chunks, each containing relatively little information, whereas sign language usually consists of fewer but bigger chunks, containing more information. This effect of modality has also been related to differences between speech and sign in production speed and processing manner (Brentari, 2002; Klima & Bellugi, 1979; Leuninger, et al., 2004).

In light of efficient language use, it is interesting to see how signs behave with regard to reduction in repeated references. In particular, we may wonder whether signs are reduced in ways which are comparable to speech and/or to co-speech gestures. On the one hand, considering that signs, like words, usually convey lexical meaning, it might be the case that reduction in sign is similar to reduction in speech, for example with regard to the semantics that are expressed. On the other hand, signs, unlike words but like co-speech gestures, are a means of communication in the visual domain, and there may be aspects of reduction that are modality-specific and thus alike between signs and co-speech gestures. Of course, it could also be the case that signs are not reduced in a way comparable to speech or to co-speech gestures, but that signs, if they are reduced, are reduced in a sign-specific manner.

Only a handful of previous studies have looked at reduction in sign language. Tyrone and Mauk (2010), studying phonetic reduction in American Sign Language, looked at the production of the sign WONDER in two phonetic contexts and at three signing rates. Their results show that sign lowering (which can be seen as a form of efficiency) occurs with increasing signing rate and can (but does not necessarily) occur in specific phonetic contexts. Another experimental study, by Mauk, Lindblom and Meier (2008) focusing on undershoot (a phenomenon comparable to reduction, which occurs "when an expected phonetic target is not achieved [...]" 2008:4) in American Sign Language also found that signing rate had an effect on the exact location in which a sign was produced, with this effect differing depending on the linguistic context. Other studies on variation in sign language, by Schembri and colleagues (2009) and by Russell, Wilkinson and Janzen (2011), looked at naturally occurring data and also found that sign location can vary, with signs being produced at lower locations than their citation form. However, none of these studies takes repetition into account as one of the factors influencing sign production.

Present study

In the present study we look at signs of Sign Language of the Netherlands (NGT), to see whether reduction in repeated references, as previously found for speech and gesture, also occurs in sign language. Considering that NGT is a fully fledged sign language and behaves in many respects like a spoken language, we hypothesize that, as in speech, reduction in repeated references will occur. The question is of course how reduction in signs can be measured. Considering the fact that the aim is to compare possible reduction in sign language with reduction in speech and gesture, we measure reduction by combining methods that have been used previously in studies on speech and on gesture. We will look at sign characteristics that we consider comparable with some of the aspects of speech that have been studied previously when looking at reduction (as discussed above), namely number of words, utterance duration and word duration. We will also take precision into account, which has been done in previous studies (as discussed above) on gesture. Therefore, in the present study on sign language we use methods that can be applied both to sign language and to spontaneous co-speech gestures. We will analyse the number of signs, the utterance and sign duration and sign precision. We conducted a production task to analyse the number of signs, utterance duration and sign duration. Following the work presented in chapter 3 of this thesis, we conducted an additional perception task to analyse sign precision. Details of both production and perception tasks are given below.

Experiment I: production experiment

To study reduction in repeated references in Sign Language of the Netherlands (NGT), a data set was created consisting of recordings of participants taking part in a director-matcher task (as in the previous chapters of this thesis). In this task, the director had to describe a number of objects in such a way that the matcher could identify them from a range of similar looking figures. In the stimuli, there were several figures that had to be described multiple times, leading to repeated references to the same item.

Participants

The director-matcher task was done by a total of 14 signers of NGT. The group of participants consisted of 5 male and 9 female speakers, with an average age of 46 years old (range 26-60 years old). Of the 14 participants, 9 were deaf since birth. The average length of time that the participants had been signing NGT was 23.5 years (range 2-50 years). Two participants learned NGT from birth, three learned NGT before the age of 5, and 9 participants learned NGT after age 10. The participants who had been signing NGT the longest were not necessarily the signers who were born deaf. Participants took part twice in the experiment; first they were randomly assigned the role of either director or matcher and they switched roles after doing the experiment once, so that each participant acted as director once.

Stimuli

Two picture grids, each containing 16 pictures, were used by each director. For each picture grid used by the director, an alternative grid was constructed for the matcher, containing the same items as on the director's picture grid, but for the matcher the items were numbered and presented in a different order (for example grids, see figures 4.1 and 4.2 below). The picture grids showed either pictures of people, or of furniture items. The two different domains (people and furniture) were used since previous studies on referring expressions have shown them to be efficient domains for making people produce referring expressions (Koolen, Gatt, Goudbeek, & Krahmer, 2011; Van Deemter, et al., 2012; Van der Sluis & Krahmer, 2007). The items of the furniture picture grid were the same as those used in the TUNA and D-TUNA corpus (Koolen, et al., 2011; Van Deemter, et al., 2012), the items of the people picture grid were inspired by the items from the people domain in these same corpora but consisted of better quality pictures in full colour.

Figure 4.1. Example of a people picture grid. The picture with the square surrounding it is the target object of that particular trial.

Figure 4.2. Example of a furniture picture grid. The object with the square surrounding it is the target object of that particular trial.

Each picture grid was used for 15 trials, adding up to a total of 30 trials for each director. For the first 15 trials, one of the people picture grids was used, for the last 15 trials one of the furniture picture grids was used. Since the participants would do the experiment twice, once in the role of director and once in the role of matcher, two sets of picture grids (so four picture grids in total) were used, with different pictures on each picture grid. A participant would see one of a set of picture grids when taking part in the role of the director and the other set of picture grids when taking part in the experiment in the role of the matcher. In each trial, there was one target object (marked by a red square), which was surrounded by 15 distractor objects, and which had to be described by the director. The crucial manipulation in the task was that several pictures had to be described repeatedly: in each of the picture grids, two pictures had to be described twice, and two pictures had to be described three times. Repeated references to the same object were never one straight after the other, which means that descriptions of other objects were given in between the initial and repeated descriptions of the critical objects. The use of different sets of picture grids means that although a matcher would become familiar with the type of pictures of people and furniture used in the experiment as they were described by the director, he or she would have to describe a new set of pictures of people and furniture once the roles were switched. Throughout the experiment, it was clear to the directors that several pictures had to be described repeatedly. An example of an initial object description (in this case of a large red chair, as shown in figure 4.2) can be seen in (1), where (Dutch) words in capitals represent the NGT signs, which are translated literally in the second line, and paraphrased in English in the third line.

(1) STOEL, ROOD, NIET LINKS, SCHUIN RECHTS, BEETJE GROTER
CHAIR, RED, NOT LEFT, SIDEWAYS TO.THE.RIGHT, LITTLE.BIT BIGGER
'a red chair, not positioned to the left but sideways to the right, one that's a little bit bigger'

For the purpose of the current analyses, the first and third descriptions of the four objects that had to be described three times were annotated and analysed. The items that were described twice as well as the second descriptions of the items that were described three times were not analysed. The four objects that were analysed were never described in the first or last trial within a set of 15 trials of one picture grid and there were always at least two trials in between trials dealing with the same object. The focus on these initial and repeated descriptions means that the current analyses are based on a data set which consists of eight descriptions (one initial and one repeated description for two pictures from each domain grid) for each of the 14 participants (directors), leading to a total of 112 object descriptions.

Procedure

The director and the matcher faced one another across a table. A camera was positioned behind the matcher filming the upper body and hands of the director (see figure 4.3 for an example).

The director had a laptop screen to her side and the matcher had a picture card in front of him. The director and matcher could see each other directly, but could not see each other's screen or card. The participants were given written instructions and had the opportunity to ask questions to the experimenter, after which the experiment started. The director was then presented with a trial on the computer screen (as in figures 4.1 and 4.2).

Figure 4.3. Camera view of the director, camera is positioned behind the matcher.

The director was asked to provide a description of the target object in such a way that the matcher could distinguish it from the 15 distractor objects. The matcher had a picture card filled with 16 numbered objects (see figure 4.4 for an example) in front of him, which was not visible to the director. The matcher's card showed the same objects as on the director's screen, but these objects were ordered differently for the director and the matcher and only the matcher's objects were numbered. The difference in ordering meant that the director could not use the location of the target object on the grid as part of the description, and this was explicitly communicated to the participants. The difference in ordering also meant that the matcher could not use the director's eye gaze on the screen as a potential cue as to which item was being described. Based on the director's target description, the matcher had to write down the number of the object the director was describing. Once the correct object was found, the director went on to the next trial.

The matcher could make clear to the director whether the target object had been found or not, but there was no free conversation between the director and the matcher, following similar instructions in Alibali et. al (2001) Mol et. al (2009) and Hoetjes et. al (2015). This lack of conversation means that our analyses are based on complete, uninterrupted descriptions given by the director. After 15 trials from the people domain, the director was shown a second grid containing 16 objects from the other domain (the furniture domain), and the matcher was presented with a new card showing these objects (again in a different order from the order in which they were presented to the director and only numbered for the matcher, as in figure 4.4). The participants did not have any problems in conducting the task (all matchers were successful in selecting the correct picture in all cases), did not produce any restarts or question-like realizations, and were not interrupted by the matcher. The entire task took the participants about 20 minutes. In two cases there was only one participant present, therefore in these cases the experimenter (who could speak NGT) fulfilled the role of the matcher (the experimenter obviously never acted as the director).

Figure 4.4. Picture card of one of the furniture grids as presented to the matcher.

Data analysis

As mentioned above, analysis has taken place for the first (initial) and third (repeated) references to the objects that had to be described three times, leading to a total of 112 object descriptions. The aim was to look at how long the descriptions took and how many signs were used in these descriptions, as well as at how the signs themselves were produced. We used the multimodal annotation programme ELAN (Wittenburg, et al., 2006) to annotate the descriptions. We annotated the duration of the target descriptions by selecting the beginning and end point of each trial in ELAN, using a beep sound that was present at the beginning of each trial as the starting and cut-off point. We looked at the number of lexical signs that were produced and their duration by selecting the beginning and end point of each sign in ELAN. A sign was considered as such on the basis of Kendon 's (1980) movement based convention of the gesture phrase, meaning that a sign can, but does not have to, contain up to five gesture phases (preparation, pre-stroke hold, stroke, post-stroke hold, and retraction) and always includes a stroke. We also analysed sign precision. In order to measure sign precision, a separate perception test was used, which will be discussed below under Experiment II.

All data annotated in ELAN was exported into SPSS. The statistical procedure consisted of two repeated measures ANOVAs, one by participants (F_1) and one by items (F_2). On the basis of these results, the *minF'* (Clark, 1973) was calculated, which indicates whether the results can be generalised over both participants and items. The experiment consisted of a 2 x 2 x 2 design, with factors domain (levels: people, furniture), repetition (levels: initial, repeated), and picture (levels: one, two). We only report where results are significant.

To check whether the fact that in two cases the matcher was the experimenter might have had an effect on the participants' behaviour, we performed an additional analysis with the type of interlocutor (participant or the experimenter) as a between subjects variable. There were no significant main effects of the type of interlocutor on any of our dependent variables and there were no interactions with type of interlocutor, which means that there were no noticeable differences in the data between participants who took part twice and those who took part only once (as director). Therefore, the type of interlocutor was excluded as a variable from our further analyses.

To check whether sign experience had an impact on our results, we divided the group of signers in two groups, one consisting of signers that had been signing NGT since before age 10, and one consisting of signers that only started to learn NGT after this age. We then performed another additional analysis with two between subjects variables added to the design, one indicating whether a participant had been born deaf or not, and one indicating whether the participants had been signing NGT since before age 10, or not. Also for these two variables there were no significant main effects on any of the dependent variables and there were also no significant interactions with either of these two variables. As with the type of interlocutor, we therefore also excluded these two variables from the final analyses.

Results

Firstly, table 4.1 shows the average duration of the target descriptions, in seconds, of initial and repeated references. Speakers took significantly less time in describing repeated references (M = 14.5, SD = 5.5, 95% CI = (11.3, 17.6)) than initial references (M = 24.2, SD = 8.4, 95% CI = (19.4, 29.1)), $F_1(1,13) = 35.14$, $p < .001$, $\eta_p^2 = .730$; $F_2(1,4) = 22.30$, $p < .01$, $\eta_p^2 = .848$; $minF'(1,10) = 13.64$, $p < .01$. Table 4.1 also shows the results for the average number of signs that were produced during initial and repeated references, and here the picture is similar. Speakers produced significantly fewer signs in repeated references (M = 5.6, SD = 1.2, 95% CI = (4.9, 6.3)) than in initial references (M = 8.2, SD = 2.1, 95% CI = (6.9, 9.4)), $F_1(1,13) = 42.51$, $p < .001$, $\eta_p^2 = .766$; $F_2(1,4) = 16.59$, $p < .05$, $\eta_p^2 = .806$; $minF'(1,7) = 11.93$, $p < .05$[10]. Finally, table 4.1 illustrates the average sign duration, in seconds, of signs in initial and repeated references, and again the same general pattern can be observed. The average duration of signs was shorter in repeated references (M = 1.2, SD = .20, 95% CI = (1.1, 1.3)) than in initial references (M = 1.5, SD = .28, 95% CI = (1.3, 1.6)), $F_1(1,13) = 15.1$, $p < .01$, $\eta_p^2 = .537$.; $F_2(1,4) = 20.17$, $p < .05$, $\eta_p^2 = .834$; $minF'(1,14) = 8.63$, $p < .05$.

Table 4.1. Overview of mean results for dependent variables (duration of description in seconds, number of signs, and sign duration in seconds) for initial and repeated references, in the production experiment.

	Initial (SD)	Repeated (SD)
Duration	24.2 (8.4)	14.4 (5.5)
Number of signs	8.2 (2.1)	5.6 (1.2)
Sign duration	1.5 (.28)	1.2 (.20)

In sum: we find systematic effects of repetition, in that repeated references are shorter, contain fewer signs, and shorter signs than initial references. These effects were the

[10] The assumption of sphericity is by definition always met in our design. However, the number of signs was not normally distributed, and for sake of conservativeness necessary corrections have been applied.

same for both domains (furniture and people) and for all pictures; in particular, we found no significant interaction between the factors repetition and domain or repetition and picture, although references to people contained more signs in general than references to furniture items, in line with speech results from previous work (Koolen, et al., 2011).

To illustrate, figures 4.5 and 4.6 show a case of reduction in the description of a target object from the furniture domain. In the initial description, the participant takes longer and uses more signs and seemingly more precise signs than in the repeated description.

BANK, DRIE ZITPLAATSEN, SCHUIN, GROOT, RECHTS, OPZIJ
SOFA, THREE SEATS, ASKEW, BIG, TO.THE.RIGHT, TO.THE.SIDE

Figure 4.5. Still and gloss (in Dutch, followed by English translation) of initial description of a sofa, lasting 48 seconds. Sign depicted in still is BANK ('sofa'), with a fairly large extension and well defined edges (as indicated by the arrows).

BANK, GROEN, OMGEDRAAID, DRIE ZITPLAATSEN
SOFA, GREEN, TURNED.AROUND, THREE SEATS

Figure 4.6. Still and gloss (in Dutch, followed by English translation) of repeated description of the same sofa, lasting 17 seconds. Sign depicted in still is BANK ('sofa'), with a smaller extension than in figure 4.5 and without well defined edges (as indicated by the arrows).

Conclusion experiment I: Production experiment

The results from experiment I show that several aspects of NGT were reduced in repeated references. Repeated references produced by signers of NGT were shorter than initial references, and repeated references in NGT contained fewer and shorter signs than initial references. These results suggest that, at least for the aspects taken into account here, repeated references in NGT behaved as previous studies found for repeated references in speech. Repeated references by signers of NGT, containing predictable information, were produced in a more efficient way than initial references. Experiment I has also shown that it is possible to adapt methods used to study reduction in speech (and gesture) in order to look at reduction in sign language.

Experiment II: perception test

As part of our analyses we wanted to also analyse sign precision. Since it is difficult to define objective measures with which to measure sign precision, a perception test was set up in which participants had to judge, in a forced choice task, which sign they considered to be the most precise, looking at pairs of video clips with signs produced in either initial or repeated references. This perception test was administered with two groups of participants: deaf NGT signing participants, and hearing Dutch participants with no knowledge of NGT. We will discuss each participant group separately.

NGT signing participants

Six NGT signing participants (1 male, 5 female, age range 18-56 years old, M = 33 years and 8 months) who were all deaf and had been signing NGT for over 10 years took part without receiving any form of compensation.

Stimuli

The participants were presented with a PowerPoint presentation in which they saw 40 trials, consisting of 40 pairs of video clips. Each pair of video clips was presented on one slide (as shown in figure 4.7).

Figure 4.7. Example of presentation manner of one pair of video clips to a participant in the perception test.

Both video clips showed the same sign produced by the same signer of NGT about the same object, as described in the director-matcher task, except in one video clip the sign was produced in an initial reference and in the other video clip the sign was produced in a repeated reference. The signs were selected on the basis of their availability in the data set from the director-matcher task, meaning that a sign had to be produced in both an initial and in a repeated description in order to be included in this data set. Signs that were only produced in an initial or only in a repeated description were left out, and signs that were produced in a different context in the initial description compared to the repeated reference were also left out (for example an initial sign which was produced at the beginning of the initial description where the repeated sign was produced at the end of the repeated description). Pairs of signs could be from either the people or the furniture domain and were clustered together in the perception test on the basis of their semantic meaning. This means that participants in the perception test were first presented with a number of video clip pairs showing tokens of one sign, e.g. DESK, followed by a cluster of video clip pairs showing tokens of another sign, e.g. WOMAN. The order in which the participants were presented with initial versus repeated signs in the video clip pairs was counterbalanced over pairs of video clips (so it was not the case that for each pair the first video clip they saw was always the sign produced in an initial reference).

Procedure

The participants were given written instructions (in Dutch) and had the opportunity to ask questions to the experimenter, after which the experiment started. The participants had to watch the pairs of video clips, one video clip at a time, and were allowed to watch a video clip more than once if they wanted to. The task was to choose for each pair of video clips which sign they considered to be the most precise (the sign in video clip A or B). The task was a self-paced forced choice task and even though the participants were allowed to watch the video clips more than once, they were encouraged to go with their first intuition. Participants were not allowed to go back to a previous stimulus item once they had made their decision. The only instruction they were given was to choose which sign they considered to be the "most precise". No details were given to suggest what the participants should base this judgment on.

Data analysis

For each trial, one point was administered when a participant considered the sign from the initial reference to be the most precise and zero points were administered when the participant considered the sign from the repeated reference to be the most precise. We conducted a binomial test to check whether the distribution between scores of 0 and 1 was equal, or not.

Results

A binomial test showed that initial gestures were chosen in 124, or 52%, of 240 cases. The distribution between scores was equal, p=.651. This shows that signs produced in initial references were not considered to be more precise than signs produced in repeated references.

Discussion

There was no effect of repetition on sign precision as judged by the NGT signing participants. However, the comments given by some of the participants during the experiment (such as "this one could be a bit bigger") suggest that maybe the NGT signers interpreted *precision* as *intelligibility,* and reached ceiling effect because both initial and repeated signs were intelligible. Alternatively, the NGT signers did not interpret precision as intelligibility, but were simply not very good at phonetic discrimination, possibly because the difference between the initial and the repeated signs was not communicatively relevant. To check whether any of these suggestions might be true, we conducted the same experiment with non-NGT signers. This strategy was chosen, since research has shown (e.g. Brentari, Gonzalez, Seidl, & Wilbur, 2011) that nonsigners can have a high degree of sensitivity to visual prosodic cues of a sign language. In Brentari et al.'s work, for example, nonsigners were highly accurate in identifying breaks and non-breaks between signs. The assumption here is that participants with no knowledge of NGT will not be influenced or distracted by the lexical meaning of the signs since the meaning of the signs is not communicatively relevant for them.

Nonsigning participants

Twenty-seven participants took part in the perception test. The participants were Dutch first year university students (9 male, 16 female, age range 18-30 years old, M = 21 years and 4 months), who had no knowledge of NGT and who took part as partial fulfilment of course credits. The stimuli, procedure, and data analysis were the same as for the NGT signing participants.

Results

A binomial test showed that initial gestures were chosen significantly more often (in 726, or 67%, of 1080 cases) to be the most precise than repeated gestures, $p < .001$. Contrary to the NGT signing participants, the participants with no knowledge of NGT judged signs taken from initial references to be more precise than signs taken from repeated references.

When we analyse the data of all participants together, by conducting an independent samples t-test, we find that there is a significant difference between the NGT signing and non-NGT signing participants, $t(31) = 4.068$, $p < .001$.

Conclusion experiment II: perception test

The results from experiment II show that, for the nonsigners, signs produced in repeated reference were considered to be less precise than signs produced in initial references. However, we also found that the NGT signing participants did not consider the signs from repeated references as less precise than the signs from initial references, perhaps because they were distracted or influenced by the lexical meanings of the signs.

In general, the results of experiment II lend further support for the findings of experiment I, namely that it is possible to use methods from speech and gesture research to study reduction in sign language, and that there indeed appears to be reduction in repeated references in NGT.

General discussion and conclusion

Summarising the results from the production and perception experiments, we found evidence suggesting that there is reduction in repeated references in sign language. We showed that repeated references were shorter, contained fewer and shorter signs, and that signs produced in repeated references were considered to be less precise by nonsigners than signs in initial references.

The present results on sign language can be tied in with previous findings, both for speech and for gesture, that language users tend to be efficient by reducing predictable information. Relating the results to previous work on speech, we showed that repeated references were shorter and contained fewer signs than initial references, in line with work by Clark and Wilkes-Gibbs (1986) and Galati and Brennan (2010). The result that signs in repeated references were shorter can also be related to previous work on speech by Aylett and Turk (2004), Lam and Watson (2010) and Bell et al. (2009) where it was found that predictable speech (through redundancy or repetition) had a shorter duration than unpredictable speech. Our finding that signs in repeated references were considered to be less precise can be viewed as an extension of the work by Bard et al. (2000), who found that repeated references had a less clear pronunciation than initial references.

When we compare the results from the present study with previous work on co-speech gestures, we can also see certain connections. It has been found that gestures with common ground are less precise than gestures without common ground (Gerwing & Bavelas, 2004). This can be related to our finding that signs in repeated references were considered to be less precise. Work presented in chapter 3 on the effect of repeated references on gestures (Hoetjes, et al., 2011, 2015) found that repetition may cause a reduction in the number of gestures, as was found in the present study for the number of signs. Moreover, their finding that gestures in repeated references were considered to be less precise than gestures in initial references, can be directly mapped onto the present results for signs. Importantly, the reduction found in the current study can be tied in with work on language efficiency and cannot be explained through a general reduction over time (as discussed in e.g. Singleton, Morford, & Goldin-Meadow, 1993; Supalla, 2008) with participants becoming more 'sloppy' in the course of the experiment. We only found a main effect of repetition, even though initial and repeated pictures occurred throughout the experiment, and we found no effect of picture, even though the different pictures occurred in different positions in the experiment. In short, the present study is the first study on sign language that shows that signers of NGT behave similarly when describing repeated references to what previous studies have found for speech and gesture in speakers of spoken languages.

Due to the fact that not a great amount of previous work has been done on reduction in sign language, the method used in the current study was inspired by relevant previous work on speech and gesture. We looked at fairly course-grained and modality-independent (i.e. applicable to speech, gesture and sign) measures such as duration of the description and number of signs and not at more sign-specific aspects such as exact sign location (as has been done by e.g. Tyrone & Mauk, 2010). We studied overall differences between initial and repeated descriptions and did not make pair-wise comparisons between signs from initial references and signs from repeated references.

Despite this course-grained approach and the fact that our measures were not based on sign characteristics per se, we were still able to see that reduction in sign language occurs. Our results show that it is possible to use such modality-independent methods to study reduction in repeated references. Studying reduction in sign language is an interesting addition to previous work on speech and gesture. Considering that signs can be argued to be both like speech (they are lexical) and like gesture (they are produced with the hands) it is interesting to see that in the present study, signs also behave like what previous studies found for speech as well as what previous studies found for gesture.

Naturally, the current study has certain limitations and leaves room for further research. Presently, we only looked at fairly coarse-grained measures, namely the duration of the target descriptions, the number of signs, the duration of these signs and the perceived precision of the signs. Although these course-grained measures were useful for a first study on repeated references in sign language, it can be argued that many more aspects of the signed references need to be studied. The present analyses have shown that signers of NGT reduce their descriptions in repeated references, but the finer details of the reduction that may be present in the description have not yet been taken into account. With 'finer details' one could think of characteristics of the signs themselves such as the exact handshape of the sign, or the sign location, as was studied previously by Tyrone and Mauk (2010), but one could also think of details about the entire reference. Aspects of the entire reference that could be studied are, for example, which lexical signs were used for which reference, showing how much vocabulary overlap there was between the initial and the repeated references. One could also look at whether sign duration does not only depend on whether a sign is produced in an initial or in a repeated reference but also at whether the exact position of a particular sign within the referring expression has an effect. The current analyses were based on the overall references, and, at least for the production experiment, we did not presently compare lexically identical sets of initial and repeated signs, something which could be done in future research. Considering that the present study is a first look at reduction in repeated references in sign language, we also do not know to what extent the results from the current study can be generalised to other repeated references and other forms of predictable information in general, or whether the results might be specific to this type of task.

When we focus on the perception experiment (Experiment II), we can discuss several things. Firstly, the results showed an effect of repetition on sign precision, but only for the nonsigners. The NGT signing participants did not consider signs from initial references to be more precise than signs from repeated references. It might be the case that the reductions in sign precision were either not picked up, or were mentally compensated by the NGT signing participants, because these reductions might not be linguistically or communicatively relevant for the NGT signers. Because of this, we decided to also run the experiment with nonsigners, whose precision judgments could not be affected by lexical interpretations. In addition, there are reasons to assume that the use of nonsigners is indeed a reasonable approach. As mentioned above, previous research has shown (e.g. Brentari, et al., 2011) that nonsigners are very capable in recognizing visual prosodic cues of a sign language.

A second point of discussion is that the participants of the perception test were not told what to base their precision judgment on (and in chapter 3 we showed that giving an explicit definition of "precision" did not influence participants' responses). Not giving participants explicit instructions was done on purpose since the term "precision" can mean different things to different people, which is exactly why it is difficult to measure objectively. However, to avoid the possibility that different participants might have interpreted the task differently, it would be possible in future work, especially when using NGT signing participants, to set up the task slightly differently, for example by asking participants to recognise a sign's meaning, as in Bard et al.'s (2000) work on speech, instead of judging a sign's precision. Another option would be to explain the entire setup of the experiment to the participants and ask them to judge which sign was produced in an initial reference and which one in a repeated reference.

In sum, the analyses done presently showed us not only that we can use analyses from related work on speech and gesture and adapt them to analyse signs in repeated references, but also that signers of NGT do reduce their repeated references. In fact, the ways in which these repeated references are reduced in NGT are quite similar to what has been found previously for speech and gesture. It is well known that speakers of non-signed languages are communicatively efficient by reducing predictable information, both in speech and in co-speech gestures. This study has shown, for the first time, that signers design their utterances to be efficient in similar ways.

Acknowledgments

We would like to thank Axelle Schmit, Kristel Bartels, Elsa Jonkers and Manon Yassa for help in collecting and annotating the data, Martijn Goudbeek and Carel van Wijk for help with the statistics, and Annette Hohenberger and Claude Mauk for their constructive comments on a previous version of this manuscript. An earlier version of this work was presented at the 2011 Gesture Workshop in Athens, Greece, and at the 34th Annual Conference of the Cognitive Science Society in Sapporo, Japan (2012). We would like to thank the audience for useful comments and suggestions.

5

On what happens in gesture when communication is unsuccessful

Abstract

Previous studies found that repeated references in successful communication are often reduced, not only at the acoustic level, but also in terms of words and manual co-speech gestures. In the present study, we investigated whether repeated references are still reduced in a situation when reduction would not be beneficial for the communicative situation, namely after the speaker receives negative feedback from the addressee. In a director-matcher task (experiment I), we studied gesture rate, as well as the general form of the gestures produced in initial and repeated references. In a separate experiment (experiment II) we studied whether there might (also) be more gradual differences in gesture form between gestures in initial and repeated references, by asking human judges which of two gestures (one from an initial and one from a repeated reference following negative feedback) they considered more precise. In both experiments, mutual visibility was added as a between subjects factor. Results showed that after negative feedback, gesture rate increased in a marginally significant way. With regard to gesture form, we found little evidence for changes in gesture form after negative feedback, except for a marginally significant increase of the number of repeated strokes within a gesture. Lack of mutual visibility only had a significant reducing effect on gesture size, and did not interact with repetition in any way. However, we did find gradual differences in gesture form, with gestures produced after negative feedback being judged as marginally more precise than initial gestures. The results from the present study suggest that in the production of unsuccessful repeated references, a process different from the reduction process as found in previous studies in repeated references takes place, with speakers appearing to put more effort into their gestures after negative feedback, as suggested by the data trending towards an increased gesture rate and towards gestures being judged as more precise after feedback.

This chapter is based on:

Hoetjes, M., Krahmer, E. & Swerts, M. (2015). On what happens in gesture when communication is unsuccessful. *Speech Communication, 72*, 160-175.

Introduction

People often refer to objects and persons during a communicative exchange. In many cases, the same target is referred to repeatedly in the discourse, and these references may be multimodal, using both speech and manual co-speech gesture. It is well established that repeated references in successful communication tend to be reduced variants of initial references, consisting of fewer words and fewer gestures. For example, a speaker who wants to point out a particular person for an addressee might produce an initial description such as "that tall girl with the long blond hair", accompanied by two gestures, first one indicating the height of the girl, followed by another one indicating the length of the girl's hair. Later on in the conversation, the speaker might refer back to the same girl by saying "the tall girl from before", accompanied by only one gesture, say, indicating the girl's height.

These reduction effects have been explained in terms of increased common ground (e.g., Clark & Wilkes-Gibbs, 1986; Galati & Brennan, 2014; Gerwing & Bavelas, 2004; Holler & Stevens, 2007; Jacobs & Garnham, 2007). The initial description introduces an entity in common ground, after which a reduced reference can be sufficient. The emergence of common ground is the result of a process often referred to as information grounding (Clark & Schaeffer, 1989; Traum, 1994), and generally understood as involving two phases: a presentation phase, in which a speaker sends a message to the addressee, and an acceptance phase, in which the addressee signals whether the message came across in good order or not. If our addressee knows which tall, long-haired girl the speaker is referring to, he can signal this using a positive "go on" signal (using the terminology of Krahmer, Swerts, Theune, & Weegels, 2002). This can, for example, be an explicit backchannel cue such as "OK", but it may also be a more implicit signal, because the addressee correctly identifies the target girl, e.g., by looking at her.

Now, consider what would happen if the initial reference is somehow not successful, which our addressee would indicate during the acceptance phase using a negative, "go back" signal (e.g., "Sorry, which girl?"). Then, how would our speaker realise her second, repeated reference to said girl? We know from other studies that speakers tend not to reduce their utterances (in terms of number of words or articulatory effort) in response to negative feedback, but we know remarkably little about whether, and if so, how, speakers' gestures would change. To the best of our knowledge only a handful of earlier studies asked this question, of which Holler and

Wilkin (2011) is arguably the most detailed. However, these authors present their work as “a first glimpse of speakers’ gestural behaviour in response to addressee feedback” (Holler & Wilkin, 2011, p. 3534), and point out that more work is “urgently needed” (ibid.).

In the present study we address the above questions by comparing gestures produced in initial references with those in repeated references following negative feedback. The experiments that were conducted for this purpose are based on the experimental paradigm of our previous work on successful repeated references, presented in chapter 3 of this thesis (Hoetjes, et al., 2011, 2015). As in this previous work (as well as in various other studies, including the aforementioned Holler & Wilkin, 2011), we concentrate on two aspects: the gesture rate and the qualitative form of the gestures. Before describing our current study in detail, we provide an overview of relevant background literature.

Background

Reduction in successful repeated references

Repeated references occur in discourse whenever a particular person or object is mentioned or described more than once. These references are never exactly the same. The differences in the ways in which references are realised are not only due to naturally occurring variability in speech, but are also influenced by the mere fact that the information status of the referent changes when it gets repeated. For instance, when an object is mentioned a second time, it already belongs to the discourse model of speaker and addressee, and can be assumed to be common ground (that is, when communication was successful). Research has found that when information is given or predictable, such as is the case in repeated references and increased common ground, speech is often reduced.

For example, Lieberman (1963) found that words produced in contexts in which they were predictable had a shorter duration and a lower pitch peak (F0). In addition, they were less intelligible when they were taken out of context. In a similar vein, references to given information have been found to be less intelligible when taken out of context and presented in isolation (e.g., Bard, et al., 2000; Fowler & Housum, 1987), and to have a shorter duration and a lower pitch peak (e.g., Aylett & Turk, 2004; Brown, 1983; Fowler & Housum, 1987; Lam & Watson, 2010), than references to information that is new in the discourse.

Reduction in repeated references at the lexical level has also been well established. For example, Clark and Wilkes-Gibbs (1986) showed that when speakers repeatedly (and successfully) refer to the same object, they lexically reduce their references (e.g. from an initial description such as "a person who's ice skating, except they're sticking two arms out in front", to a sixth description of the same figure as "the ice skater", Clark & Wilkes-Gibbs, 1986, p. 12). This robust finding has often been explained in terms of the creation of a conceptual pact (Brennan & Clark, 1996), which occurs as more common ground emerges between speakers.

These findings relate to spoken language, but human speakers are known to produce speech in tandem with a variety of visual cues, of which manual gestures are our main focus of attention in this study. Such manual speech-accompanying or co-speech gestures (which we will call gestures for short) can generally be defined as symbolic movements of the arms and hands that people produce when they speak (Kendon, 1980, 2004; McNeill, 1992). Most researchers agree that there is a close, co-expressive relationship between speech and gesture (Kendon, 1972, 1980, 2000, 2004; McNeill, 1985, 1992; McNeill & Duncan, 2000), with speech and gesture arguably going "hand-in-hand" (e.g., Kita & Özyürek, 2003; So, et al., 2009). To take one, more or less arbitrary, example, consider the study reported by So et al. (2009), who asked English speakers to retell stories to an experimenter. So and colleagues found that speakers often used gestures to identify a referent in the story, by producing it in the same location used for the previous gesture for this referent. However, importantly, they did this most often when the referent was also uniquely specified in the accompanying speech. This led these authors to conclude that for referential identification, speech and gesture indeed appear to go hand-in-hand.

Based on this, one could hypothesize that reduction in speech during successful communication is accompanied by reduction in gesture. This is indeed what a number of studies have investigated, and to some degree the results are consistent with this hypothesis. For instance, it is generally found that repeated multimodal references contain fewer gestures than initial ones (e.g., de Ruiter, et al., 2012; Holler, et al., 2011; Levy & McNeill, 1992; Masson-Carro, Goudbeek, & Krahmer, 2014), just as they contain fewer words. However, when looking at the ratio of gestures to words a more complex picture emerges. Gesture rate (often computed as the ratio of gestures per 100 words, although various alternatives have been proposed, see chapter 3 for discussion) has a long tradition in gesture research, going back to, at least, Cohen and Harrison (1973). It has frequently been used as a dependent variable in gesture studies, because it allows us to gain more insight into the relative contribution of gesture to speech. Some studies found evidence for a decrease in gesture rate when information is shared or repeated (Galati & Brennan, 2014; Jacobs & Garnham, 2007), suggesting that gestures become gradually less important, but others found that it increases (Holler, et al., 2011) or that it stays the same (de Ruiter, et al., 2012). A smaller number of studies have also considered the form of gestures, and generally these studies found evidence for gestures being smaller and less precise when relating to information in common ground (Galati & Brennan, 2014; Gerwing & Bavelas, 2004; Holler & Stevens, 2007; Vajrabhaya & Pederson, 2013). Gerwing and Bavelas (2004), for example, argue that gestures relating to given information are "sloppier" and more "elliptical", much like words expressing given information are articulated less clearly.

In chapter 3 of this thesis we presented a large-scale study with the aim to gain more insight in gesture behaviour during the production of repeated references, also in view of the mixed results of earlier studies. This was done using a variant of the director-matcher referential communication task (e.g., Clark & Wilkes-Gibbs, 1986; de Ruiter, et al., 2012; Holler & Stevens, 2007; Krauss & Weinheimer, 1966), in which speakers were asked to refer to Greebles (Gauthier & Tarr, 1997), which are hard to describe figures with different shapes and protrusions. During the experiment, the director (speaker) described various Greebles to the matcher (addressee), some of which were described multiple times, allowing the authors to compare initial, second and third references. We found, among other things, that the gesture rate (per 100 words) did not differ significantly between the three descriptions. In addition, no reliable qualitative differences in form were found (looking at gesture duration, gesture size, whether the gesture was produced with one hand or with two hands and at the number of repeated strokes). However, in an additional judgment study, we found that gestures produced during initial descriptions were judged to be more "precise" (as defined by Gerwing & Bavelas, 2004) than those produced during repeated descriptions.

The impact of (negative) feedback

The studies on reduction in referential communication in speech and gesture discussed above all involve situations in which the communication was successful. This was generally the case because the speaker received positive, "go on" feedback, that was either explicit (e.g. via backchannel cues from the addressee) or implicit (e.g. because the addressee selected the right "target"). However, referential communication is not always successful, which an addressee may indicate by responding to an initial description with negative, "go back" feedback. Various studies have revealed that negative feedback signals are marked, in that they are associated with more prosodic effort, for instance because they are realised with a higher pitch, longer duration and more pauses than comparable positive feedback signals (Krahmer, et al., 2002; Shimojima, Katagiri, Koiso, & Swerts, 2002). This makes intuitive sense, since it is more important for the speaker to pick up negative than positive feedback from the addressee.

Speakers can respond to negative feedback in various ways, also co-depending on the nature of the feedback. For example, the speaker might repeat the words, but rather than reduce these, she is likely to articulate them with more prosodic effort (louder, higher, etc), resulting, potentially, in hyper-articulated speech (Lieberman, 1963; Lombard, 1911; Oviatt, MacEachern, & Levow, 1998). In addition, she may reformulate the original utterance and/or add further information to it (Litman, Swerts, & Hirschberg, 2006). In this study, we investigate whether, and if so, to what extent, a speaker's gestural behaviour changes as well in response to negative feedback. Given the aforementioned close relationship between speech and gesture, it can be hypothesized that gestures produced during a repeated description following negative feedback are not reduced, but what the precise effect will be on the gesture rate and gesture form is difficult to predict. The outcome does have important implications for theories about speech-gesture production, as it will inform us about the relative importance of the gesture modality during communicative problems.

So far, only a handful of studies have looked at gesture production in response to feedback. Jacobs and Garnham (2007, experiment 2), for example, found an effect of the level of attentiveness of the listener on gesture production. They had participants narrate a comic strip to either an attentive or inattentive confederate listener. The attentive listener was instructed to behave in an attentive manner while each strip was explained, using appropriate verbal and non-verbal (positive) feedback, while the other was instructed to display "inattentive behaviour". Jacobs and Garnham found that speakers produced more gestures when the listener seemed attentive rather than inattentive. In a somewhat similar vein, Galati and Brennan (2014) point out that speakers take into account verbal and non-verbal addressee feedback, which in turn may shape the speaker's gestures (see also Kuhlen & Brennan, 2010). However, in their study, Galati and Brennan conclude that feedback could not solely account for the way speakers changed their gestures when talking to different addressees (p. 447). While studies such as these indicate that speakers' overall gestural behaviour may be influenced by (lack of) feedback from an addressee, they do not provide insights into the question of how speakers adapt their gestures, both in terms of frequency and form, in response to specific instances of (negative) feedback.

As far as we know, the only study that addresses this question in any detail is Holler and Wilkin (2011). These authors first point to a small number of descriptive studies, describing examples from earlier work which indeed suggest that individual gestures can be adjusted due to feedback from the addressee (Kendon, 2004; Streeck, 1993, 1994). This serves as a starting point for Holler and Wilkin's experimental study, in which they asked participants to retell a fragment from a German television series for children to a confederate addressee who provided scripted feedback at four predetermined points in the narrative. Feedback always took the form of a question, which could either be a request for clarification or confirmation of a detail, or an expression of global non-understanding, asking the speaker to repeat or clarify what was said. Notice that all of these could be classified as "go back" feedback signals, in that they indicate that the addressee requires more information about what the speaker said before. Holler and Wilkin compared utterances before and after feedback, focusing on the gesture rate and the form of gestures. They found that speakers gestured at a numerically slightly higher rate before than after feedback, although this difference was not statistically significant. They then zoomed in on the effects of the four feedback signals separately, and found, again, that for three out of four types of feedback, gesture rate before and after feedback did not differ significantly. The fourth one (seeking confirmation) did lead to a significantly lower gesture rate. Concerning the analysis of gesture form, Holler and Wilkin compared 100 pairs of gestures produced before and after feedback, and found that in the majority (60%) of the cases gestures were likely to be "more communicative" after feedback, which means that they were either larger, more precise (in the sense of Gerwing & Bavelas, 2004), produced in a visually more prominent place or more likely to be displayed from a character perspective (see Holler & Wilkin, 2011, p. 3531, for details).

Holler and Wilkin (2011) point out that their study offers the first insights into how addressee feedback influences gesture production, but they also highlight a number of issues that should be taken up in future research. One concerns the nature of the feedback that was provided; even though feedback was scripted, there was some variation in the behaviour of the confederate, for instance "in terms of whether she used a gesture or not" (Holler & Wilkin, 2011, p. 3534). Given earlier studies on mimicry in gesture production (see e.g., Mol, et al., 2012, for an overview and discussion), this could have influenced the gestures produced after feedback. In addition, they point out that it is unclear to what extent their findings can be generalised to different languages (the language they studied was English), other kinds of feedback, and other variables capturing the form of the speaker's gestural behaviour.

On the role of visibility

Gesture researchers have often used visibility in their experimental designs to get a better understanding of the extent to which gestures are produced for an addressees or whether they are (also) produced for the speaker, i.e., may serve more cognitive needs (see Bavelas & Healing, 2013, for discussion). The general reasoning is that if speakers would produce gestures to further their addressees' understanding, one would expect speakers to produce fewer gestures when addressees cannot see them (see e.g., Alibali, et al., 2001, for this argumentation). Indeed, various studies have found that gesture rates decrease when participants cannot see each other (e.g., Alibali, et al., 2001; Bavelas, et al., 2008). In addition, visibility may also influence the form of the gesture (Bavelas, et al., 2008; Gullberg, 2006). For example, Bavelas et al. (2008) found that speakers, describing an elaborate dress on a picture in a mutual visibility condition, used larger gestures, as if they were positioning the dress around themselves, while speakers describing the dress over the telephone tended to produce gestures on the same scale as on the picture.

In line with our previous study on repeated references (see chapter 3), and following many other studies (e.g., Alibali, et al., 2001; Bard, et al., 2000; Bavelas, et al., 2008; de Ruiter, et al., 2012; Hoetjes, Krahmer, & Swerts, 2014b; Holler, et al., 2011; Mol, et al., 2009), we include visibility as an additional variable in the design of our production experiment (experiment I). We do this in such a way that one group of participants will be able to see each other (mutual visibility), while the other group are prevented to do so using a screen (no visibility). We include visibility in our design for two reasons: first, because it enables comparison with our previous study on repeated references in successful communication, and, second, to study whether the impact of negative feedback on gesture production, both in terms of gesture rate and in terms of gesture form, is more speaker- or more addressee-oriented[11].

The present study

In this paper, we study the influence of negative feedback on the production of repeated multimodal Dutch referring expressions. For this, we use the same general set-up as employed in chapter 3, in which speakers, in a director-matcher task, had to refer to hard-to-describe objects with different shapes and protrusions (the aforementioned Greebles). Using the same set-up has two main advantages. Firstly, we know from the aforementioned study that referring to Greebles elicits a substantial number of spontaneous (mostly representational) gestures, both in initial and repeated descriptions. Secondly, and arguably more important, it serves as a kind of baseline, in that it allows us to compare speech-gesture production in successful repeated descriptions with unsuccessful ones, after negative feedback from the addressee.

Feedback (both positive and negative) can come in many variants. Here we opt for a simple variant: after a speaker (the director) has described a target object, the addressee (the matcher, who is a confederate of the experimenter) either selects the correct referent (‘go on’, which is signalled using a pleasant high ping sound) or (in a limited number of critical, repeated trials) a wrong one (‘go back’, signalled using a low buzzing sound). The current set-up enabled us to have a large level of control over the negative feedback, which was identical for all participants. In this way we could collect

[11] Note, however, that manipulating visibility does not necessarily distinguish between speaker and addressee functions, see Bavelas, et al., 2008; Holler, et al. 2011.

initial (before feedback) and repeated descriptions (after negative feedback) for all speakers for the same targets. This allowed us to study how speakers (which are the unit of analysis in our study, cf. Bavelas & Healing, 2013) adjust their gesture behaviour on the basis of negative feedback.

As mentioned above, following the work presented in chapter 3 of this thesis and many other related studies (e.g., Alibali, et al., 2001; Bavelas, et al., 2008; de Ruiter, et al., 2012; Hoetjes, et al., 2014b; Holler, et al., 2011; Mol, et al., 2009), we added visibility as an additional variable to the design, in such a way that one group of participants could see each other during the experiment, while the other group was prevented from doing so by an opaque screen which was placed in between them.

For the critical trials, the initial (pre-feedback) as well as the second and third (post-negative-feedback) descriptions were manually transcribed and the accompanying gestures coded. As motivated above, this allowed us to compare the gesture rate before and after negative feedback across multiple descriptions. In addition, we studied whether the form of the gestures changed as a function of feedback, using the coding scheme employed in chapter 3, looking at duration and size of the gestures, number of hands involved (one or two) and number of stroke repetitions. Additionally, precision of gestures was assessed using a separate judgment study with naïve participants.

By looking at both gesture rate and gesture form before and after negative feedback, we can further our understanding of the role that co-speech gestures play during communication. Gesture rates have often been used in gesture studies, because they inform us about the relative importance of speech and gesture in a multimodal utterance. For example, if gesture rate per word would increase after negative feedback, this would imply that speakers rely more on the gestural modality than on speech in the case of communication problems. In a similar vein, by comparing gesture form before and after negative feedback, we may learn how important gestures are for speakers and how much effort they put into them, and compare this to speech processes after negative feedback. For example, if speakers would produce more precise gestures after negative feedback, this would suggest they put more effort in the gestural part of their utterances. Earlier research on successful communication has often suggested that speech and gesture go “hand-in-hand”. In this paper, we ask whether the same pattern can be observed in the case of communication problems, or whether negative feedback has a different impact on gesture and speech production. This offers potentially important information for speech-gesture production models, which aim to explain how speakers produce speech and gesture in tandem (see e.g., Chu & Hagoort, 2014; Hoetjes, et al., 2015; Hostetter & Alibali, 2008; Wagner, et al., 2014, for recent discussion).

Experiment I: Production of gestures before and after negative feedback

Participants

Participants were 38 undergraduate students (9 male, 29 female, age range 18-30 years old, M = 21 years and 7 months), who took part as partial fulfilment of course credits. The participants took part in the experiment in the role of director, and a confederate took part in the role of matcher. This confederate was the same person (female, 20 years old) for all 38 director participants. The participants had no knowledge of, and had not taken part in our previous study on repeated references, which is presented in chapter 3 of this thesis.

Stimuli

The stimulus materials consisted of picture grids of Greebles[12] (see figure 5.1 for an example Greeble and see Gauthier & Tarr, 1997, for a more detailed description of the Greebles and their properties), which are abstract, small, yellow objects that are hard to describe. The Greebles, which were initially designed to study human face recognition, vary in terms of their gender (“Glip”, “Plok”), their main body shape (“Samar”, “Galli”, “Radok”, “Tasio”), their different types of protrusions (“Boges”, “Quiff”, “Dunth”), and the different shapes and sizes of these protrusions.

[12] Images courtesy of Michael J. Tarr, Center for the Neural Basis of Cognition and Department of Psychology Carnegie Mellon University, http://www.tarrlab.org/

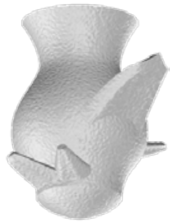

Figure 5.1. Example of a Greeble, turned upside down as compared to their presentation in Gauthier and Tarr (1997).

We successfully used the same Greeble objects in our previous study on reduction in repeated references (see chapter 3), and this is the main reason for reusing them in the current study. The Greebles were originally selected because they are quite abstract, and because they only differ from each other with regard to their shape and protrusions. The assumption was that, since speakers would naturally be unfamiliar with the specialised Greeble vocabulary mentioned above (e.g. “Glip”), these shapes and protrusions would have to be described in detail, using both speech and gesture. This way, we could collect repeated object shape descriptions, which were likely to contain repeated gestures illustrating the same Greeble-parts. As in the previous study, the Greebles were turned upside down as compared to the way in which they were presented in Gauthier and Tarr (1997), to make them look less like animate objects (which might cause participants to produce fewer shape descriptions because it would facilitate lexical descriptions such as “angry-looking” or “with the long nose”). We created two picture grids, each containing 16 Greebles. There were 10 trials per picture grid, thus 20 trials in total. In each trial, there was one target object, marked by a red square surrounding it, and 15 distractor objects surrounding the target object (see figure 5.2 for an example of a picture grid). The order in which the directors were presented with the two picture grids was counterbalanced across participants.

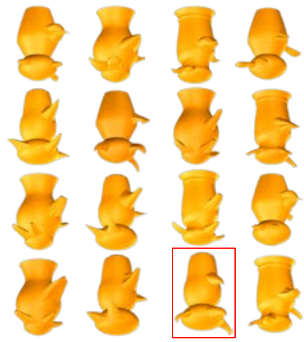

Figure 5.2. Example of one of the picture grids (picture grid 2). The target picture of this particular trial is the one in the bottom row, third from left (surrounded by a square).

The experimental manipulation (and the crucial difference with our previous study, in which we used these same stimuli) was that several Greebles had to be described repeatedly due to apparent communication problems. In each of the picture grids, two Greebles had to be described three times, of which the second and the third description were produced following negative feedback. To make sure that these critical trials did not stand out, an additional seven Greebles per grid had to be described once, and one Greeble had to be described twice (once after negative feedback). These were the filler items. The repeated references to the same object had to be given one straight after the other, when negative feedback provided by the matcher made it clear to the participant that an incorrect object had been chosen (see procedure below). The participants did not know in advance that in some of the trials they would have to take several attempts at describing a picture. This means that the participants thought they had to produce 10 descriptions for each picture grid (one per trial), when in reality they had to produce 15 descriptions for each picture grid. The Greebles that had to be described repeatedly were always preceded and followed by a filler item. To avoid order effects we made sure that the objects that had to be described repeatedly were never in the first or the last trial of the picture grid. We analysed all three descriptions of the objects that had to be described three times (i.e. a total of twelve descriptions for each participant, since four objects had to be described three times).

Procedure

The experiment consisted of a director-matcher task that was performed in a lab, where the director and the matcher were seated at a table opposite each other (see figure 5.3 for an example of the setup). After entering the lab, the participants (both the director and the confederate matcher) were given written instructions and had the opportunity to ask questions, after which the experiment started. The fact that the matcher was a confederate was to some extent communicated to the director: the director was told that the matcher was someone who had done the experiment before and was helping out because another participant had not shown up. In order to make sure that the director would do her best in providing good descriptions of each target and could not rely on previous experience of the matcher, she was told that the order in which objects were discussed was different for each participant pair (which was not actually the case). The instructions did not mention the use of gesture, so all gesture production was spontaneous.

The director was presented with the trials on a computer screen (which was positioned to her side), and the task for the director was to provide a description of the target object in such a way that it could be distinguished by the matcher from the 15 distractor objects. The director was told that, on the basis of her target description, the matcher picked the object that she thought was being described. After the matcher had picked one of the objects, a sound would tell the director whether the matcher had chosen the correct object or not (a low buzzing sound was played for incorrect object identification and a high ping sound was played for correct object identification). In terms of the coding scheme of Stivers and Enfield (2010), our negative feedback can be seen as an "other-initiation of repair", comparable to the feedback for scene 3 in Holler and Wilkin (2011) and the "What?"/"Sorry"/"huh?" negative feedback used in Healey, Mills and Eshgi (2013). When the sound indicating incorrect object identification was played, the director would describe the same target object again, until the matcher had identified the correct object. After this, the director could move on to the next trial. After 10 trials (and a total of 15 descriptions), the director was shown a second picture grid containing 16 new objects, and continued for another 10 trials (i.e. 15 descriptions).

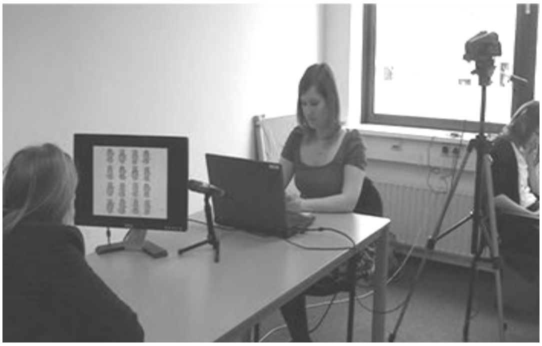

Figure 5.3. Example of experimental setup. The director is seen from the back, viewing one of the picture grids. The confederate matcher is seated across from the director, and the experimenter is seated to the side (just visible on the right next to the camera). The director and the matcher can see each other in this example, but for half of the participants a large (around A0 size) opaque screen was placed between the director and the matcher.

The director was told that the matcher was shown the same objects on her screen (which was positioned in front of her) as on the director's screen, but that these objects were ordered differently for the director and the matcher. It was explained that this meant that the director could not use the object's location in the grid for her target descriptions. In reality however, and unknown to the director, the director and the matcher both viewed the same picture grid and all the matcher had to do was play one of the sounds after the director had given a description of the target object of that particular trial. The participants were debriefed at the end of the experiment, and none of the participants expressed any suspicions concerning the experimental set-up.

The feedback given by the matcher only consisted of the sounds that were played after each trial, although she occasionally added appropriate post-feedback comments such as "hmm, that was the wrong one." The matcher offered no other verbal or non-verbal feedback, and displayed a neutral facial expression throughout the experiment. In addition, the matcher did not interrupt the director, gesture, or ask any questions. This allowed us to collect descriptions before and after negative feedback that were as comparable as possible, to ensure that any effects could be attributed to our manipulation, and not to possible differences in verbal interaction (see Holler & Wilkin, 2009, p. 273 for a similar argument, and Alibali et al., 2001, and Mol et al., 2009, for comparable instructions).

The entire experiment was filmed, with one camera positioned behind the matcher (filming the director) and another camera positioned to the side of the director (filming the entire setup, as in figure 5.3). For half of the participants, a large opaque screen was placed between the director and the matcher, meaning that, in these cases, the director and the matcher could not see each other throughout the entire experiment. Other than that the mutual visibility and no visibility conditions were identical.

Data analysis

The video recordings were digitised and the recordings showing the director were annotated using the multimodal annotation programme ELAN (Wittenburg, et al., 2006). The subsequent (speech and) gesture annotation and data analysis were based on previous research on (reduction in) repeated references, especially the research reported in chapter 3.

As a manipulation check, and to enable computation of gesture rate, we first conducted an analysis of the speech. All speech produced within one of the critical references (using the moment when the matcher played one of the sounds as the cutoff point) was transcribed orthographically. Hesitations, false starts, repetitions and corrections were all transcribed and included in the word count. Importantly, the distribution of disfluent elements was equal over the various conditions, so that these did not bias the gesture rates reported below. Contractions were counted as single words, but we encountered only one of these in our data ("zo'n" – such a). We analysed the number of words per trial, the duration (in seconds) per trial, and, based on these, we computed the speech rate (in number of words per second) per trial. Based on earlier research we expected the speech rate to go down after negative feedback (Krahmer, et al., 2002; Shimojima, et al., 2002), and this thus offers a manipulation check.

The gesture annotation was identical to the one employed in chapter 3 of this thesis, and relied on the gesture phases distinguished by Kendon (1980, 2004), see e.g., also McNeill (1992), Bressem and Ladewig (2011) and Wagner, Malisz and Kopp (2014). According to this view, gesture production consists of a number of phases. Starting from a stable, rest position, gesture production begins with a preparation phase, in which the hand moves away from the rest position, after which the stroke occurs, which is usually regarded as the obligatory, main part of the gesture, containing most effort as well as most semantic information. Before or after the stroke, a motionless phase may occur, which is usually referred to as the hold phase. Finally, the hands may return to a rest position during the retraction phase. For the gesture analyses, all stroke phases of all gestures produced in the descriptions of the objects that had to be described three times were selected.[13] The first video frame in which the most effortful movement started was taken as the onset of the stroke, while the offset of the stroke was taken to be the first video frame in which the stroke phase turned into a post-stroke hold, or retraction, phase.

Various authors have emphasised the importance of distinguishing different kinds of gestures during analyses (e.g., Alibali, et al., 2001; Bavelas, et al., 2008; de Ruiter, et al., 2012). Based on McNeill (1992), a distinction can be made between iconic, deictic and beat gestures. Iconic gestures, in our data, are gestures that depict a particular feature of the target object, such as its main shape or the shape of one of the protrusions ("shaped like [this]", where the word 'this' is accompanied by an iconic shape gesture). Deictic gestures are pointing gestures, generally used to indicate a specific location of one of the object's protrusions ("and [here] there is a pointy bit"). Beat gestures consist of simple rhythmic movements without any semantic relation to the speech they accompany. In our previous study, also using Greeble stimuli (see chapter 3), we found that over 95% of the gestures produced by directors were iconics (and, importantly, that figure did not change depending on whether it was an initial or repeated description), making separate analyses for different kinds of gestures impossible. The same applies to the current dataset, in which the affordances of the Greeble stimuli (consisting of distinct shapes and protrusions), resulted in our speakers producing iconic gestures almost exclusively. Therefore we decided, as in chapter 3, to not distinguish between the different types of gestures in our gesture analyses.

We computed gesture rate per description by dividing the number of gestures by the number of words. For the sake of readability, rates were multiplied with 100, so that

[13] Given the smaller size of the dataset in this study as compared to the one in chapter 3 (Hoetjes, et al., 2015), we decided to include all gestures in the detailed analysis, whereas in chapter 3 only one gesture per description was annotated in detail (even though all were counted and taken into account for analyses of gesture rate).

the gesture rate can be interpreted as the number of gestures per 100 words. In addition, we analysed several aspects of the form of the gestures. When a director did not produce a gesture in a description, this was treated as a missing value in our analyses on gesture form. The following four aspects of gesture form were taken into account. We measured the duration of the stroke, in seconds. We measured the size of each gesture by coding whether the stroke was produced with a finger (code 1), the hand (code 2), the forearm (code 3) or the entire arm (code 4), with a higher code assuming that the smaller articulators could also be used (e.g. code 3 includes 1 and 2). We coded whether the gesture was produced with one hand or with two hands (resulting in a range from 1-2, with e.g. 1.3 indicating that 30% of gestures were two-handed). Finally, we annotated the level of repetition within each gesture by counting the number of repeated strokes. A stroke was considered to be repeated when (nearly) identical strokes followed each other without a retraction phase in between.

To assess annotation reliability, a second annotator, who was not aware of the experimental conditions, coded gesture duration, gesture size, number of hands and number of repeated strokes for a subset of the data, consisting of the first gesture of all participants who produced at least one gesture (N=34 gestures, 2.5% of the data). The annotators agreed on only 44% of cases on gesture duration[14] (Kappa= .042), but on 88% of cases on the size of the gesture (Kappa= .821), 97% of cases on the number of hands that were used (Kappa= .941), and on 73% of cases on the number of repeated strokes (Kappa= .277). The low level of agreement on gesture duration meant that we decided to disregard gesture duration from our further analyses[15]. The other levels of agreement indicate that these annotations were reliable, and range from 'fair', for repeated strokes, to 'almost perfect agreement', according to Landis and Koch's (1977) characterization. Therefore, we used the first author's annotations for the statistical analysis.

Speech and gesture analyses were conducted for all three reference descriptions of the objects that had to be described three times. The statistical procedure consisted of

[14] There was agreement on gesture duration when there was a maximal difference of 5 frames, or 200ms, between annotators.

[15] Leaving out the analyses for gesture duration did not change the overall picture as presented in the results section since there were no significant effects of repetition or visibility on gesture duration.

two repeated measures ANOVAs, one by participants (F_1) and one by items (F_2). On the basis of these, *minF'* was computed (Clark, 1973), so that the results can be generalised over participants and items simultaneously, while keeping the experiment-wise error rate low (Barr, et al., 2013, p. 268). The experiment consisted of a 2 x 3 design, with factors Visibility (levels: screen, no screen) and Repetition (levels: initial, second, third), with initial references produced before feedback and second and third references produced after negative feedback from the matcher. We used post hoc analyses and only report where results are significant after correcting for multiple comparisons using the Bonferroni procedure.

Results

We first discuss effects of repetition and visibility on speech, followed by our main focus: effects of repetition and visibility on gesture rate, and on gesture form.

Effects on speech In table 5.1, we show the means and standard errors of the dependent speech variables for all three object descriptions. Firstly, inspection of table 5.1 reveals that the second references (after negative feedback) were shorter in duration than the initial references, while third references (also following negative feedback) were in turn longer than the second references, but shorter than the initial ones. This effect of repetition was significant, $F_1(2,72) = 17.17$, $p < .001$, $\eta_p^2 = .323$; $F_2(2,9) = 7.20$, $p < .05$, $\eta_p^2 = .616$; $minF'(2,18) = 5.07$, $p < .05$. Post hoc Bonferroni analyses showed that all three references differed from each other (all $p < .05$).

Secondly, we found that the second references contained fewer words than the initial references. The third references contained more words than the second references, but fewer than the initial references (see table 5.1). This effect of repetition was significant, $F_1(2,72) = 29.22$, $p < .001$, $\eta_p^2 = .448$; $F_2(2,9) = 15.91$, $p < .01$, $\eta_p^2 = .780$; $minF'(2,21) = 10.29$, $p < .001$. Post hoc Bonferroni analyses showed that the initial references differed from the second references and from the third references (both $p < .01$). The second and third references did not differ significantly from each other.

Thirdly, as expected, we saw that speech rate (measured in number of words per second) was lower for each following reference (see table 5.1). Again, this effect of repetition was significant, $F_1(2,72) = 30.61$, $p < .001$, $\eta_p^2 = .460$; $F_2(2,9) = 18.19$, $p < .01$,

η_p^2 = .802; *minF'* (2,22) = 11.40, $p < .001$. Post hoc Bonferroni analyses showed that all references differed from each other (all $p < .01$).

Table 5.1. Overview of means and standard errors (SE) for dependent variables in speech (duration in seconds, number of words, and speech rate in number of words per second), as a function of Repetition (three levels). Star indicates significant *minF'*.

	Initial (SE)	Second (SE)	Third (SE)
Duration*	39.7 (2.5)	28.9 (1.6)	33.2 (1.8)
Number of words*	85.0 (6.0)	55.4 (3.4)	58.7 (3.9)
Speech rate*	2.1 (.05)	1.9 (.05)	1.7 (.05)

Turning to the effect of visibility on speech, we found that for all three speech variables, a lack of mutual visibility between the director and the matcher caused numbers to go down (see table 5.2), although these reductions were only numerical, and not statistically significant. There were no significant interactions between repetition and visibility.

Table 5.2. Overview of means and standard errors (SE) for dependent variables in speech (duration in seconds, number of words, and speech rate in number of words per second), as a function of Visibility (two levels).

	Visibility (SE)	No visibility (SE)
Duration	35.3 (2.3)	32.5 (2.3)
Number of words	72.5 (5.5)	60.2 (5.5)
Speech rate	2.0 (.06)	1.8 (.06)

Effects on gesture rate In table 5.3, the means and standard errors of all the dependent variables in gesture for all three object descriptions can be found. Below we discuss them in more detail, starting with number of gestures and gesture rate.

First, we counted the number of gestures per trial. In absolute numbers, fewer gestures were produced in the second references (following negative feedback) than in the initial references (before negative feedback), and more gestures were produced in the third references (also following negative feedback) than in the second references (see table 5.3). However, this effect of repetition was only significant over participants and not in the *minF'* analysis, and hence cannot be considered statistically reliable, F_1 (2,72) = 4.88, $p < .05$, $\eta_p^2 = .119$; F_2 (2,9) = 1.5, $p = .27$, $\eta_p^2 = .250$; *minF'*(2,15) = 1.14, $p =$ .34.

Given that the number of words also varies from one description to the next, the gesture rate (number of gestures per 100 words) is more important to analyse, and table 5.3 shows that after each instance of negative feedback a higher gesture rate is observed. This effect was significant over participants and items, and marginally significant in *minF'*, F_1 (2,72) = 7.1, $p < .01$, $\eta_p^2 = .165$; F_2 (2,9) = 4.8, $p < .05$, $\eta_p^2 = .516$; *minF'*(2,24) = 2.86, $p = .077$. Post hoc Bonferroni analyses showed that the gesture rate of the initial references differed from the gesture rate of the third references ($p < .01$).

Table 5.3. Overview of means and standard errors (SE) for dependent variables in gesture (number of gestures, gesture rate (in number of gestures per 100 words), gesture duration (in seconds), gesture size (range 1-4), number of hands (range 1-2, with e.g. 1.4. indicating that 40% of gestures were two-handed), and stroke repetition (number of repeated strokes)), as a function of Repetition (three levels). Star indicates marginally significant minF'.

	Initial (SE)	Second (SE)	Third (SE)
Number of gestures	3.3 (.49)	2.6 (.38)	3.3 (.52)
Gesture rate *	4.1 (.67)	4.8 (.79)	5.3 (.74)
Gesture size	2.9 (.10)	2.9 (.09)	2.9 (.09)
Number of hands	1.5 (.06)	1.4 (.06)	1.3 (.05)
Stroke repetition*	.33 (.06)	.50 (.10)	.55 (.09)

In table 5.4 (see below), the means and standard errors of all the dependent gesture variables in the two visibility conditions can be seen. There was a numerical, but not statistically significant, decrease both in the absolute number of gestures, and in gesture rate, when there was no mutual visibility. There were no significant interactions between repetition and visibility on number of gestures or on gesture rate.

Effects on gesture form When we look at aspects of gesture form (see again table 5.3), the statistical analyses showed no significant effect of repetition after negative feedback on gesture size or the number of hands that were used to produce the gestures. We did find a marginally significant effect of repetition on the number of repeated strokes, F_1 (2, 54) = 3.236, p = .06, η_p^2 = .107; F_2 (2,9) = 13.645, p < .05, η_p^2 = .752; $minF'$(2,62) = 2.61, p = .08, with an increase for each instance of negative feedback. However, post hoc Bonferroni analyses showed that the three descriptions did not differ significantly from each other.

Turning to the effect of visibility on gesture form (see table 5.4), we firstly found that there was no effect of visibility on the number of hands or on the number of repeated strokes. There was, however, an effect of visibility on gesture size, F_1(1,27) = 9.009, p < .01, η_p^2 = .250; F_2 (1,9) = 77.642, p < .001, η_p^2 = .896; $minF'$(1,32) = 8.072, p <.01, with gestures produced when there was a screen between the director and the matcher being smaller than gestures produced when there was no screen between the director and the matcher. There were no significant interactions between repetition and visibility for any of the aspects of gesture form that were analysed.

Table 5.4. Overview of means and standard errors (SE) for dependent variables in gesture (number of gestures, gesture rate (in number of gestures per 100 words), gesture size (range 1-4), number of hands (range 1-2, with e.g. 1.4. indicating that 40% of gestures were two-handed), and stroke repetition (number of repeated strokes)), as a function of Visibility (two levels). Star indicates significant *minF'*.

	Visibility (SE)	No visibility (SE)
Number of gestures	3.4 (.63)	2.8 (.63)
Gesture rate	5.1 (1.0)	4.3 (1.0)
Gesture size*	3.1 (.10)	2.7 (.11)
Number of hands	1.4 (.07)	1.3 (.07)
Stroke repetition	.41 (.09)	.52 (.10)

Summarising the findings of experiment I, we found that references after negative feedback had a lower speech rate and a marginally significant higher gesture rate than initial references. In addition, gestures after negative feedback contained marginally more repeated strokes. When there was no visibility between the director and the matcher, gestures were smaller.

Experiment II: Precision judgment

In addition to the gesture measure analyses of the production experiment (experiment I), a separate precision judgment study was run to see whether there might (also) be differences in form between initial gestures and repeated gestures following negative feedback which are more gradual in nature than could be established using the discrete annotations of the data obtained in the production experiment. In this precision judgment experiment, as the name suggests, participants judged the precision of gestures. The setup of this precision judgment experiment, as was the case for the production experiment, closely follows the method used in our previous work on repeated, successful references, as presented in chapter 3.

Participants

Twenty nine participants (15 male, 14 female, age range 16-55 years old, M = 30 years old), who had not taken part in the production experiment and who had no knowledge of our other previous experiments, took part in the precision judgment experiment, without receiving any form of compensation.

Stimuli

For the precision judgment experiment, 44 trials were constructed, consisting of 44 pairs of video clips which were selected from the dataset collected in the production experiment. The pairs of video clips consisted of one video clip of a gesture taken from an initial description, and one video clip of a gesture following negative feedback, taken either from a second or third description. We selected all gesture pairs (44) that matched the following criteria. Each pair of gestures was taken from descriptions produced by the same director and both gestures in a pair referred to the same part of the same target object. No more than two gesture pairs produced by one director were used. Also, there had to be an equal distribution between gestures from second and from third descriptions. Of the 44 pairs of video clips, 23 were pairs consisting of one gesture from an initial description and one gesture from a second description, and 21 were pairs consisting of one gesture from an initial description and one from a third description. Finally, we aimed for an equal distribution between visibility conditions. For 19 of the 44 pairs, the gestures were taken from directors who were not able to see the matcher during the production experiment, and the remaining 25 pairs were taken from directors who were able to see each other.

Video clips were presented next to each other in pairs on a computer monitor, and the order in which the clips were presented on the screen was counterbalanced over trials. We presented participants with pairs, and not triads, of gestures, because there were not a sufficient number of directors producing a gesture about the same part of the same object in all three descriptions. Note, however, that in the analyses we did also consider possible differences between gestures from second and third references.

Procedure

The participants were presented individually with the 44 pairs of video clips. For each pair of video clips, the participants had to judge which of the two gestures they considered to be 'the most precise', where we expected gestures produced during repeated descriptions (i.e. following negative feedback) to be judged more precise than gestures from initial descriptions. No instructions were given with regard to what aspect(s) of the gesture the participants should take into account when making this judgment. The experiment was a forced choice test, presented without sound, and the participants were allowed to watch a video clip more than once if they wanted to. However, they were encouraged to go with their first intuition, and participants made hardly any use of the possibilities for replaying stimuli.[16]

[16] For our study on successful repeated references, reported in chapter 3 of this thesis, we conducted a very similar judgment study, and also experimented with different variants. In particular, in one variant participants were shown the target Greeble along with the two gesture stimuli, and were explained what was intended with one gesture being more "precise" than another ("for example when it provides more information about the shape of the object or when a gesture is more complex", following Gerwing & Bavelas, 2004). Neither of these adaptations influenced our previous findings, which is why we opt for the simplest variant (without Greeble picture and explanation of precision) here.

Data analysis

In each trial, in line with our expectation, a score of one (1) was assigned when the gesture following negative feedback was chosen by the participant to be the most precise, and a score of zero (0) when the participant chose the initial (pre-feedback) gesture to be the most precise. A binomial test was performed to see whether repeated gestures, after negative feedback, were chosen more often than initial gestures to be the most precise one of the two; in addition, a chi square analysis was conducted on the total scores (i.e. number of times that the gesture following negative feedback was chosen to be the most precise), with repetition (pairs of initial and second gestures versus pairs of initial and third gestures) and visibility (mutual visibility versus no mutual visibility) as independent variables.

Results

Repeated gestures were chosen to be the most precise in 673, or 53%, of 1276 cases, and initial gestures were chosen to be the most precise in 603, or 47%, of cases. This difference from chance level was marginally significant, $p = .053$.

Table 5.5 shows the distribution of scores for the number of times a gesture following negative feedback was chosen to be the most precise, as a function of repetition (second or third description) and visibility. A chi-square test of independence was conducted to examine the relation between repetition and visibility. We found a significant relation between repetition and visibility, $\chi^2(1) = 15.303$, $p < .001$. A chi-square test of goodness-of-fit showed that there was an equal distribution between repeated gestures from second references and from third references, $\chi^2(1)= 1.618$, $p = .203$. However, there was not an equal distribution between gestures taken from contexts of mutual visibility and gestures taken from contexts without visibility. Gestures following negative feedback which were produced with mutual visibility were chosen more often to be the most precise than gestures following negative feedback which were produced without mutual visibility, $\chi^2(1)= 25.499$, $p < .001$.

Table 5.5. Distribution of scores (and percentages) for number of times a repeated gesture (i.e. following negative feedback) was chosen to be the most precise, as a function of repetition (i.e. was the repeated gesture from a second or from a third description) and visibility (i.e. was the gesture produced with mutual visibility, or not).

	Second description	Third description	Total
Visibility	216 (32%)	186 (28%)	402 (60%)
No visibility	104 (15%)	167 (25%)	271 (40%)
Total	320 (47%)	353 (53%)	673 (100%)

General discussion

When a speaker describes an object or person, the addressee may or may not be able to determine which object or person is referred to. In the former case, when referential communication is successful, the addressee may either explicitly or implicitly indicate this to the speaker using a "go on" feedback cue, and the interaction continues. But in the latter case, when communication is unsuccessful, the addressee will signal this using a more marked "go back" feedback cue (e.g., Krahmer, et al., 2002; Shimojima, et al., 2002). We know that these negative "go back" cues have an impact on the next utterance of the speaker, which is more likely to be articulated with increased prosodic effort (higher pitch, louder volume, slower speech rate) and to be reformulated or rephrased (e.g., Litman, et al., 2006; Oviatt, et al., 1998, among many others). But what is the effect of negative, "go back" feedback on gesture production? Only a very limited number of studies have addressed this question so far, of which Holler and Wilkin (2011) is the most explicit, also in stressing that more research in this field is urgently needed.

In this chapter, we investigated what happens in gesture when referential communication is unsuccessful. Specifically, we studied repeated references to hard to describe objects (Greebles) with different shapes and protrusions, comparing initial descriptions with descriptions produced after negative feedback. Our experimental method was a variation of earlier work, presented in chapter 3 of this thesis, on successful referential communication to these Greebles, and we know from these studies that the Greebles reliably elicit spontaneous shape gestures, both during initial and repeated references. In general, we rely on a variant of the director-matcher referential communication paradigm (e.g., Clark & Wilkes-Gibbs, 1986; de Ruiter, et al., 2012; Holler & Stevens, 2007; Krauss & Weinheimer, 1966), combined with a visibility manipulation such that some participant pairs could see each other (mutual visibility), while others could not. Crucially, in a number of cases, an initial object description was followed by two, consecutive instances of negative, “go back” feedback, indicating that the addressee was not able to match the correct Greeble object to the description of the speaker. As in various earlier studies using the referential communication paradigm (including Hoetjes, et al., 2015; Holler & Wilkin, 2011), we look at both the gesture rate (in number of gestures per 100 words), before and after negative feedback, as well as the influence of feedback on the way directors produce gestures. Our analysis of gesture form consisted of both a detailed analysis of ‘discrete’ properties of the gestures (their size, number of hands involved and number of stroke repetitions), as well as a separate precision judgment experiment, in which naïve judges were asked to determine which of two gestures (one produced before and one after negative feedback) they considered to be the most “precise”.

We found, first of all, a marginally significant increase in gesture rate in repeated references following negative feedback, indicating that our speakers started to rely relatively more on the gesture modality when facing referential communication problems. This is different from the pattern that was observed in chapter 3, where gesture rate did not change across repeated, successful references. In general, many studies looking at gesture rate in successful communication found that gesture rate remains either the same or is reduced when speakers present information that is repeated or otherwise given in unproblematic interactions (e.g., de Ruiter, et al., 2012; Galati & Brennan, 2014; Jacobs & Garnham, 2007, see chapter 3 of this thesis for further discussion). Interestingly, the exception is formed by the work of Holler and colleagues, who found that gesture rate increases with repetition in successful communication (Holler, et al., 2011), but not after addressee feedback (Holler & Wilkin, 2011). In general, it is difficult to compare gesture rate across different studies (in which speakers are performing different tasks and talk about different things, which in turn may trigger different kinds of gestures), which is one of the main reasons why we opted for re-using the paradigm of our earlier study. In addition, due to the fact that the findings of the present study did not reach significance, it is difficult to relate them to previous findings on gesture rate.

However, gesture rate alone is perhaps not sufficiently informative when studying gesture production, a point also made recently by Bavelas and Healing (2013). Gesture form is important as well. Concerning form we found that gestures produced after negative feedback were somewhat more likely to contain repeated strokes (Experiment I) and to be judged as marginally more precise (Experiment II). Again, these patterns are clearly different from what we observed in chapter 3, where repeated (successful) references did not contain more strokes (in fact, no changes in 'discrete' gesture form were found), and where gestures from repeated references were *less* likely to be judged as precise than those in initial references.

On balance, the picture that emerges is that references after negative feedback (and in contrast to successful repeated references) showed a tendency towards relying more on gesture (increased gesture rate), and that these gestures showed a tendency towards being produced with more effort (more stroke repetition, more precision), but more research is needed to support this pattern due to the marginality of the statistical effects. This pattern of results seems to be consistent with earlier findings on the influence of negative feedback on speech and language (e.g., Litman, et al., 2006; Oviatt, et al., 1998), and notice, incidentally, that the decrease in speech rate which we observed matches these earlier findings as well.

It is informative to look at some examples of the kind of descriptions that our participants actually produced in this experiment. Example 5.1 illustrates the increase in gesture rate in the present study.

Example 5.1. Repeated descriptions of the same object by participant number 36 (in the no visibility condition), in translation from Dutch original, followed by original number of words, number of gestures and gesture rate. The moment at which a gesture was produced is placed between square brackets (dots indicate silence).

Initial description, before feedback

"Uh this one is [again wide in the middle] and thin at the top and the bottom. Uh the circle is a bit average uh in the circle there are three uh points. And at the top there is one and it edges a little [yes it is on the right side but it] also

> stands a bit to the front. Uh let me think. Uh there are one, two, three, four, four of this shape I think and this is the only one where three [of those] points are at the bottom".

89 words, 3 gestures, gesture rate 3.37

Second description, after negative feedback

> "Yes no that is not true I uh am saying it wrong. Yes there are [two where three] are uh let's have a good look, yes there are two which have three of those uh points at the bottom, only with that one it is again uh uh [it again has the shape of an uh] […] of such a [yes] [such a handle] of something and the others are a bit more pointy".

71 words, 5 gestures, gesture rate 7.04

Third description, after repeated negative feedback

> "Uh let's see. The difference still with those others is that that point at the top that that one does not have those [uh uh] how do you call that [that sort of detail in it], has [detail in it].

37 words, 3 gestures, gesture rate 8.11

Inspection of this example confirms, first of all, that talking about Greebles is hard, but it also illustrates what causes the increase in gesture rate that we observed. While speakers use fewer words in descriptions after negative feedback, they continue to rely on shape gestures, since these express the most distinguishing properties of the target Greeble.

Figure 5.4 illustrates increased gesture precision after negative feedback, as compared to before feedback was given. Notice that the gesture after negative feedback is produced at a higher location and shows a larger displacement of the speaker's hands than the gesture before feedback, consistent with the notion that after negative feedback, gestures are produced with more effort.

Figure 5.4. Example of a pair of gestures produced about the same object by the same participant (in the visibility condition), illustrating gesture precision. The gesture on the left is an initial gesture, produced before feedback, the gesture on the right is a gesture produced after negative feedback, which was judged to be more precise. Arrows indicate path and direction of each gesture.

Since it was used in many relevant earlier studies (most notably for our current purposes in the study presented in chapter 3 of this thesis, but also, for instance, in Alibali, et al., 2001; Bard, et al., 2000; Bavelas, et al., 2008; de Ruiter, et al., 2012; Holler, et al., 2011; Mol, et al., 2009), we included mutual visibility as a factor in our current experiments as well. As in the study in chapter 3, and many other studies, we found that gestures produced without visibility were smaller than those produced when there was mutual visibility between director and matcher (see figure 5.5). Perhaps more interestingly, we found in the judgment study that when there was mutual visibility, gestures produced after negative feedback were somewhat more likely to be judged as precise than initial, pre-feedback gestures. This suggests that our directors put more effort in their post-feedback gestures when these could be seen by their addressee, which in turn might imply that these gestures were communicatively intended. Notice that this is also in accordance with Holler and Wilkin's (2011) finding that gestures after feedback were "more communicative".

Figure 5.5. Example of gestures produced in context of visibility (on the left), and in context without mutual visibility (on the right, part of the opaque screen is just visible), illustrating gesture size. Arrows indicate path and direction of each gesture. In the gesture on the left, the entire arm is moving, whereas in the gesture on the right only the hands are moving.

As mentioned before, not many studies have investigated the effect of feedback on gesture production, especially not with regard to the question of how speakers adapt the frequency and form of their gestures. One notable exception, as discussed, is the study on the effect of addressee feedback on gesture production by Holler and Wilkin (2011). As we have seen, our findings, in particular those related to gesture form, appear to be consistent with theirs; after (negative) feedback, gestures appear to be more effortful and communicative. It is interesting to observe that this convergence of results is obtained despite differences in experimental set-up which were partly motivated from their suggestions for further research (Holler & Wilkin, 2011, p. 3534): different kinds of feedback (even though all, as said, are intuitive "go back" signals) which were administered in a different way, different gesture analyses, and different languages. Additionally, while in the current study we compared initial references with *two* instances following negative feedback, Holler and Wilkin (2011) offered at most *one* instance of negative feedback for an individual referent or event. Moreover, we added a visibility manipulation, as well as a separate gesture precision judgment experiment, adding further evidence that gestures after (negative) feedback are somewhat more precise, in particular when they were visible for the addressee.

Various avenues for future research remain. We opted for artificial negative feedback (a low buzzing sound), identical for all participants, administered by a matcher who otherwise remained neutral in her verbal and non-verbal feedback, and did not further interact with the directors. This kind of high level feedback, which may be likened to a “huh?” or “sorry?”, indicating that the previous utterance from the director was not successful, has been used before and has the advantage for current purposes that it allowed us to collect comparable descriptions, including gestures, before and after negative feedback, to see how speakers (our unit of analysis, cf. Bavelas & Healing, 2013) adapt their gestures after negative feedback. However, we cannot rule out the possibility that occasionally the matcher did produce some unintentional nonverbal feedback, which the director could subsequently have picked up. In addition, the matcher timed the occurrence of the negative feedback to produce it at the contextually appropriate time, but this also may have introduced some timing differences across trials. In follow up research, it would be important to see whether the findings obtained in the current, controlled set-up, generalize to more natural situations. Ideally, this would involve spontaneous interactions between pairs of naïve participants, rather than between participants and a confederate, to rule out any undesired experimental side effects of using the latter (cf. Kuhlen & Brennan, 2013). This could involve, for example, communication about Greebles as well, in which miscommunications (of various kinds) may occur in a more natural way.

It is to be expected that, in such a setting, different kinds of feedback and, related, different kinds of interaction, could lead to different gesture patterns. Imagine, just by way of example, that a director describes (in speech and gesture) a Greeble from the Radok family, with a cylindrical main shape. In the current experiment, such an utterance would be followed by general negative feedback. But now consider a different, more specific form of negative feedback, in which the matcher asks (incorrectly) “you mean the one with a vase shape?” (i.e., a “Galli”), indicating this vase shape using a gesture. This “go back” signal from the matcher would likely also initiate a repair from the director (“No, cylindrical.”), and may result in a pair of spontaneous cylindrical gestures before and after feedback (comparable to the pairs collected with the current paradigm, except that the negative feedback was specific rather than general). It would be very interesting to compare such pairs (assuming they can be collected in sufficiently large numbers) using a more natural variant of the methodology of the current paper, where we predict that, crucially, the post feedback gestures will be realized with more effort (e.g., more repeated strokes along a virtual cylinder) and are more likely to be judged as precise compared to the pre-feedback counterpart, perhaps to a larger extent than found in the current study.

Related, it would be interesting to see whether our current findings can be generalised to other types of gesture. In the present study, almost all gestures that were produced by directors were representational, and specifically iconic, ones. This was to be expected, since the stimuli were selected on the basis of their differences in shape and protrusions and thus afforded in particular the production of iconic gestures. A question is whether an increase in gesture rate and gesture form similar to what we found in the present study could be seen if the gestures in question were, for example, deictic or beat gestures (or metaphoric gestures or emblems, for that matter). There has been at least one study investigating deictic gestures in repeated references (de Ruiter, et al., 2012), but this study did not focus on miscommunication, and studied gesture rate, and not gesture form. It would be interesting to include negative feedback in that type of study, either in the controlled manner ("beep!") of the current study, or the less-controlled, but more natural alternative just sketched ("You mean this one?", while pointing to an incorrect object).

Finally, a last aspect that could be studied in future work concerns the gesture rate, where our findings (marginally significant increase in gesture rate after negative feedback) do not match those of Holler and Wilkin (2011) (no increase after feedback). As we discussed in detail in chapter 3, the study of gesture rate (as a dependent variable in different kinds of studies) has given rise to a complex pattern of results, which may partly be due to different ways in which gesture rates have been computed in the past. In future research, it would seem to be important to more systematically compare different ways of computing gesture rates, to get a better understanding of what these rates may tell us, and why the results can differ from one study to the next. In addition, as we already pointed out above, it becomes increasingly important to combine analyses of gesture rate with analyses of gesture form, to get a better understanding of the gestures that speakers produce.

Conclusion

In this study, we asked what happens in gesture when referential communication is unsuccessful. We conducted a director-matcher task in which directors had to produce repeated references about the same object after negative feedback which indicated that communication was unsuccessful. We found that after negative feedback, there was a marginally significant increase in gesture rate and gestures were produced with somewhat more repeated strokes (also marginally significant in *minF'*). In addition, a separate precision judgment test showed that after negative feedback, gestures were somewhat more likely to be rated as most precise, compared to gestures produced before negative feedback was given. Taken together, we suggest that this means that when communication was unsuccessful in our task, speakers showed a tendency towards relying more on gesture, and the gestures they produced trended towards being more effortful. In addition, the visibility manipulation suggests that our directors put more effort in their gestures when these could be seen by the addressee, which in turn might imply that these particular gestures were communicatively intended. All in all, the picture that emerges is rather different from our earlier reduction findings for successful repeated references, as presented in chapter 3; when communication is successful and information becomes more predictable, speakers can permit themselves to put less effort in their repeated references, both in speech (e.g., less clear articulation, fewer words) and in gesture (e.g., less precision). When communication is not successful, speakers have to make an extra effort, in an attempt to restore communicative success. We already knew that this increased effort has an impact on speech; the current paper suggests that it has a comparable effect on gesture production as well.

Acknowledgments

We would like to thank Elsa Jonkers and Kristel Bartels for their help in collecting the data for both experiments, Elsa Jonkers for additional help in transcribing and annotating the data, Ingrid Masson Carro for help in the reliability analyses, Nick Wood for technical support, and Martijn Goudbeek, Judith Holler and an anonymous reviewer for helpful comments and suggestions. An earlier version of this study was presented at the 36th Annual Conference of the Cognitive Science Society, Quebec (2014). We would like to thank the audience for useful comments.

6

General discussion and conclusion

In this thesis, we reported on four empirical studies that each discusses an aspect of reference production, with a specific focus on the production, perception and interpretation of gestures and signs. In this concluding chapter, we first provide a summary of all four studies, followed by a discussion of methodological and theoretical implications, and we end with suggestions for future work and a general conclusion.

Summary of the empirical chapters

In the first empirical and our most explorative chapter, entitled ‘Does our speech change when we cannot gesture?’ (chapter 2), the goal was to study whether people’s speech changes when they cannot use their hands. This study was inspired by work by Dobrogaev (1929), who claimed that when people cannot gesture, their speech becomes less fluent and more monotonous. The study in chapter 2 aimed to find support for this claim by conducting a production experiment in which speakers had to give addressees instructions on how to tie a tie. During half of the experiment, participants had to sit on their hands and thus could not gesture. Other factors that might influence the ease of communication, such as mutual visibility and previous experience, were also taken into account. We studied the effect of not being able to gesture on speech duration, on number of words, speech rate, number of filled pauses, and several acoustic measures. The results showed no support for the claim that the inability to gesture affects speech fluency or causes participants to talk more monotonously. Also, (lack of) mutual visibility did not have an effect on the dependent variables. However, we did find an effect of previous experience on several of the speech measures, with numbers going down (for duration, number of words, and number of filled pauses) with each repeated instruction. In addition to the production experiment, we conducted a perception study in which participants were presented with sound fragments from the production experiment, half of which were produced together with a gesture, and half of which were not. The question was whether participants would be able to hear whether a gesture was produced during the sound fragments, or not. The results showed that people were not able to hear whether someone gestures or not. In short, the results of this study showed no support for the claim that speech becomes less fluent and more monotonous when people cannot gesture.

It is, of course, difficult to decide what to conclude from the lack of significant findings related to not being able to gesture. In general, it is probably fair to say that if

Dobrograev's claim reflects a strong effect, we should have been able to find at least some evidence for it in one of the variables of interest. Although we do not want to maintain that gestures do not have any impact on acoustic or prosodic speech properties (see e.g., Krahmer & Swerts, 2007), it is worth noting that earlier studies finding positive evidence have mostly been based on more experimentally controlled speech materials, whereas our data were more spontaneously elicited, which means that other factors may have overruled the effect of gesturing per se.

The study in chapter 2 did not provide any support for the hypothesis that speech becomes less fluent and more monotonous when we cannot gesture, but it did show that speech is influenced by what we in this chapter called 'previous experience'. This result is in line with previous studies (e.g. Clark & Wilkes-Gibbs, 1986) showing that speech is reduced in cases of repeated references (since our variable 'previous experience' resulted in the production of repeated references), and paved the way for our next three studies, which all focused on the effect of repeated references, not in speech, but in the visual modality (sign and gesture).

In our second study entitled 'Reduction in gesture during the production of repeated references' (chapter 3), we focused explicitly on the effect of repetition, not only in speech, but especially in gesture. The study in chapter 3 consisted of three experiments, which all contributed to answering the question to what extent gesture reduction in repeated references is comparable to other forms of linguistic reduction in repeated references. In all three experiments, mutual visibility was taken into account so that we could also study to what extent reduction in gesture is more speaker- or listener driven. Given conflicting findings about the effect of repetition on gesture rate in previous studies (discussed in detail in chapter 3), in the first of the three experiments we systematically compared two measures of gesture rate: gesture rate per word and gesture rate per semantic attribute. We did this by conducting a production experiment in which speakers of Dutch had to repeatedly describe abstract objects to a listener and by annotating how many gestures were produced in the initial, second and third references to the same object. In addition to studying the effect of repetition on gesture rate, in the first experiment we manually annotated several discrete aspects of gesture form to see whether repetition (also) impacts the form of gestures. The hypothesis was that the form of gestures in repeated references would be reduced as compared to the form of gestures from earlier references, in line with similar observations for lexical (e.g., Brennan & Clark, 1996; Clark & Wilkes-Gibbs, 1986) and acoustic (e.g., Aylett & Turk, 2004; Bard, et al., 2000; Fowler, 1988) repetition. Apart from looking at discrete aspects of gesture form, in our second experiment we studied more gradual aspects of gesture form, using a perception experiment in which we presented naïve participants with pairs of video clips showing gestures from initial and repeated descriptions produced in the production experiment. The task for the participants was to decide which gesture was more precise, the one from an initial reference or the one from a repeated reference. The hypothesis was that reduction in gesture would be visible in less precise gestures in repeated references. Finally, in our third experiment, we studied how effective initial and repeated gestures are at communicating information. In this experiment, we presented participants with a gesture, as produced in an initial or repeated description during the production experiment, and asked them to decide which of a pair of objects the gesture represented. Here the hypothesis was that if gestures are reduced in repeated references, then they will be less communicative, which will in turn make the task for the participants harder.

The results from the first experiment presented in chapter 3 showed that with regard to gesture rate, we found no reduction in repeated references in terms of gesture rate per word, but we did find a U-shaped reduction pattern for gesture rate per attribute, with a reduction in second references, and an increase back to the level of the initial references in third references. The discrete annotations of gesture form showed no reliable effect of repetition on gesture form. However, the results from the second experiment showed that there were gradual differences in gesture form between initial and repeated references: participants judged gestures from repeated references as less precise than those from initial references. However, this decrease in precision in gestures produced in repeated references did not make these gestures less informative: results from the final experiment showed that participants were equally successful for initial and repeated gestures in deciding during which object description a gesture was produced. Besides effects of repetition, in chapter 3 we found that there were also some effects of visibility on gesture production. Gestures produced when there was no mutual visibility between the speaker and the listener were fewer in number, shorter in duration, smaller in size and less informative. We concluded in chapter 3 that there is reduction in gesture in repeated references, although with regard to gesture rate this may depend on the exact measure taken and with regard to gesture form there may be only gradual, and not discrete, reduction.

A question remaining after our second study was whether the findings of chapter 3 can be generalised, i.e. whether repeated references necessarily result in reduction. In our third and fourth study we studied this question by focusing on repeated references in two contexts that are different from the one in chapter 3: a context that differs with regard to the use of the visual modality, namely sign language (chapter 4) and a context which differs with regard to the status of the referring expression in the discourse, namely when communication is not successful (chapter 5).

In chapter 4, entitled 'Do repeated references result in sign reduction', we studied reduction in repeated references by speakers of Sign Language of the Netherlands (NGT). Again, we conducted both a production and a perception experiment, to see whether speakers of NGT reduce repeated references in sign language in similar ways as speakers of Dutch do in their (speech and) gesture. The hypothesis was that speakers of NGT will reduce their repeated references just like speakers of Dutch do, because speakers tend to be efficient language users, and it is efficient to reduce speech whenever possible (e.g., Jaeger, 2010). In this study we conducted a production experiment in which we asked speakers of NGT to describe pictures of objects to an addressee, and, like in the study reported in chapter 3, several objects had to be described repeatedly. We studied the duration of the initial and repeated descriptions, the number of signs that were used in the descriptions, and the duration of the signs themselves. A perception experiment with the same setup as the one used in chapter 3 was used to study whether signs produced in repeated references were perceived as less precise than signs produced in initial references.

The results from the production experiment showed systematic effects of repetition; repeated references were shorter, contained fewer signs, and shorter signs than initial references. The results from the perception experiment showed that non-signing participants (but not signing participants) considered signs produced in repeated references to be less precise than signs produced in initial references. Although the variables that were taken into account in this chapter were fairly coarse-grained and more research on the details of reduction in sign languages is clearly needed, the results from this study do suggest that a similar reduction process occurs in repeated references in NGT as has been found previously in speech and in gesture (e.g. in chapter 3). This means that we can generalise some of the findings from our study on reduction in gesture in repeated references to a context in which the visual modality serves a different role, namely in sign language.

In our final empirical study entitled 'On what happens in gesture when communication is unsuccessful' (chapter 5), we studied whether repeated references are still reduced (as they were in chapter 3) in a context in which reduction may not be beneficial for the communicative situation, namely when communication is unsuccessful. The hypothesis was that reducing a repeated reference would not be efficient or useful when communication is unsuccessful, and that this would become apparent by repeated references that are not reduced, but remain the same, or are increased with regard to the dependent variables taken into account. In this study, the fact that communication was unsuccessful was indicated by negative feedback given to the speaker by the addressee after an object description. As in the other studies, we conducted both a production and a perception experiment. Like in chapter 3, we conducted a production experiment in which speakers had to describe abstract objects to an addressee. Repeated descriptions of the same object had to be given after the addressee indicated to the speaker that she was unable to locate the intended object. In the production experiment, we studied gesture rate and discrete measures of gesture form. Again, more gradual differences in gesture form were studied in a separate perception experiment, in which participants had to judge which of two gestures (one from an initial and one from a repeated reference following negative feedback) they considered to be the most precise. As with the study reported in chapter 3, in both the production and the perception study reported in chapter 5 we included mutual visibility between the speaker and the addressee as a between subjects factor.

Results showed that after negative feedback, there was a marginally significant increase in gesture rate per word in repeated references. We found little evidence for a change in discrete measures of gesture form, apart from a marginally significant increase in the number of repeated strokes after negative feedback. With regard to the gradual differences in gesture form, we found that gestures that were produced after negative feedback were judged as marginally more precise than gestures from initial references which were produced before any negative feedback. Although the effects were fairly subtle, the results from this study do show that repeated references following negative feedback are not necessarily reduced, which is markedly different from the reduction patterns we observed in chapter 3 and 4. As in the previous studies, lack of mutual visibility between the speaker and the addressee had a reducing effect, although in this study only on the size of the gestures that were produced. Concluding, our final study showed that in the production of unsuccessful repeated references, a different process than found for successful repeated references can take place, with gesture production in repeated references not being reduced (which makes them less communicative), but with speakers appearing to keep their repeated references constant or by putting more effort in their repeated references, and thus making the repeated references more communicative.

Methodological implications

There are several methodological implications resulting from the work presented in this thesis. Firstly, there are some methodological implications with regard to the use of gesture metrics. In general, the gesture metric to use in a particular study naturally depends on the research question and on the type of data that is available. This means that many different gesture metrics can be used and have been used in previous studies (see a discussion of this in chapter 3). This can make comparison between different studies quite difficult. In this thesis, we attempted to address this issue in two ways: firstly by directly comparing different metrics that measure the same variable, namely gesture rate, within one study, and secondly by using the same metrics across studies whenever possible, allowing comparison between studies. We will briefly discuss both ways.

In chapter 3, one of the goals of the study was to compare two ways in which gesture rate can be measured. We compared number of gestures per word and number of gestures per semantic attribute. We saw that the two measures showed different results: repetition did not affect the gesture rate per word, but it did affect the gesture rate per attribute. What methodological implication can this difference between gesture rate metrics have? As we discussed in chapter 3, if, in a particular task or semantic domain, there is no one-to-one relationship between number of words and number of attributes, then computing both gesture rate per word and gesture rate per attribute can be informative. Which measure to use depends on the researcher's position with regard to the relationship between speech and gesture, since the two measures of gesture rate reflect differences in this relationship (namely, at the word level or at a semantic level).

However, computing gesture rate per attribute means that a semantic representation for a task needs to be defined, which can be complicated, and semantic annotation may be time consuming and therefore not always feasible.

In this thesis we used many of the same gesture metrics across chapters, so that comparison between studies was possible. This was especially relevant for chapters 3 and 5, and, to a lesser extent, also chapter 4 (see the discussion sections of chapters 4 and 5 for comparison between their respective results and those from chapter 3). Naturally, it makes most sense to use the same gesture metrics when studies are concerned with the same topic, as was the case in three of the studies presented in this thesis (chapters 3, 4 and 5). However, many previous studies with related topics have used differing gesture annotation methods (see discussion of this in the introduction of chapter 3). This may have to do with habit, personal preferences, but also with the fact that it can be quite difficult to establish reliable measures. In our case we saw, for example, in chapter 5 that reaching high interrater reliability was hard to obtain for some of our metrics, most notably for gesture duration. The low reliability results for duration (which is a continuous variable) in this study also suggest that the gesture field is in need of other measures to establish consistency between coders than the kappa statistics that have been used for the more categorically distinct variables (like gesture type).

Secondly, there are some methodological implications resulting from using a combination of production and perception experiments. As mentioned in the introductory chapter of this thesis, the assumption behind doing this was that by studying both production and perception we can separate what the speaker does (production) from what an addressee picks up (perception). By studying both production and perception, we were able to paint a more complete picture than if only one of the two had been taken into account. In some cases (chapters 2 and 4), the results from the production and the perception experiment were in agreement with each other, which meant that the perception study strengthened the production study, and vice versa. For instance, in chapter 4 we found that speakers of NGT reduced their repeated references, and this was visible not only in production (fewer and shorter signs), but also in perception (less precise signs). However, in chapters 3 and 5 we saw that using a combination of production and perception studies can also be useful because differences in results can show subtleties that might otherwise be missed. In chapter 3, for example, we saw in the production experiment that there was no effect of repetition on discrete aspects of gesture form. However, the perception experiment (experiment II) showed that there was an effect of repetition on more gradual aspects of gesture form. Likewise, in chapter 5 we saw in the production experiment that repetition did not affect many aspects of discrete gesture form, but the perception experiment showed that there was a (marginal) effect of repetition on gradual gesture form. We believe that studying a combination of gesture production and gesture perception has given added value to our studies and we would like to encourage this methodology for future work.

Thirdly, there may be some implications resulting from including visibility as a factor in the design of three of our four studies. As mentioned in the introductory chapter, the assumption was that including visibility as a factor would give us more insight into the extent to which gestures are produced for the speaker, or whether they are (also) produced for the addressee. Following previous studies (see Bavelas & Healing, 2013) we assumed that gestures can serve both a cognitive and a communicative role, and that the gestures that are (still) produced when there is no visibility between the speaker and the addressee mainly serve to help the speaker. Importantly for our studies, a lack of mutual visibility may affect not only gesture rate, but also gesture form (Bavelas, et al., 2008; Gullberg, 2006). The main goal of our studies with regard to visibility was to study to what extent changes in gesture, with regard to both rate and form, are more speaker- or more addressee-oriented. We found that in most, but not all, studies, lack of visibility between speaker and addressee caused reduction in aspects of gesture production. As we state in the discussion sections of chapters 3 and 5, we claim that the effects that we found of visibility on gesture form suggest that visible gestures were likely to be produced with the addressee in mind, and this could be relevant for interpreting speech-gesture models (which will be discussed below).

However, there has been criticism on studies with visibility designs such as ours, suggesting that the visibility manipulation may have confounded with addressee responsiveness (Bavelas & Healing, 2013). The idea here is that especially in cases where there is no free dialogue between the speaker and the addressee, there can still be nonverbal feedback between the speaker and the addressee when there is visibility, but when there is no visibility between the speaker and the addressee, not only will the addressee not see any gestures that might be produced, but there is also less possibility for the addressee to give any other nonverbal feedback to the speaker. This criticism can be applied to our studies also, since all of our studies were fairly controlled with respect to this factor, as there was no extended conversation or interaction between the speaker and the addressee. As mentioned in the introductory chapter, this was done on purpose, so that data from different stimuli and different participants was as comparable as possible and any differences could be attributed to the experimental manipulations. Although this was a conscious choice, preventing free interaction and focusing only on the speaker (who was our 'unit of analysis', Bavelas & Healing, 2013) is not in line with some previous work on more interactive processes of reference production (e.g. Brennan & Clark, 1996; Clark & Wilkes-Gibbs, 1986). This will be further discussed below in the section on future work.

Theoretical implications

The work presented in this thesis also has some theoretical implications, which we will now discuss.

When we take the various studies together, a general picture emerges whereby speakers design their multimodal utterances to be efficient. Earlier research has shown that when information is predictable, for example because it is repeated, speakers reduce this information, for instance by producing fewer words, which are in turn reduced acoustically as well. We have shown that a similar process applies to gesture production. Speakers have a tendency to produce fewer gestures when they are repeating information, and the gestures themselves are also reduced in some respects (chapter 3). This applies not only to co-speech gestures, but also to signs produced in sign language (chapter 4). However, this process only occurs when communication is successful. When communication is not successful, and previously introduced information cannot be considered given, repeated references are not reduced, but remain constant in their execution or are produced with more effort (chapter 5). In such a situation, speakers may ‘over-articulate’ their speech and gestures, in a way behaving similarly to when they would address a child or a non-native speaker, or someone who is naïve with respect to a specific domain. Presumably this is done in an increased attempt to make the message successful, while keeping the overall discourse as efficient as possible. The suggestion that speakers design their utterances to be efficient has been proposed previously, especially by Jaeger (2010) in his hypothesis of

Uniform Information Density (see also chapter 4). The idea here is that speakers prefer to distribute information evenly across the speech signal, at "all levels of linguistic representation" (Jaeger, 2010, p. 24). The assumption is that speakers try to ensure that the overall level of information remains the same at different points in time. This can be seen, for example, in speakers using words that are reduced in duration when these words have been used before in the discourse and thus contain given information, allowing a longer duration for words in the discourse containing new information. Jaeger tested this hypothesis at the level of syntactic production. In our studies we have shown that the hypothesis of Uniform Information Density can be applied to gesture production also.

As mentioned in the introductory chapter, in this thesis we did not aim to take a particular stand with regard to existing speech-gesture models. However, we would nevertheless like to shortly discuss whether the results from the three empirical chapters involving gesture show support for (a) particular speech-gesture model(s).

In chapter 2, we saw that speech fluency and monotony did not change when people could not gesture. As we mentioned in the discussion section of that chapter, it is difficult to determine what this means for existing speech-gesture models. When we consider the speech-gesture models based on Levelt's (1989) blueprint of the speaker, and assume that speech fluency and monotony are determined in the articulator, we cannot rule out or find strong support for one particular speech-gesture model, because a separation between speech and gesture at the level of the articulator is consistent with all models arguing that speech and gesture do not separate until before or in the conceptualizer.

In chapters 3 and 5 we saw that some of our results can be interpreted in light of the different speech-gesture models. We argued that the gesture rate per word measure can be related to the speech production process at the level of the formulator, which is fairly late in Levelt's (1989) speech production model. In chapter 3 we saw that gesture rate per word stayed the same in repeated references, indicating that in repeated references words and gestures were reduced to the same extent. This suggests that speech and gesture go 'hand-in-hand' (So, et al., 2009), which is in line with McNeill and Duncan's (2000) Growth Point Theory. However, the results on gesture rate per word in chapter 5 showed that when communication is unsuccessful, there is a marginally significant increase in the number of gestures per word in repeated references, caused by gesture production staying constant, while speech production (number of words) was reduced. This suggests that speakers may rely relatively more on gesture when communication is difficult, and indicates that speech and gesture production do not necessarily need to go hand in hand throughout the entire speech production process, but can also vary more independently, which seems more in line with speech-gesture models that assume a more independent production of speech and gesture, such as in the Sketch Model (de Ruiter, 2000).

In chapters 3 and 5 we saw that a lack of visibility caused some aspects of gesture production to be reduced. Our conclusions, based on this effect of visibility, suggest that at least certain gestures are communicative and are thus (also) produced with the addressee in mind. Most speech-gesture models leave implicit whether gesture production is done for the speaker or for the addressee. There is one model, the Process Model, (Krauss, et al., 2000) that assumes that gesture production has a facilitative function, and not a communicative one. Our claim in chapters 3 and 5, that the effect of visibility shows that gestures were (also) intended communicatively, does not support the Process Model.

Future work

In this thesis, each empirical chapter proposed its own suggestions for future work. Here, we would like to discuss in some more detail suggestions for future work which span more than one chapter.

Firstly, in all chapters involving gesture production, we analysed all speech-accompanying gestures, and did not distinguish between different types of gestures. In the introductory chapter we described the different types of gestures, and explained that they can, and usually have, different roles in a discourse. Different types of gesture may also behave differently in specific contexts, due to their differences both in form and in function. In the gesture studies reported in this thesis, we used stimuli that specifically afforded the production of iconic gestures. Due to the affordances of the stimuli it was indeed the case that overwhelmingly representational, and more specific, iconic, gestures were produced. Therefore, conducting separate analyses for different gesture types was practically impossible. However, it might be the case that the findings presented in this thesis only hold for iconic gestures (see also the discussion sections in chapters 2, 3 and 5), and that other types of gestures impact speech production (as in chapter 2) differently, or might be influenced differently by repetition (as in chapters 3 and 5) or lack of visibility (as in chapters 2, 3 and 5). One can imagine that the gradual form of, for example, beat gestures might not change to the same extent as what we saw for the gestures in chapters 3 and 5, for the simple reason that due to the generally simple form of beat gestures not much change in form may be possible at all. To study different types of gesture in future work, a task must be used which differs from the tasks used in this thesis insofar that it must either specifically afford the production of, for example, beat gestures or deictic gestures, or be set up in such a way that all types of gestures are likely to occur. For example, getting participants to produce mainly deictic gestures could be achieved by, say, asking participants to describe objects from an array of objects that is mutually visible yet placed fairly far away from both participants (as in de Ruiter, et al., 2012).

Secondly, although in all studies we analysed a range of dependent variables, not all studies included all conceivable variables. Perhaps most notably, in chapter 5, we only reported on gesture rate per word and only performed one perception study, looking at gesture precision. This was sufficient for our research purposes (to show that gesture behaviour in repeated references following negative feedback differs from that in 'ordinary' repeated references), but it would be interesting to look at gesture rate per attribute for this data set, as well as at the perceived informativity of gestures following negative feedback, as we did in chapter 3. We conjecture that gesture rate per attribute will increase after negative feedback, and that gestures following negative feedback are more informative than those preceding it, based on the idea that speakers rely relatively more on gesture in difficult communicative settings. Indeed, first pilot results on this data set point in this direction, and we will report on this in future research. In a somewhat similar vein, knowing more details about the way in which gesture and speech production are related at the lexical level in our studies offers a further interesting line for future research. Here, too, we have started some more qualitative pilot work, that might help inform us on the exact relationship between speech and gesture. In particular, we have started analysing the number of disfluencies that were produced in the data from the production experiment in chapter 3. Preliminary results show that participants who gestured less often produced more disfluencies. Additional qualitative analyses such as these could be taken up in future work.

Thirdly, a natural suggestion for future work is to conduct studies such as the ones presented in this thesis, but using a more natural setting, in which there is free interaction between the speaker and the addressee, and also in which no confederates are used (as was the case in chapter 5). As we mentioned in the discussion sections of chapters 3 and 5, we believe that by avoiding free interaction and by focusing only on the speaker, the results from our studies might have been on the conservative side, as compared to if there would have been free interaction. In free interaction, explicit feedback from the addressee could ensure that truly shared conceptual pacts (Brennan & Clark, 1996) are created, and, depending on the conversational context, these might consist of repeated references that are reduced more and are thus even less informative (when taken out of context) than in chapter 3, or consist of repeated references that are realised with more effort, and are thus even more communicative than in chapter 5. By conducting future work in a more natural communicative setting, the ecological validity of our findings can be strengthened, and it will be possible to better relate the work presented in this thesis to studies that assume collaborative reference production.

Conclusion

In this thesis, we studied reference in speech, gesture, and sign. In all four studies reported in this thesis, we focused on different aspects of 'talking hands'. In chapter 2 we saw that although we know that speech and gesture are closely related, when the hands cannot do the talking, no clear changes were observed in speech. In chapter 3 we saw that when the hands can do the talking, they can go hand in hand with speech and can be reduced in repeated references. In chapter 4 we saw that in sign language, when the hands have to do all the talking, the hands can be communicatively efficient in the same ways as we saw in speech and gesture in chapter 3. In chapter 5 we found that repeated references are not necessarily reduced, as was the case in chapters 3 and 4, but that in cases of miscommunication, when the hands are talking but not heard, an increase in aspects of gesture production can occur. In conclusion, we found that, when speakers *have* their hands at their disposal, they let them take part in the talking. This thesis started by describing the 'secret language' of the hands that my six-year-old self thought politicians were speaking. It turns out that this 'secret language' is not so secret after all, with speech and gesture both contributing to the same message.

Summary

Chapter 2

In the first empirical and our most explorative chapter, entitled 'Does our speech change when we cannot gesture?', the goal was to study whether people's speech changes when they cannot use their hands. This study was inspired by work by Dobrogaev (1929), who claimed that when people cannot gesture, their speech becomes less fluent and more monotonous. The study in chapter 2 aimed to find support for this claim by conducting a production experiment in which speakers had to give addressees instructions on how to tie a tie. During half of the experiment, participants had to sit on their hands and thus could not gesture. Other factors that might influence the ease of communication, such as mutual visibility and previous experience, were also taken into account. We studied the effect of not being able to gesture on speech duration, on number of words, speech rate, number of filled pauses, and several acoustic measures. The results showed no support for the claim that the inability to gesture affects speech fluency or causes participants to talk more monotonously. Also, (lack of) mutual visibility did not have an effect on the dependent variables. However, we did find an effect of previous experience on several of the speech measures, with numbers going down (for duration, number of words, and number of filled pauses) with each repeated instruction. In addition to the production experiment, we conducted a perception study in which participants were presented with sound fragments from the production experiment, half of which were produced together with a gesture, and half of which were not. The question was whether participants would be able to hear whether a gesture was produced during the sound fragments, or not. The results showed that people were not able to hear whether someone gestures or not. In short, the results of this study showed no support for the claim that speech becomes less fluent and more monotonous when people cannot gesture.

Chapter 3

In our second study entitled 'Reduction in gesture during the production of repeated references', we focused explicitly on the effect of repetition, not only in speech, but especially in gesture. The study in chapter 3 consisted of three experiments, which all contributed to answering the question to what extent gesture reduction in repeated references is comparable to other forms of linguistic reduction in repeated references. In all three experiments, mutual visibility was taken into account so that we could also study to what extent reduction in gesture is more speaker- or listener driven. Given conflicting findings about the effect of repetition on gesture rate in previous studies, in the first of the three experiments we systematically compared two measures of gesture rate: gesture rate per word and gesture rate per semantic attribute. We did this by conducting a production experiment in which speakers of Dutch had to repeatedly describe abstract objects to a listener and by annotating how many gestures were produced in the initial, second and third references to the same object. In addition to studying the effect of repetition on gesture rate, in the first experiment we manually annotated several discrete aspects of gesture form to see whether repetition (also) impacts the form of gestures. The hypothesis was that the form of gestures in repeated references would be reduced as compared to the form of gestures from earlier references, in line with similar observations for lexical (e.g., Brennan & Clark, 1996; Clark & Wilkes-Gibbs, 1986) and acoustic (e.g., Aylett & Turk, 2004; Bard, et al., 2000; Fowler, 1988) repetition. Apart from looking at discrete aspects of gesture form, in our second experiment we studied more gradual aspects of gesture form, using a perception experiment in which we presented naïve participants with pairs of video clips showing gestures from initial and repeated descriptions produced in the production experiment. The task for the participants was to decide which gesture was more precise, the one from an initial reference or the one from a repeated reference. The hypothesis was that reduction in gesture would be visible in less precise gestures in repeated references. Finally, in our third experiment, we studied how effective initial and repeated gestures are at communicating information. In this experiment, we presented participants with a gesture, as produced in an initial or repeated description during the production experiment, and asked them to decide which of a pair of objects the gesture represented. Here the hypothesis was that if gestures are reduced in repeated references, then they will be less communicative, which will in turn make the task for the participants harder.

The results from the first experiment presented in chapter 3 showed that with regard to gesture rate, we found no reduction in repeated references in terms of gesture rate per word, but we did find a U-shaped reduction pattern for gesture rate per attribute, with a reduction in second references, and an increase back to the level of the initial references in third references. The discrete annotations of gesture form showed no reliable effect of repetition on gesture form. However, the results from the second experiment showed that there were gradual differences in gesture form between initial and repeated references: participants judged gestures from repeated references as less precise than those from initial references. However, this decrease in precision in gestures produced in repeated references did not make these gestures less informative: results from the final experiment showed that participants were equally successful for initial and repeated gestures in deciding during which object description a gesture was produced. Besides effects of repetition, in chapter 3 we found that there were also some effects of visibility on gesture production. Gestures produced when there was no mutual visibility between the speaker and the listener were fewer in number, shorter in duration, smaller in size and less informative. We concluded in chapter 3 that there is reduction in gesture in repeated references, although with regard to gesture rate this may depend on the exact measure taken and with regard to gesture form there may be only gradual, and not discrete, reduction.

Chapter 4

In chapter 4, entitled 'Do repeated references result in sign reduction', we studied reduction in repeated references by speakers of Sign Language of the Netherlands (NGT). Again, we conducted both a production and a perception experiment, to see whether speakers of NGT reduce repeated references in sign language in similar ways as speakers of Dutch do in their (speech and) gesture. The hypothesis was that speakers of NGT will reduce their repeated references just like speakers of Dutch do, because speakers tend to be efficient language users, and it is efficient to reduce speech whenever possible (e.g., Jaeger, 2010). In this study we conducted a production experiment in which we asked speakers of NGT to describe pictures of objects to an addressee, and, like in the study reported in chapter 3, several objects had to be described repeatedly. We studied the duration of the initial and repeated descriptions, the number of signs that were used in the descriptions, and the duration of the signs themselves. A perception experiment with the same setup as the one used in chapter 3 was used to study whether signs produced in repeated references were perceived as less precise than signs produced in initial references.

The results from the production experiment showed systematic effects of repetition; repeated references were shorter, contained fewer signs, and shorter signs than initial references. The results from the perception experiment showed that non-signing participants (but not signing participants) considered signs produced in repeated references to be less precise than signs produced in initial references. Although the variables that were taken into account in this chapter were fairly coarse-grained and more research on the details of reduction in sign languages is clearly needed, the results from this study do suggest that a similar reduction process occurs in repeated references in NGT as has been found previously in speech and in gesture (e.g. in chapter 3).

Chapter 5

In our final empirical study entitled 'On what happens in gesture when communication is unsuccessful', we studied whether repeated references are still reduced (as they were in chapter 3) in a context in which reduction may not be beneficial for the communicative situation, namely when communication is unsuccessful. The hypothesis was that reducing a repeated reference would not be efficient or useful when communication is unsuccessful, and that this would become apparent by repeated references that are not reduced, but remain the same, or are increased with regard to the dependent variables taken into account. In this study, the fact that communication was unsuccessful was indicated by negative feedback given to the speaker by the addressee after an object description. As in the other studies, we conducted both a production and a perception experiment. Like in chapter 3, we conducted a production experiment in which speakers had to describe abstract objects to an addressee. Repeated descriptions of the same object had to be given after the addressee indicated to the speaker that she was unable to locate the intended object. In the production experiment, we studied gesture rate and discrete measures of gesture form. Again, more gradual differences in gesture form were studied in a separate perception experiment, in which participants had to judge which of two gestures (one from an initial and one from a repeated reference following negative feedback) they considered to be the most precise. As with the study reported in chapter 3, in both the production and the perception study reported in chapter 5 we included mutual visibility between the speaker and the addressee as a between subjects factor.

Results showed that after negative feedback, there was a marginally significant increase in gesture rate per word in repeated references. We found little evidence for a change in discrete measures of gesture form, apart from a marginally significant increase in the number of repeated strokes after negative feedback. With regard to the gradual differences in gesture form, we found that gestures that were produced after negative feedback were judged as marginally more precise than gestures from initial references which were produced before any negative feedback. Although the effects were fairly subtle, the results from this study do show that repeated references following negative feedback are not necessarily reduced, which is markedly different from the reduction patterns we observed in chapter 3 and 4. As in the previous studies, lack of mutual visibility between the speaker and the addressee had a reducing effect, although in this study only on the size of the gestures that were produced. Concluding, our final study showed that in the production of unsuccessful repeated references, a different process than found for successful repeated references can take place, with gesture production in repeated references not being reduced (which makes them less communicative), but with speakers appearing to keep their repeated references constant or by putting more effort in their repeated references, and thus making the repeated references more communicative.

References

Alibali, M., Heath, D. C., & Myers, H. J. (2001). Effects of visibility between speaker and listener on gesture production: Some gestures are meant to be seen. *Journal of Memory and Language, 44*, 169–188.

Alibali, M., Kita, S., & Young, A. (2000). Gesture and the process of speech production: We think, therefore we gesture. *Language and cognitive processes, 15*, 593-613.

Altmann, G. (1980). Prolegomena to Menzerath's law. *Glottometrika, 2*, 1-10.

Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., et al. (1991). The HCRC Map Task Corpus. *Language and Speech, 34*(4), 351-366.

Anderson, A. H., Bard, E. G., Sotillo, C., Newlands, A., & Doherty-Sneddon, G. (1997). Limited visual control of the intelligibility of speech in face-to-face dialogue. *Perception and Psychophysics, 59*(4), 580-592.

Arnold, J. E. (2008). Reference production: Production-internal and addressee-oriented processes. *Language and Cognitive Processes, 23*(4), 495-527.

Arnold, J. E., Kahn, J., & Pancani, G. (2012). Audience design affects acoustic reduction via production facilitation. *Psychonomic Bulletin & Review, 19*(3), 505-512.

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: a functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech, 47*(1), 31-56.

Bangerter, A. (2004). Using pointing and describing to achieve joint focus of attention in dialogue. *Psychological Science, 15*, 415-419.

Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language, 42*, 1-22.

Bard, E. G., & Aylett, M. (2005). Referential form, duration, and modelling the listener in spoken dialogue. In J. Trueswell & M. Tanenhaus (Eds.), *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions* (pp. 173-191). Cambridge: MIT Press.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255-278.

Bavelas, J., Chovil, N., Lawrie, D. A., & Wade, A. (1992). Interactive gestures. *Discourse Processes, 15*, 469-489.

Bavelas, J., Gerwing, J., Sutton, C., & Prevost, D. (2008). Gesturing on the telephone: Independent effects of dialogue and visibility. *Journal of Memory and Language, 58*, 495–520.

Bavelas, J., & Healing, S. (2013). Reconciling the effects of mutual visibility on gesturing: A review. *Gesture, 13*(1), 63-92.

Beattie, G., & Coughlan, J. (1999). An experimental investigation of the role of iconic gestures in lexical access using the tip-of-the-tongue phenomenon. *British Journal of Psychology, 90*, 35-56.

Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language 60*, 92-111.

Bernardis, P., & Gentilucci, M. (2006). Speech and gesture share the same communication system. *Neuropsychologia, 44*, 178-190.

Boersma, P., & Weenink, D. (2010). Praat: doing phonetics by computer (Version 5.1.25) [Computer program]. Retrieved January 20, 2010, from http://www.praat.org/

Bolinger, D. (1983). Intonation and gesture. *American Speech, 58*(2), 156-174.

Brennan, S., & Clark, H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology, 22*(6), 1482-1493.

Brentari, D. (1998). *A prosodic model of sign language phonology*: MIT Press.

Brentari, D. (2002). Modality and structure in signed and spoken languages *Modality differences in sign language phonology and morphophonemics* (pp. 35–64): Cambridge University Press.

Brentari, D., Gonzalez, C., Seidl, A., & Wilbur, R. B. (2011). Sensitivity to visual prosodic cues in signers and nonsigners. *Language and Speech, 54*(1), 49-72.

Bressem, J., & Ladewig, S. H. (2011). Rethinking gesture phases: Articulatory features of gestural movement? *Semiotica, 184*(1/4), 53–91.

Brown, G. (1983). Prosodic structure and the given / new distinction. In A. Cutler & D. R. Ladd (Eds.), *Prosody: Models and Measurements* (pp. 67–78). New York: Springer-Verlag.

Butterworth, B., & Beattie, G. (1978). Gesture and silence as indicators of planning in speech. In R. N. Campbell & P. T. Smith (Eds.), *Recent advances in the psychology of language: Formal and experimental approaches*. New York: Plenum.

Cavé, C., Guaïtella, I., Bertrand, R., Santi, S., Harlay, F., & Espesser, R. (1996). *About the relationship between eyebrow movements and F0 variations.* Paper presented at the The Fourth International Conference on Spoken Language Processing, Philadelphia, USA.

Chown, K. (2008). Reflex theory in a linguistic context: Sergej M. Dobrogaev on the social nature of speech production. *Studies in East European Thought, 60*, 307-319.

Chu, M., & Hagoort, P. (2014). Synchronization of speech and gesture: Evidence for interaction in action. *Journal of Experimental Psychology: General, 143*(4), 1726-1741.

Chu, M., & Kita, S. (2007). *Individual difference in the use of spontaneous gestures in a mental rotation task.* Paper presented at the International Society for Gesture Studies Conference (ISGS): Integrating Gestures, Evanston, IL, USA.

Clark, H. (1973). The Language-as-Fixed-Effect fallacy: A critique of language statistics in psychological research. *Journal of verbal learning and verbal behavior, 12*, 335-359.

Clark, H. (1996). *Using language*. Cambridge: Cambridge University Press.

Clark, H., & Brennan, S. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine & J. S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127-149): American Psychological Association.

Clark, H., & Krych, M. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language, 50*, 62-81.

Clark, H., & Schaeffer, E. F. (1989). Contributing to discourse. *Cognitive Science, 13*, 259-294.

Clark, H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22*, 1-39.

Cohen, A. (1977). The communicative functions of hand illustrators. *Journal of Communication, 27*, 54-63.

Cohen, A., & Harrison, R. P. (1973). Intentionality in the use of hand illustrators in face-to-face communication situations. *Journal of Personality and Social Psychology, 28*(2), 276-279.

Crasborn, O. (2001). *Phonetic implementation of phonological categories in Sign Language of the Netherlands*. PhD dissertation. Utrecht: LOT.

de Ruiter, J. P. (2000). The production of gesture and speech. In D. McNeill (Ed.), *Language and gesture* (pp. 284-311). Cambridge: Cambridge University Press.

de Ruiter, J. P. (2006). Can gesticulation help aphasic people speak, or rather, communicate? *Advances in Speech-Language Pathology, 8*(2), 124-127.

de Ruiter, J. P. (2007). Postcards from the mind: the relationship between speech, imagistic gesture, and thought. *Gesture, 7*(1), 21-38.

de Ruiter, J. P., Bangerter, A., & Dings, P. (2012). The interplay between gesture and speech in the production of referring expressions: Investigating the trade-off hypothesis. *Topics in Cognitive Science, 4*(2), 232–248.

Dobrogaev, S. M. (1929). Ucnenie o reflekse v problemakh iazykovedeniia [Observations on reflexes and issues in language study]. *Iazykovedenie i Materializm* 105-173.

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

Emmorey, K., & Casey, S. (2001). Gesture, thought, and spatial language. *Gesture, 1*(1), 35-50.

Enfield, N. (2001). 'Lip-pointing': A discussion of form and function with reference to data from Laos. *Gesture, 1*(2), 185-211.

Ernestus, M., & Warner, N. (2011). An introduction to reduced pronunciation variants [Editorial]. *Journal of Phonetics, 39*(SI), 253-260.

Fenk-Oczlon, G. (2001). Familiarity, information flow, and linguistic form. In J. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 431–448). Amsterdam/Philadelphia: John Benjamins.

Fenk, A., & Fenk-Oczlon, G. (1993). Menzerath's law and the constant flow of linguistic information. In R. Köhler & B. Rieger (Eds.), *Contributions to quantitative linguistics* (pp. 11-31). Dordrecht: Kluwer Academic Publishers.

Ferreira, V. S. (2008). Ambiguity, accessibility, and a division of labor for communicative success. *Learning and Motivation, 49*, 209–246.

Finlayson, S., Forrest, V., Lickley, R., & Mackenzie Beck, J. (2003). Effects of the restriction of hand gestures on disfluency. In R. Eklund (Ed.), *Proceedings of DiSS'03, Disfluency in Spontaneous Speech Workshop* (pp. 21-24). Goteborg.

Flecha-García, M. L. (2010). Eyebrow raises in dialogue and their relation to discourse structure, utterance function and pitch accents in English. *Speech Communication, 52*(6), 542-554.

Fowler, C. A. (1988). Differential shortening of repeated content words produced in various communicative contexts. *Language and Speech, 31*(4), 307-319.

Fowler, C. A., & Housum, J. (1987). Talkers' signaling of 'new' and 'old' words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language, 26*(5), 489-504.

Fridlund, A. J. (1994). *Human facial expression: An evolutionary view*. San Diego: Academic Press.

Galati, A., & Brennan, S. (2010). Attenuating information in spoken communication: For the speaker, or for the addressee? *Journal of Memory and Language, 62*, 35–51.

Galati, A., & Brennan, S. (2014). Speakers adapt gestures to addressees' knowledge: implications for models of co-speech gesture. *Language, Cognition and Neuroscience, 29*(4), 435-451.

Gatt, A., Krahmer, E., van Deemter, K., & van Gompel, R. P. G. (2014). Models and empirical data for the production of referring expressions. *Language, Cognition and Neuroscience, 29*(8), 899-911.

Gauthier, I., & Tarr, M. (1997). Becoming a "Greeble" expert: exploring mechanisms for face recognition. *Vision Research, 37*, 1673-1682.

Gee, J., & Goodhart, W. (1988). American Sign Language and the human biological capacity for language. In M. Strong (Ed.), *Language Learning and Deafness* (pp. 49-74). Cambridge: Cambridge University Press.

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

Goudbeek, M., & Krahmer, E. (2012). Alignment in interactive reference production: Content planning, modifier ordering and referential overspecification. *Topics in Cognitive Science, 4*, 269-289.

Graham, J. A., & Heywood, S. (1975). The effects of elimination of hand gestures and of verbal codability on speech performance. *European Journal of Social Psychology, 5*, 189–195.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics 3: Speech arts* (pp. 41-58). New York: Academic Press.

Gullberg, M. (2006). Handling discourse: gestures, reference tracking, and communication strategies in early L2. *Language Learning, 56*(1), 155-196.

Gullberg, M., De Bot, K., & Volterra, V. (2008). Gestures and some key issues in the study of language development. *Gesture, 8*(2), 149-179.

Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language, 69*, 274-307.

Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.

Healey, P., Mills, G. J., & Eshgi, A. (2013). Making things worse to make them better: The role of negative evidence in the coordination of referring expressions *Proceedings of Pre-Cogsci 2013: Production of referring expressions: Bridging the gap between cognitive and computational approaches to reference*. Berlin.

Hoetjes, M., Koolen, R., Goudbeek, M., Krahmer, E., & Swerts, M. (2011). GREEBLES Greeble greeb. On reduction in speech and gesture in repeated references. In L. Carlson, C. Hoelscher & T. F. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 3250-3255). Boston: Cognitive Science Society.

Hoetjes, M., Koolen, R., Goudbeek, M., Krahmer, E., & Swerts, M. (2015). Reduction in gesture during the production of repeated references. *Journal of Memory and Language, 79-80*, 1-17.

Hoetjes, M., Krahmer, E., & Swerts, M. (2014a). Do repeated references result in sign reduction? *Sign Language and Linguistics, 17*(1), 56-81.

Hoetjes, M., Krahmer, E., & Swerts, M. (2014b). Does our speech change when we cannot gesture? *Speech Communication*(57), 257-267.

Holler, J., & Stevens, R. (2007). The effect of common ground on how speakers use gesture and speech to represent size information. *Journal of Language and Social Psychology, 26*(1), 4-27.

Holler, J., Tutton, M., & Wilkin, K. (2011). Co-speech gestures in the process of meaning coordination *Proceedings of the 2nd GESPIN - Gesture & Speech in Interaction Conference*. Bielefeld.

Holler, J., & Wilkin, K. (2009). Communicating common ground: how mutually shared knowledge influences speech and gesture in a narrative task. *Language and Cognitive Processes, 24*(2), 267-289.

Holler, J., & Wilkin, K. (2011). An experimental investigation of how addressee feedback affects co-speech gestures accompanying speakers' responses. *Journal of Pragmatics, 43*, 3522–3536.

Hostetter, A. B., & Alibali, M. (2008). Visible embodiment: gestures as simulated action. *Psychonomic Bulletin & Review, 15*(3), 495-514.

Hostetter, A. B., & Alibali, M. (2010). Language, gesture, action! A test of the Gesture as Simulated Action framework. *Journal of Memory and Language, 63*, 245-257.

Hostetter, A. B., Alibali, M. W., & Kita, S. (2007). Does sitting on your hands make you bite your tongue? The effects of gesture prohibition on speech during motor descriptions. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th annual meeting of the Cognitive Science Society* (pp. 1097-1102). Mawhah, NJ: Erlbaum.

Iverson, J., M., & Goldin-Meadow, S. (1998). Why people gesture when they speak. *Nature, 396*, 228.

Jacobs, N., & Garnham, A. (2007). The role of conversational hand gestures in a narrative task. *Journal of Memory and Language, 56*, 291-303.

Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology, 61*(1), 23-62.

Jaeger, T. F., & Tily, H. (2011). Language Processing Complexity and Communicative Efficiency. *WIREs: Cognitive Science, 2*(3), 323-335.

Johnson, R. E., & Liddell, S. K. (2010). Toward a phonetic representation of signs: Sequentiality and contrast. *Sign Language Studies, 11*(2), 241-274.

Kaland, C., Krahmer, E., & Swerts, M. (2014). White bear effects in language production: evidence from the prosodic realisation of adjectives. *Language and Speech, 57*(4), 470-486.

Kelly, S. D., Manning, S. M., & Rodak, S. (2008). Gesture gives a hand to language and learning: Perspectives from cognitive neuroscience, developmental psychology and education. *Language and Linguistics Compass, 2*(4), 569-588.

Kendon, A. (1972). Some relationships between body motion and speech. In A. W. Seigman & B. Pope (Eds.), *Studies in dyadic communication* (pp. 177-216). New York: Pergamon Press.

Kendon, A. (1980). Gesture and speech: two aspects of the process of utterance. In M. R. Key (Ed.), *Nonverbal communication and language* (pp. 207-227). The Hague: Mouton.

Kendon, A. (1986). Some reasons for studying gesture. *Semiotica, 62*(1/2), 3-28.

Kendon, A. (2000). Language and gesture: unity or duality? In D. McNeill (Ed.), *Language and gesture* (pp. 47-63). Cambridge: Cambridge University Press.

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

Kendon, A. (2007). Some topics in gesture studies. In A. Esposito, M. Bratanic, E. Keller & M. Marinaro (Eds.), *Fundamentals of verbal and nonverbal communication and the biometric issue* (pp. 3-19). Amsterdam: IOS Press.

Kita, S. (2000). How representational gestures help speaking. In D. McNeill (Ed.), *Language and gesture* (pp. 162-185). Cambridge: Cambridge University Press.

Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language, 48*, 16-32.

Klima, E. S., & Bellugi, U. (1979). *The signs of language*. Cambridge, MA: Harvard University Press.

Koolen, R., Gatt, A., Goudbeek, M., & Krahmer, E. (2011). Factors causing overspecification in definite descriptions. *Journal of Pragmatics, 43*(13), 3231-3250.

Krahmer, E., & Swerts, M. (2004). More about brows: a cross-linguistic study via analysis-by-synthesis. In C. Pelachaud & Z. Ruttkay (Eds.), *From Brows to*

Trust: Evaluating Embodied Conversational Agents (pp. 191-216): Kluwer Academic Publishers.

Krahmer, E., & Swerts, M. (2007). The effects of visual beats on prosodic prominence: acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language, 57*, 396-414.

Krahmer, E., Swerts, M., Theune, M., & Weegels, M. (2002). The dual of denial: Two uses of disconfirmations in dialogue and their prosodic correlates. *Speech Communication, 36*, 133-145.

Krauss, R. M. (1998). Why do we gesture when we speak? *Current Directions in Psychological Science, 7*, 54-60.

Krauss, R. M., Chen, Y., & Chawla, P. (1996). Nonverbal behavior and nonverbal communication: What do conversational hand gestures tell us? In M. Zanna (Ed.), *Advances in experimental social psychology* (pp. 389-450). Tampa: Academic Press.

Krauss, R. M., Chen, Y., & Gottesman, R. F. (2000). Lexical gestures and lexical access: A process model. In D. McNeill (Ed.), *Language and gesture* (pp. 261-283). New York: Cambridge University Press.

Krauss, R. M., Dushay, R., Chen, Y., & Rauscher, F. (1995). The communicative value of conversational hand gestures. *Journal of Experimental Social Psychology, 31*, 533-552.

Krauss, R. M., & Hadar, U. (1999). The role of speech-related arm/hand gestures in word retrieval. In R. Campbell & L. Messing (Eds.), *Gesture, speech, and sign* (pp. 93-116). Oxford: Oxford University Press.

Krauss, R. M., & Weinheimer, S. (1966). Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology, 4*, 343-346.

Kuhlen, A., & Brennan, S. (2010). Anticipating distracted addressees: How speakers' expectations and addressees' feedback influence storytelling. *Discourse Processes, 47*(7), 567- 587.

Kuhlen, A., & Brennan, S. (2013). Language in dialogue: When confederates might be hazardous to your data. *Psychonomic Bulletin & Review, 20*(1), 54-72.

Lam, T. Q., & Watson, D. G. (2010). Repetition is easy: Why repeated referents have reduced prominence. *Memory and Cognition, 38*(8), 1137-1146.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(159-174).

Leuninger, H., Hohenberger, A., Waleschkowski, E., Menges, E., & Happ, D. (2004). The impact of modality on language production: Evidence from slips of the tongue and hand. In T. Pechmann & C. Habel (Eds.), *Multidisciplinary approaches to language production* (pp. 219-277). Berlin, New York, Amsterdam: Mouton de Gruyter.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge: MIT Press.

Levelt, W. J. M., Richardson, G., & La Heij, W. (1985). Pointing and voicing in deictic expressions. *Journal of Memory and Language, 24*, 133-164.

Levy, E., & McNeill, D. (1992). Speech, gesture and discourse. *Discourse Processes, 15*, 277-301.

Liddell, S. K. (2003). *Grammar, gesture, and meaning in American Sign Language*: Cambridge University Press.

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

Lieberman, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech, 6*(3), 172-187.

Lindblom, B. (1990). Explaning variation: a sketch of the H and H theory. In W. Hardcastle & A. Marchal (Eds.), *Speech production and speech modelling* (pp. 403-439). Dordrecht: Kluwer Academic Publishers.

Liszkowski, U. (2005). Human twelve-month-olds point cooperatively to share interest with and helpfully provide information for a communicative partner. *Gesture, 5*(1-2), 135-154.

Litman, D., Swerts, M., & Hirschberg, J. (2006). Characterizing and predicting corrections in spoken dialogue systems. *Computational Linguistics, 32*, 417-438.

Lombard, E. (1911). Le signe de l'elevation de la voix. *Ann. Maladies Oreille Larynx, Nez. Pharynx 37*, 101-119.

Masson-Carro, I., Goudbeek, M., & Krahmer, E. (2014). On the automaticity of reduction in dialogue: Cognitive load and repeated multimodal references. In P. Bello, M. Guarini, M. McShane & B. Scassellati (Eds.), *Proceedings of the*

36th Annual Meeting of the Cognitive Science Society. Quebec City: Cognitive Science Society.

Matthews, D., Butcher, J., Lieven, E., & Tomasello, M. (2012). Two- and four-year-olds learn to adapt referring expressions to context: Effects of distracters and feedback on referential communication. *Topics in Cognitive Science, 4*(2), 184-210.

Mauk, C. E., Lindblom, B., & Meier, R. P. (2008). Undershoot of ASL locations in fast signing In J. Quer (Ed.), *Signs of the Time. Selected Papers from TISLR 8* (pp. 3-24). Seedorf: Signum.

Mayberry, R. I., & Jaques, J. (2000). Gesture production during stuttered speech: Insights into the nature of gesture-speech integration. In D. McNeill (Ed.), *Language and Gesture* (pp. 199-213). Cambridge: Cambridge University Press.

McClave, E. (1998). Pitch and manual gestures. *Journal of Psycholinguistic Research, 27*(1), 69-89.

McNeill, D. (1985). So you think gestures are nonverbal? *Psychological Review, 92*, 350-371.

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

McNeill, D. (2005). *Gesture and thought*. Chicago: The University of Chicago Press.

McNeill, D. (2006). Gesture and communication. In K. Brown (Ed.), *The Encyclopedia of Language and Linguistics* (pp. 58-66): Elsevier.

McNeill, D., & Duncan, S. (2000). Growth points in thinking-for-speaking. In D. McNeill (Ed.), *Language and gesture* (pp. 141-161). Cambridge: Cambridge University Press.

Melinger, A., & Kita, S. (2007). Conceptualisation load triggers gesture production. *Language and Cognitive Processes, 22*(4), 473-500.

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

Mol, L. (2011). *Language in the hands*. PhD dissertation: Tilburg University.

Mol, L., Krahmer, E., Maes, A., & Swerts, M. (2009). The communicative import of gestures. Evidence from a comparative analysis of human–human and human–machine interactions. *Gesture, 9*(1), 97-126.

Mol, L., Krahmer, E., Maes, A., & Swerts, M. (2012). Adaptation in gesture: Converging hands or converging minds? *Journal of Memory and Language, 66*, 249–264.

Mol, L., Krahmer, E., & van de Sandt-Koenderman, M. (2013). Gesturing by speakers with aphasia: How does it compare? *Journal of Speech, Language, and Hearing Research, 56*, 1224-1236.

Morsella, E., & Krauss, R. M. (2005). Muscular activity in the arm during lexical retrieval: implications for gesture-speech theories. *Journal of Psycholinguistic Research, 34*(4), 415-427.

Müller, C. (1998). *Redebegleitende Gesten. Kulturgeschichte – Theorie – Sprachvergleich*. Berlin: Berlin Verlag.

Olson, D. R. (1970). Language and thought: Aspects of a cognitive theory of semantics. *Psychological Review, 77*, 257-273.

Oviatt, S., MacEachern, M., & Levow, G.-A. (1998). Predicting hyperarticulate speech during human-computer error resolution. *Speech Communication, 24*, 87-110.

Özyürek, A. (2002). Do speakers design their co-speech gestures for their addressees? The effects of addressee location on representational gestures. *Journal of Memory and Language, 46*, 688–704.

Pine, K., Bird, H., & Kirk, E. (2007). The effects of prohibiting gestures on children's lexical retrieval ability. *Developmental Science, 10*(6), 747-754.

Rauscher, F. H., Krauss, R. M., & Chen, Y. (1996). Gesture, speech and lexical access: The role of lexical movements in speech production. *Psychological Science, 7*, 226-230.

Rimé, B., Schiaratura, L., Hupet, M., & Ghysselinckx, A. (1984). Effects of relative immobilization on the speaker's nonverbal behavior and on the dialogue imagery level. *Motivation and Emotion, 8*(4), 311-325.

Russell, K., Wilkinson, E., & Janzen, T. (2011). ASL sign lowering as undershoot: A corpus study. *Laboratory Phonology, 2*(2), 403-422.

Samuel, A. G., & Troicki, M. (1998). Articulation quality is inversely related to redundancy when children or adults have verbal control. *Journal of Memory and Language, 39*(2), 175-194.

Sandler, W. (1989). *Phonological Representation of the Sign: Linearity and Nonlinearity in American Sign Language*. Dordrecht: Foris.

Sandler, W., & Lillo-Martin, D. (2006). *Sign Language and Linguistic Universals*. Cambridge: Cambridge University Press.

Schembri, A., McKee, D., McKee, R., Pivac, S., Johnston, T., & Goswell, D. (2009). Phonological variation and change in Australian and New Zealand Sign languages: The location variable. *Language variation and change, 21*, 193-231.

Shannon, C. (1948). A mathematical theory of communications. *Bell systems technical journal, 27*(4), 623-656.

Shimojima, A., Katagiri, Y., Koiso, H., & Swerts, M. (2002). Informational and dialogue-coordinating functions of prosodic features of Japanese echoic responses. *Speech Communication, 36*, 113-132.

Singer, M. A., & Goldin-Meadow, S. (2005). Children learn when their teachers' gestures and speech differ. *Psychological Science, 16*, 85-89.

Singleton, J. L., Morford, J. P., & Goldin-Meadow, S. (1993). Once is not enough: Standards of well-formedness in manual communication created over three different timespans. *Language, 69*, 683-715.

So, W. C., Kita, S., & Goldin-Meadow, S. (2009). Using the hands to identify who does what to whom: gesture and speech go hand-in-hand. *Cognitive Science, 33*, 115-125.

Stivers, T., & Enfield, N. (2010). A coding scheme for question–response sequences in conversation. *Journal of Pragmatics, 42*, 2620–2626.

Stokoe, W. C. (1960). Sign language structure: An outline of the visual communication systems of the American deaf. *Studies in Linguistics: Occasional papers, 8*.

Streeck, J. (1993). Gesture as communication I: its coordination with gaze and speech. *Communication Monographs, 60*(4), 275-299.

Streeck, J. (1994). Gesture as communication II: the audience as co-author. *Research on Language and Social Interaction, 27*(3), 239-267.

Supalla, T. (2008). Sign language archeology: Integrating historical linguistics with fieldwork on young sign languages. In R. M. de Quadros (Ed.), *Sign Languages: Spinning and unraveling the past, present and future. Proceedings of the The Ninth International Conference on Theoretical Issues in Sign Language Research.* (pp. 575-583). Florianopolis, Brazil: Editora Arara Azul. Petrópolis/RJ. Brazil.

Traum, D. R. (1994). *A computational theory of grounding in natural language conversation.* Unpublished PhD dissertation, University of Rochester.

Tyrone, M. E., & Mauk, C. E. (2010). Sign lowering and phonetic reduction in American Sign Language. *Journal of Phonetics, 38*, 317-328.

Vajrabhaya, P., & Pederson, E. (2013). *Repetition vs. listener accommodation: A case study of co-speech gesture in retellings.* Paper presented at the New Ways of Analyzing Variation 42 Conference, Pittsburgh.

Van Deemter, K., Gatt, A., van der Sluis, I., & Power, R. (2012). Generation of referring expressions: Assessing the Incremental Algorithm. *Cognitive Science, 36*(5), 799-836.

van der Hulst, H. (1993). Units in the analysis of signs. *Phonology, 10*(2), 209-241.

Van der Sluis, I., & Krahmer, E. (2007). Generating multimodal referring expressions. *Discourse Processes, 44*(3), 145-174.

Wagner, P., Malisz, Z., & Kopp, S. (2014). Gesture and speech in interaction: An overview. *Speech Communication, 57*, 209-232.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). *ELAN: a Professional Framework for Multimodality Research.* Paper presented at the LREC 2006, Fifth International Conference on Language Resources and Evaluation, Genoa, Italy.

Zipf, G. K. (1936). *The psychobiology of language*. London: Routledge.

Acknowledgments

For me, writing this thesis was like running a marathon. As is usually the case in a marathon, it takes a while to run all those miles, and, like in a marathon, some of the miles were nice and easy, but others were more difficult, and near the end there was an unexpected visit from the man with the hammer. But being able to write these acknowledgments means that the finish line is almost there! Oh, if only you knew how many times I went running and spent most of my run, crossing the river Waal, or dodging deer in the Nijmegen forest, thinking about what I would write in this acknowledgments section, who I would thank, and in what way exactly. It brings me so much pleasure to be able to finally put those thoughts down on paper and to thank those people who helped me along the way in my personal race towards getting my PhD degree.

Marc and Emiel, I couldn't have wished for better supervisors, who complement each other in so many ways. It might be that more PhD students say that they wouldn't have made it to the end without their supervisors' help, but in my case it really is true, you two helped me so much, especially in the last few years, and this thesis would truly not exist in its current shape if it weren't for your constant encouragement and optimism. Marc, thank you for always staying positive, for your silly jokes, but also for being the one to often give that one insightful comment which puts everything in (a new) perspective. Emiel, thank you for your pragmatism, your quick and detailed reading and comments on my texts, for letting me know more than once that your door is always open, and of course, your tips on which English novels to read.

I would also like to thank the members of my thesis committee, Sonia Frota, Albert Gatt, David House, Stefan Kopp, and Fons Maes, for taking the time to read my work, for commenting on it in detail, for travelling to Tilburg for the defense, and most of all for approving this dissertation! I am very grateful.

This dissertation was written in the context of the Vici project "Bridging the gap between computational linguistics and psycholinguistics: The case of referring expressions." Over the course of about 5 years, we held regular Vici meetings, which were always fun, generally involved cookies or chocolates, and (sometimes) also a serious discussion of our work. Thank you fellow Vici members for making those years so enjoyable and fruitful! Thank you Emiel for being the cornerstone of the project, and thank you Marc for making sure every meeting was not only useful, but also contained at least one bad joke. Thank you Jette for doubling the number of female members, and thank you Albert; the Maltesers we often had during the meetings have taken on a new meaning.

In particular I want to thank the project members Martijn Goudbeek and Ruud Koolen, who were there with me from the start and were (and are) some of my favourite colleagues. I especially enjoyed the time when all three of us were living in Nijmegen and we so often took the train to Tilburg together (sometimes it would've made more sense to hold the Vici meetings in there). Martijn, the project would not have been the same without your sense of humour (it wasn't always clear whether you were Marc's sidekick or the other way around). Thank you also for always being available for any (statistics) questions, and, more recently, it has been a pleasure teaching together with you. Ruud, thank you for being the perfect roommate, for all those years! We could spend all day working quietly, but luckily we always found time to catch up and chat, sometimes about our research, but mainly about teaching issues, how to deal with certain students, our (lack of) career plans, and life in general.

Several people who joined the Vici group for shorter or longer periods of time helped me with my research. I want to especially thank Elsa Jonkers and Kristel Bartels for all the hard and precise work they did collecting and annotating data for several chapters reported in this thesis. You were invaluable student assistants! Many thanks also to Joost Driessen, Axelle Schmit and Manon Yassa for helping me by collecting and transcribing data. And thank you Ingrid Masson Carro, for reading and commenting on the introductory chapter of this thesis.

Nick Wood, thank you, not only for perfectly digitising my video recordings, but also for insisting that you should be the one to do this for *all* of my data. You saved me so much time! Thank you also for the lovely dinners we had at your and Leah's place during which academia was put in perspective.

Many, many thanks to Marlous Bervoets for creating the cover of this thesis!

Thank you Judith Holler, Marianne Gullberg, and Jan de Ruiter for your detailed reading of my work and your feedback on the work presented in this thesis. Judith I want to thank also for her useful review of chapter 5, and Jan I want to thank also for his constructive feedback on chapter 3. I would like to thank Marianne also for introducing

me to the field of gesture studies, and for showing me the ropes back when I was working on my MA thesis.

One of the best things about working at Tilburg University was the great atmosphere at the department. There was never any competition between colleagues (or if there was I did not notice it) and my colleagues were (and are!) truly nice people. Thank you all for making my time as a PhD student a pleasurable one, and for being supportive when things were tough. Some colleagues I want to mention in particular. Lauraine, thank you for giving the department such a homely feel, and for all the work you do behind the scenes. Thank you Mariek, Maaike, Janneke, Monique, Lieke and Marije, for being great co-teachers, for enjoyable train rides (Lieke, Maaike and Monique), and for your interest in how things were going thesis-wise (especially Janneke and Mariek). Of course, I want to thank the 'original' PhDinner group (although none of you are a PhD student anymore, and many of you have moved elsewhere): Ruud, Lisette, Martijn Balsters, Martijn Goudbeek, Suleman, Constantijn, Jorrig, Jette, Mandy and Lisanne, thank you for being a great bunch of people, and for the lovely food and drinks we had at your (and my) place. Lisette, Constantijn, Jorrig, Ruud, Martijn Goudbeek, Jette, Phoebe, Sterling, Ingrid, Karin, Yan, Hans, Mandy (sorry if I forgot anyone), thank you for travelling across the world with me to various conferences and for the fun we had in places like Norwich (never been on such a short plane flight in my life), Boston (never been in such a sweltering place in my life), Sapporo (never felt so out of place in my life), and Quebec (never had such an 'inexplicable' headache in my life).

Some people deserve to be mentioned more than once (and as things get more personal, you will find that I am switching to Dutch here). Mandy Visser en Lisanne van Weelden, dames, waar te beginnen? Ten eerste vind ik het natuurlijk fantastisch (en vanzelfsprekend) dat juist jullie mijn paranimfen zijn. Het is een hele geruststelling dat als ik dan toch van mijn stokje ga tijdens de verdediging (en jullie weten dat de kans hierop reëel is), jullie klaar staan om het van me over te nemen. Maar dat is slechts een klein deel van mijn dankbaarheid naar jullie. In de loop van de jaren zijn jullie voor mij veel meer geworden dan een paar leuke collega's. Bij jullie kan ik altijd mijn eieren kwijt, even tussendoor tijdens het werk, maar zeker ook tijdens een van onze borrels, etentjes, of logeerpartijen (op allerlei exotische locaties zoals Malle, Sapporo, Quebec en niet te vergeten Aalten). Jullie zijn echt een enorme steun geweest; met name de laatste paar jaar toen het allemaal niet altijd even makkelijk was, was het heel fijn om op het werk een paar mensen te hebben rondlopen die precies weten hoe de vork in de steel zit. En wat zijn jullie toch ook een gouden combinatie: Lisanne, die in staat is om maanden uit 1 tas te leven, en door niets geraakt lijkt te worden (maar ondertussen een klein hartje heeft), en Mandy, die voor een weekje weg al meer schoenen meeneemt dan dat er dagen in de week zijn, en door alles geraakt lijkt te worden (maar ondertussen sterker is dan ze zelf vermoed). De tijd van elkaars collega zijn in Tilburg is dan weliswaar voorbij, maar ik heb er wel een stel supervriendinnen bij! Dank voor alles lieve schatten!

Strict genomen hebben mijn vrienden misschien niet veel bijgedragen aan dit proefschrift. Toch wil ik een aantal bij dezen laten weten hoezeer ik ze waardeer; in het dagelijkse leven ben ik nou niet echt scheutig met het verwoorden van mijn emoties, dus dit is mijn kans, en ik ga hem grijpen!

Om te beginnen wil ik Imke Vunderink en Lies Wijnterp bedanken. Onze levens zijn, nadat we samen zijn begonnen met studeren, (geografisch) flink uit elkaar gelopen, maar toch zijn we elkaar niet kwijtgeraakt. We kennen elkaar al zo lang en zo goed, wat is het fijn om zulke vriendinnen te hebben! Imke, dank je voor de onvoorwaardelijke vriendschap. De afgelopen jaren zijn er veel dingen in onze levens gebeurd, maar samen kunnen we altijd even alles bespreken (en waar nodig in perspectief plaatsen), of juist niet, met een glaasje wijn of een kop thee (hmmm, de wilde jaren die we hadden toen Lies nog in Nijmegen woonde zijn denk ik toch echt voorbij). Lies, wie had ooit gedacht dat wij daadwerkelijk zouden promoveren, en dan ook nog ongeveer tegelijkertijd. Wat zijn we toch een doorzetters! Tja, wat kan ik verder zeggen, je woont nu gewoon veel te ver van mij vandaan (kom terug!), en wat ben ik een slechte skyper! Ik hoop dat je weet dat mijn skypegedrag in de verste verte geen afspiegeling is van mijn waardering van onze vriendschap.

Nanna Haug Hilton, je bent op veel vlakken een voorbeeld voor me, van het aan de ene kant moeiteloos hip en modieus zijn (is er dan niets wat jou niet staat?) tot het aan de andere kant wars zijn van uiterlijk vertoon (je haren wassen in de rivier met ter plekke geleende shampoo en daarna je net gevangen vis doodslaan op een steen hebben een onvergetelijke indruk gemaakt). Na ons afstuderen ben je naast een vriendin ook een voorbeeld-collega geworden (die het nog ver gaat schoppen). Dank je daarvoor! Ik kijk uit naar nog heel veel feestjes, winkel- en eetafspraken met jou in het hoge noorden (of dat nou Zwolle, Groningen of Oslo zal zijn).

Sara Vermeulen, wat een geluk dat ik jou al mijn hele leven ken! Het feit dat wij vanaf dat we vier of vijf waren al samen kattekwaad uithaalden in Slijk-Ewijk heeft natuurlijk niet heel veel bijgedragen aan dit proefschrift. Toch wil ik je bij dezen bedanken voor onze bijzondere vriendschap. Want ook al kunnen onze wegen zich soms tijdelijk scheiden, ze komen altijd weer samen. Onze nonstop gesprekken tijdens de vele wandelingen de laatste jaren (hebben we uberhaupt wel gezien waar we liepen?) over werk, proefschriften, het leven en de liefde, zijn me heel dierbaar.

Dankzij mijn (schoon/stief)familie ben ik blijven staan, vooral tijdens het afronden van dit proefschrift, zonder dat ik ooit het gevoel heb gehad dat het erg zou zijn geweest als ik toch was gevallen. Ik ben ze allemaal heel dankbaar voor deze onvoorwaardelijke steun. Sommigen wil ik in het bijzonder bedanken.

Gijs Hoetjes en Lieke de Jong, dank jullie wel voor een tweede thuis. Gijs, altijd als ik twijfels uitte over wat dan nu toch te doen met mijn werkzame leven gaf je als tip ‘schrijf een boek’. Nou, daar is ie dan. Hij leest wat minder lekker weg dan jouw boeken, vrees ik, maar toch. Ik hoop dat je hem mooi vind. Ik denk dat jij en ik allebei beter zijn in het schrijven dan in het spreken over onze emoties. Daarom bij deze op schrift: ik ben trots dat je mijn vader bent! Lieke, dank je wel voor je hartelijkheid, gastvrijheid en voor je luisterend oor. ‘Stiefmoeder’ heeft bij mij alleen maar positieve associaties!

Elisabeth Spaan en Veronica Vasterling, dank jullie wel voor een derde thuis (het worden er steeds meer). Els, je bent de liefste moeder. Ik weet dat wat ik ook doe, je me altijd zult steunen en dat ik altijd bij je terecht kan. Dat is een heel fijn gevoel, dank je! Veronica, dank je voor je nuchtere kijk op de academische wereld, die kwam goed van pas!

Geisje, hoe ouder we worden, hoe kleiner ons leeftijdsverschil lijkt te zijn en hoe meer we aan elkaar hebben; ik zou ons vanzelfsprekende contact voor geen goud willen missen! Dank je wel dat ik altijd alles bij je kwijt kan, vaak ook nog eens tijdens mooie wandelingen of ontspannen sauna bezoekjes.

Bas, bij jou ben ik thuis. Samen beklimmen we de hoogste bergen en kruipen we uit de diepste dalen. Met jou in mijn leven is dit proefschrift maar bijzaak. En ja, er was vast een proefschrift geweest als wij elkaar niet hadden gekend, maar dan hadden in ieder geval de stimulusfilmpjes in hoofdstuk 2 er veel minder mooi uitgezien! Thank you for tying the knot!

Publication list

Journal publications

Hoetjes, M., Krahmer, E. & Swerts, M. (2015). On what happens in gesture when communication is unsuccessful. *Speech Communication, 72*, 160-175.

Hoetjes, M., Koolen, R., Goudbeek, M., Krahmer, E. & Swerts, M. (2015). Reduction in gesture during the production of repeated references. *Journal of Memory and Language, 79-80*, 1-17.

Hoetjes, M., Krahmer, E. & Swerts, M. (2014). Do repeated references result in sign reduction? *Sign Language & Linguistics, 17* (1), 56-81.

Hoetjes, M., Krahmer, E. & Swerts, M. (2014). Does our speech change when we cannot gesture? *Speech Communication*, *57*, 257-267.

Hoetjes, M. (2008). The use of gestures in placement events. *Leiden Working Papers in Linguistics, 5* (1), 24-36.

Conference proceedings

Gu, Y, Mol, L., Hoetjes, M. & Swerts, M. (2014). Does language shape the production and perception of gestures? A study on late Chinese-English bilinguals' conceptions about time. Bello, P., Guarini, M., McShane, M. & Scassellati, B. (eds.). *Proceedings of the 36th Annual Conference of the Cognitive Science Society, Quebec, Canada*. Austin TX: Cognitive Science Society, 547-552.

Gu, Y., Mol, L., Hoetjes, M. & Swerts, M. (2013). What can Chinese speakers' temporal gestures reveal about their thinking about time? *Proceedings of TiGeR 2013: The combined meeting of the 10th international Gesture Workshop (GW) and the 3rd Gesture and Speech in Interaction (GESPIN) conference.* tiger.uvt.nl

Hoetjes, M., Krahmer, E. & Swerts, M. (2014) On what happens in gesture when communication is unsuccessful. Bello, P., Guarini, M., McShane, M. & Scassellati, B. (eds.). *Proceedings of the 36th Annual Conference of the Cognitive Science Society, Quebec, Canada*. Austin TX: Cognitive Science Society, 2369-2374.

Hoetjes, M., Krahmer, E. & Swerts, M. (2012) Do repeated references result in sign reduction? Miyake, N., Peebles, D. & Cooper, R. P. (eds.). *Proceedings of the 34th Annual Conference of the Cognitive Science Society, Sapporo, Japan.* Austin TX: Cognitive Science Society, 461-466.

Hoetjes, M., Koolen, R., Goudbeek, M., Krahmer, E. & Swerts, M. (2011). GREEBLES Greeble greeb. On reduction in speech and gesture in repeated references. Carlson, L., Hoelscher, C. & Shipley, T. F. (eds.). *Proceedings of the 33rd Annual Conference of the Cognitive Science Society, Boston, United States.* Austin, TX: Cognitive Science Society, 3250-3255.

Hoetjes, M., Schmit, A., Koolen, R. , Goudbeek, M. , Krahmer, E. & Swerts, M. (2011). Reduction in repeated references in gesture and sign language. Efthomiou, E. & Kouroupetroglou, G. (eds.). *Proceedings of the 9th International Gesture Workshop.* Athens, Greece, 84-97.

Hoetjes, M., Krahmer, E. & Swerts, M. (2010). Can we hear a gesture? The influence of gestures on speech*. Proceedings of the 12th international conference on Laboratory Phonology (LabPhon).* Albuquerque: University of New Mexico, 175-176.

Hoetjes, M., Krahmer, E. & Swerts, M. (2009). Untying the knot between gestures and speech. Theobald, B. J. & Harvey, R. (Eds.), *Proceedings of the 2009 Conference on Audio Visual Speech Processing (AVSP.)* Norwich, UK: School of Computing Sciences, 90-95.

Abstracts of (conference) presentations

Hoetjes, M. , Koolen, R. (2011). *Greebles, greeble, greeb. Reduced speech and gesture in repeated references.* Talk presented at the Optimal Communication Colloquium, Radboud University Nijmegen, 12 January 2011.

Hoetjes, M., Krahmer, E., & Swerts, M. (2010). *My hands are tied: the influence of gestures on speech.* Talk presented at the 4th conference of the International Society for Gesture Studies (ISGS). July 2010, Europa-Universität Viadrina, Frankfurt am Oder, Germany.

Hoetjes, M., Krahmer, E., & Swerts, M. (2009). *Untying the knot between gesture and speech. The influence of gestures on speech.* Talk presented at the Nijmegen Gesture Centre, Max Planck Institute for Psycholinguistics, 28 August 2009.

Hoetjes, M., Krahmer, E., & Swerts, M. (2009). *It's a tie. The influence of gestures and their visibility on speech.* Talk presented at the Tabu dagen, Rijksuniversiteit Groningen, 11-12 June 2009.

TiCC PhD series

1. Pashiera Barkhuysen. *Audiovisual Prosody in Interaction*. Promotores: M.G.J. Swerts, E.J. Krahmer. Tilburg, 3 October 2008.

2. Ben Torben-Nielsen. *Dendritic Morphology: Function Shapes Structure*. Promotores: H.J. van den Herik, E.O. Postma. Co-promotor: K.P. Tuyls. Tilburg, 3 December 2008.

3. Hans Stol. *A Framework for Evidence-based Policy Making Using IT*. Promotor: H.J. van den Herik. Tilburg, 21 January 2009.

4. Jeroen Geertzen. *Dialogue Act Recognition and Prediction*. Promotor: H. Bunt. Co-promotor: J.M.B. Terken. Tilburg, 11 February 2009.

5. Sander Canisius. *Structured Prediction for Natural Language Processing*. Promotores: A.P.J. van den Bosch, W. Daelemans. Tilburg, 13 February 2009.

6. Fritz Reul. *New Architectures in Computer Chess*. Promotor: H.J. van den Herik. Co-promotor: J.W.H.M. Uiterwijk. Tilburg, 17 June 2009.

7. Laurens van der Maaten. *Feature Extraction from Visual Data*. Promotores: E.O. Postma, H.J. van den Herik. Co-promotor: A.G. Lange. Tilburg, 23 June 2009 (cum laude).

8. Stephan Raaijmakers. *Multinomial Language Learning*. Promotores: W. Daelemans, A.P.J. van den Bosch. Tilburg, 1 December 2009.

9. Igor Berezhnoy. *Digital Analysis of Paintings*. Promotores: E.O. Postma, H.J. van den Herik. Tilburg, 7 December 2009.

10. Toine Bogers. *Recommender Systems for Social Bookmarking*. Promotor: A.P.J. van den Bosch. Tilburg, 8 December 2009.

11. Sander Bakkes. *Rapid Adaptation of Video Game AI*. Promotor: H.J. van den Herik. Co-promotor: P. Spronck. Tilburg, 3 March 2010.

12. Maria Mos. *Complex Lexical Items*. Promotor: A.P.J. van den Bosch. Co-promotores: A. Vermeer, A. Backus. Tilburg, 12 May 2010 (in collaboration with the Department of Language and Culture Studies).

13. Marieke van Erp. *Accessing Natural History. Discoveries in Data Cleaning, Structuring, and Retrieval*. Promotor: A.P.J. van den Bosch. Co-promotor: P.K. Lendvai. Tilburg, 30 June 2010.

14. Edwin Commandeur. *Implicit Causality and Implicit Consequentiality in Language Comprehension*. Promotores: L.G.M. Noordman, W. Vonk. Co-promotor: R. Cozijn. Tilburg, 30 June 2010.

15. Bart Bogaert. *Cloud Content Contention*. Promotores: H.J. van den Herik, E.O. Postma. Tilburg, 30 March 2011.

16. Xiaoyu Mao. *Airport under Control*. Promotores: H.J. van den Herik, E.O. Postma. Co-promotores: N. Roos, A. Salden. Tilburg, 25 May 2011.

17. Olga Petukhova. *Multidimensional Dialogue Modelling*. Promotor: H. Bunt. Tilburg, 1 September 2011.

18. Lisette Mol. *Language in the Hands.* Promotores: E.J. Krahmer, A.A. Maes, M.G.J. Swerts. Tilburg, 7 November 2011 (cum laude).

19. Herman Stehouwer. *Statistical Language Models for Alternative Sequence Selection*. Promotores: A.P.J. van den Bosch, H.J. van den Herik. Co-promotor: M.M. van Zaanen. Tilburg, 7 December 2011.

20. Terry Kakeeto-Aelen. *Relationship Marketing for SMEs in Uganda*. Promotores: J. Chr. van Dalen, H.J. van den Herik. Co-promotor: B.A. Van de Walle. Tilburg, 1 February 2012.

21. Suleman Shahid. *Fun & Face: Exploring Non-Verbal Expressions of Emotion During Playful Interactions*. Promotores: E.J. Krahmer, M.G.J. Swerts. Tilburg, 25 May 2012.

22. Thijs Vis. *Intelligence, Politie en Veiligheidsdienst: Verenigbare Grootheden?* Promotores: T.A. de Roos, H.J. van den Herik, A.C.M. Spapens. Tilburg, 6 June 2012 (in collaboration with the Tilburg School of Law).

23. Nancy Pascall. *Engendering Technology Empowering Women*. Promotores: H.J. van den Herik, M. Diocaretz. Tilburg, 19 November 2012.

24. Agus Gunawan. *Information Access for SMEs in Indonesia*. Promotor: H.J. van den Herik. Co-promotores: M. Wahdan, B.A. Van de Walle. Tilburg, 19 December 2012.

25. Giel van Lankveld. *Quantifying Individual Player Differences*. Promotores: H.J. van den Herik, A.R. Arntz. Co-promotor: P. Spronck. Tilburg, 27 February 2013.

26. Sander Wubben. *Text-to-text Generation Using Monolingual Machine Translation*. Promotores: E.J. Krahmer, A.P.J. van den Bosch, H. Bunt. Tilburg, 5 June 2013.

27. Jeroen Janssens. *Outlier Selection and One-Class Classification*. Promotores: E.O. Postma, H.J. van den Herik. Tilburg, 11 June 2013.

28. Martijn Balsters. *Expression and Perception of Emotions: The Case of Depression, Sadness and Fear.* Promotores: E.J. Krahmer, M.G.J. Swerts, A.J.J.M. Vingerhoets. Tilburg, 25 June 2013.

29. Lisanne van Weelden. *Metaphor in Good Shape*. Promotor: A.A. Maes. Co-promotor: J. Schilperoord. Tilburg, 28 June 2013.

30. Ruud Koolen. *Need I say More? On Overspecification in Definite Reference*. Promotores: E.J. Krahmer, M.G.J. Swerts. Tilburg, 20 September 2013.

31. J. Douglas Mastin. *Exploring Infant Engagement, Language Socialization and Vocabulary Development: A Study of Rural and Urban Communities in Mozambique*. Promotor: A.A. Maes. Co-promotor: Dr. P.A. Vogt. Tilburg, 11 October 2013.

32. Philip C. Jackson, Jr. *Toward Human-Level Artificial Intelligence: Representation and Computation of Meaning in Natural Language*. Promotores: H.C. Bunt, W.P.M. Daelemans. Tilburg, 22 April 2014.

33. Jorrig Vogels. *Referential Choices in Language Production: The Role of Accessibility.* Promotores: A.A. Maes, E.J. Krahmer. Tilburg, 23 April 2014 (cum laude).

34. Peter de Kock. *Anticipating Criminal Behaviour.* Promotores: H.J. van den Herik, J.C. Scholtes. Co-promotor: P. Spronck. Tilburg, 10 September 2014.

35. Constantijn Kaland. *Prosodic marking of semantic contrasts: do speakers adapt to addressees?* Promotores: M.G.J. Swerts, E.J. Krahmer. Tilburg, 1 October 2014.

36. Jasmina Marić. *Web Communities, Immigration and Social Capital.* Promotor: H.J. van den Herik. Co-promotores: R. Cozijn, M. Spotti. Tilburg, 18 November 2014.

37. Pauline Meesters. *Intelligent Blauw*. Promotores: H.J. van den Herik, T.A. de Roos. Tilburg, 1 December 2014.

38. Mandy Visser. *Better use your head. How people learn to signal emotions in social contexts.* Promotores: M.G.J. Swerts, E.J. Krahmer. Tilburg, 10 June 2015.

39. Sterling Hutchinson. *How symbolic and embodied representations work in concert.* Promotores: M.M. Louwerse, E.O. Postma. Tilburg, 30 June 2015.

40. Marieke Hoetjes. *Talking hands. Reference in speech, gesture, and sign.* Promotores: E.J. Krahmer, M.G.J. Swerts. Tilburg, 7 October 2015.