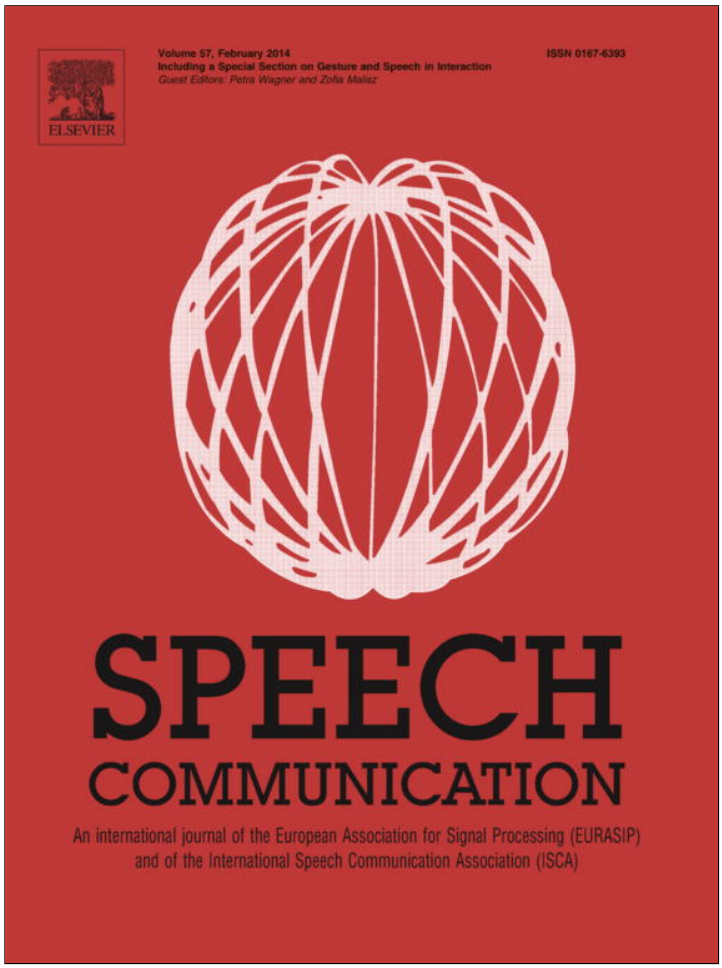

Available online at www.sciencedirect.com

ScienceDirect

Speech Communication 57 (2014) 257–267

SPEECH COMMUNICATION

www.elsevier.com/locate/specom

Does our speech change when we cannot gesture?

Marieke Hoetjes *, Emiel Krahmer, Marc Swerts

Tilburg Center for Cognition and Communication (TiCC), School of Humanities, Tilburg University, The Netherlands

Available online 14 June 2013

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; half of the participants had to perform 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 for 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.

Keywords: Gesture; Speech; Gesture prevention; Director–matcher task

1. 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 (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 et al. (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.

1.1. 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. In a recent study, Hostetter et al. (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

* Corresponding author. Address: Tilburg University, Room D404, PO Box 90153, 5000 LE, Tilburg, The Netherlands. Tel.: +31 13 466 2918.

E-mail addresses: m.w.hoetjes@tilburguniversity.edu (M. Hoetjes), e.j.krahmer@tilburguniversity.edu (E. Krahmer), m.g.j.swerts@tilburguniversity.edu (M. Swerts).

0167-6393/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.specom.2013.06.007

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 et al., 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 and 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 and Hadar, 2001; Rauscher et al., 1996), and not being able to gesture has been shown to increase disfluencies (Finlayson et al., 2003). In a study by Rauscher et al. (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 et al. (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.

1.2. Influence of (not being able to) gesture on acoustics

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.

1.3. 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 hypothesize 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 and Casey, 2001; Gullberg, 2006; Mol et al., 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. 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 and Hadar, 2001; Rauscher et al., 1996) and are thus mainly 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 largely produced for the addressee and thus mainly 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.

1.4. 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 and Casey, 2001) or used an artificial setting (e.g. Krahmer and Swerts, 2007). This means that many aspects of the direct influence of gestures on speech remain unknown.

1.5. 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 were free to move their hands compared to a control condition in which they had to sit on their hands. Two other aspects of cognitive load, mutual visibility and topic complexity (expressed in the number of attempts), were also taken into account.

We conducted 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, for example by people using more filled pauses when they cannot gesture compared to when they can gesture.

In addition to the production experiment we conducted a perception experiment, where participants were presented with pairs of sound fragments from the production experiment and were asked to choose in which sound fragment the speaker was 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.

2. Method

2.1. Production experiment

2.1.1. 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.

2.1.2. Stimuli

Directors watched different 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 separate video clip, containing instructions for a different tie knot, was cut into six fragments. Each fragment contained a short (maximally 10 or 15 s) 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 Fig. 1.

2.1.3. 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

Fig. 1. Still of the beginning of a fragment of one of the stimulus clips, in this case accompanied by the phrases 'behind' and 'up'.

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 control for (lack of) mutual visibility. Examples of the experimental setup can be seen in Fig. 2.

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 Fig. 2. The audio recorder was placed on the table, to the right of the director, as can be seen in Fig. 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 min.

2.1.4. Design

The experiment had a mixed design ($2 \times 2 \times 3$), 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 and 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

Fig. 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.

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.

2.1.5. 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 et al., 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. With 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. The number of attempts was not taken into account as a factor for the number of gestures since only the gestures produced during the first description were analyzed.

Secondly, we analyzed 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 $\times$ 3 attempts $\times$ 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, depending on the ability to gesture.

The use of filled pauses in the director's speech was also analyzed. On the basis of previous literature 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 (see Section 2.2.2 below), which was also used for the perception experiment (as described below in Section 2.2). The sound pair recordings were analyzed using Praat software (Boersma and Weenink, 2010). The minimum and maximum pitch, the mean pitch and pitch range and the mean intensity of each sound fragment were analyzed. These aspects were taken into account because previous research has suggested that speech becomes more monotonous when speakers cannot gesture. For the acoustic analyses we only 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 analyzed 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 analyzed 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 analyzed whether there was an effect of the ability to gesture on the acoustic measures. Unless noted otherwise, all tests for significance were conducted with repeated measures ANOVA. All significant main effects and interactions will be discussed.

2.2. Perception experiment

To see whether the claim, that speech becomes more monotonous when people cannot gesture, can be supported by the present data set, and, more importantly, 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.

2.2.1. 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).

2.2.2. 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 the 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 strict 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.

2.2.3. Setup

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.

2.2.4. 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.

2.3. Hypotheses

Following previous research (e.g. Alibali et al., 2001; Bavelas et al., 2008; Emmorey and 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 and Gentilucci, 2006; Krahmer and 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.

The perception task on the selected audio recordings is conducted to see whether people can *hear* when somebody is gesturing. Here we expect that a possible change in speech without gesture compared to speech with gesture can be perceived by the listeners.

3. Results

Table 1 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.

3.1. Number of gestures

We found an unsurprising main effect of ability to gesture on the mean number of gestures produced by the director ($F\ (1,\ 36) = 26.8,\ p < .001$), showing that the experimental manipulation worked. There was no effect of mutual visibility on the number of gestures (see Table 2). Noteworthy however (as can be seen in more detail in Table 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 show that there are large individual differences with regard to the number of gestures that participants produce.

3.2. Speech duration in time

The mean speech duration of all fragments was 31 s ($SD = 13.7$). There was no effect of ability to gesture on speech duration in time (see Table 1), nor was there an

Table 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 (sec)	36.2 (*16.1*)	35.1 (*11.9*)	35.6 (*14.0*)
Duration attempt 2 (sec)	29.8 (*12.6*)	30.4 (*14.6*)	30.1 (*13.6*)
Duration attempt 3 (sec)	25.4 (*11.6*)	29.0 (*15.6*)	27.2 (*13.6*)
Duration all attempts (sec)	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 attempts	.025 (*.019*)	.027 (*.019*)	.026 (*.019*)
Max pitch (Hz)	248.5 *(83)*	251.65 *(93.5)*	250 (*88.25*)
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 *(70.85)*
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
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)*

effect of mutual visibility. There was, however, a significant effect of the number of attempts, $F(2, 72) = 23.376$, $p < .001$ (see Table 1), with people getting quicker in instructing a tie knot when they have done so before.

3.3. Speech duration in words

No effects of ability to gesture or mutual visibility on the number of words produced by the director were found (see Table 1). However, there was a significant effect of the number of attempts. Significantly fewer words (for the means, see Table 1) were used for each following attempt, $F(2, 72) = 9.06$, $p < .001$, showing 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.

3.4. 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 1). There were also no interaction effects.

3.5. 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 1) were used to instruct each following attempt, $F(2, 72) = 19.756$, $p < .05$, showing that the rate of filled pauses decreases once people have instructed a tie knot before. There was also an interaction effect between the ability to gesture and the number of attempts on the rate of filled pauses, $F(2, 72) = 3.272$, $p = .044$. For the first attempt the inability to gesture lead to a decrease in the rate of filled pauses, whereas for the second and third attempt the inability to gesture lead to an increase in the rate of filled pauses (see Table 1).

3.6. Acoustic analyses

We found no significant effect of the ability to gesture on any of the dependent acoustic measures (for the means, see Table 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.

3.7. Acoustic perception test

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, $t_{(19)} = 1.843$, *n.s.*).

4. Discussion

In this study, the 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 immobilized 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 where 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, with the two tie knots which had to be instructed three times, 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 Section 1, 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 (exemplified by the number of attempts). 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 and 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 any main effects of the ability to gesture on almost all of our dependent measures, 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. The inability to gesture had a reducing effect on the rate of filled pauses in the first attempt at instructing how to tie a tie, but in the second and third attempt the inability to gesture caused the rate of filled pauses to increase. 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 that, however, that when we look at the descriptives, as given in Table 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 ges-

ture. 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 (large number of) variables that we took into account are all variables that can be considered to be the main variables 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 analyzed 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 mean number of gestures that were produced (as shown in Tables 1 and 2) show that this logic is unlikely. Moreover, when 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 realized 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 et al. (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.

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, preferably with (even) larger samples than the one collected here to further increase statistical power. Various lines of future research naturally suggest themselves, both related to the gestures that were studied and the task. Krahmer and Swerts (2007) found evidence, in a rather controlled setting, of the impact of gestures on speech production, as discussed in the introduction. 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 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 and 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.

Finally, given the inconclusive set of findings described in the introduction, it would be worthwhile to look at the various studies in more detail, and see whether any underlying generalizations have been missed. One reviewer, for instance, suggested that it could be interesting to look at whether previous studies relied on confederates or not, given the recent discussion on which impact confederate addressees might have on results in interactive studies (Bavelas et al., 2008; Kuhlen and Brennan, 2013). We leave this for future research.

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 analyzed 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). Based on the data from the present study, we might tentatively state that topic complexity has a larger effect on fluency than the ability to gesture. However, since topic complexity was not the main focus of the present study we will conclude by saying that, at least on the basis of the present dataset, tying people's hands has not helped to untie the knot between speech and gesture.

Acknowledgements

We would like to thank Bastiaan 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. We received financial support from The Netherlands Organization for Scientific Research, via a Vici grant (NWO grant 27770007), which is gratefully acknowledged. 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.

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.

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.

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.

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.

Cavé, C., Guaïtella, I., Bertrand, R., Santi, S., Harlay, F., Espesser, R. (1996). About the relationship between eyebrow movements and F0 variations. In: 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., Kita, S. (2007). Individual difference in the use of spontaneous gestures in a mental rotation task. In: Integrating Gestures, International Society for Gesture Studies Conference, Evanston, IL, USA.

Clark, H., Krych, M., 2004. Speaking while monitoring addressees for understanding. Journal of Memory and Language 50, 62–81.

Clark, H., Wilkes-Gibbs, D., 1986. Referring as a collaborative process. Cognition 22, 1–39.

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

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

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

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

Finlayson, S., Forrest, V., Lickley, R., & Mackenzie Beck, J. (2003). Effects of the restriction of hand gestures on disfluency. In: Disfluency in Spontaneous Speech (DiSS'03). Retrieved from http://www.isca-speech.org/archive/diss_03/dis3_021.html.

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.

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.

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

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: McNamara, D.S., Trafton, J.G. (Eds.), 29th Annual Meeting of the Cognitive Science Society. Erlbaum, Mawhah, NJ, pp. 1097–1102.

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

Kendon, A., 2004. Gesture. Visible Action as Utterance. Cambridge University Press, Cambridge.

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.

Krahmer, E., Swerts, M., 2004. More about brows: a cross-linguistic study via analysis-by-synthesis. In: Pelachaud, C., Ruttkay, Z. (Eds.), From Brows to Trust: Evaluating Embodied Conversational Agents. Kluwer Academic Publishers, pp. 191–216.

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.

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: Zanna, M. (Ed.), Advances in Experimental Social Psychology. Academic Press, Tampa, pp. 389–450.

Krauss, R.M., Hadar, U., 2001. The role of speech-related arm/hand gestures in word retrieval. In: Campbell, R., Messing, L. (Eds.), Gesture, Speech, and Sign. Oxford University Press, Oxford, pp. 93–116.

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

Levelt, W.J.M., 1989. Speaking: From Intention to Articulation. MIT Press, Cambridge.

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

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

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

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.

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.

Özyürek, A., 2002. Do speakers design their cospeech 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.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., Sloetjes, H. (2006). ELAN: a professional framework for multimodality research. In: LREC 2006, Fifth International Conference on Language Resources and Evaluation.