BACHELOR’S THESIS

Radboud University Nijmegen
Faculty of Arts
Department of Linguistics

Neurophysiological evidence of the gesture enhancement effect on degraded speech in children

Suzanne de Jong
s4627245

Supervisor: Prof. Dr. Asli Özyürek in collaboration with Dr. Kazuki Sekine

Second evaluator: Dr. Joost Rommers

August 3, 2019

Table of contents

Abstract

To communicate, people can use speech but also co-speech gestures. Previous studies have shown that adults integrate speech and gesture during comprehension. Two forms of evidence of gesture-speech integration are bimodal enhancement, which implies that gestures can facilitate speech comprehension, and inverse effectiveness, suggesting this gesture enhancement effect is larger in adverse situations like degraded speech. For children, only behavioral evidence of gesture-speech integration in degraded speech conditions is provided. To provide neural evidence, in the present study 6 and 7-year-old children were presented with short videos of an actor producing a verb. Sometimes the speech was clear and sometimes degraded by noise. In addition, half of the videos contained an iconic gesture that matched the speech. In line with previous behavioral studies, children were better at repeating the verb when the speech was clear and, in degraded speech, when it contained a gesture, showing the gesture enhancement effect in an adverse condition. ERP measures showed that co-speech gestures caused a larger N400 than speech without gestures, suggesting more semantic processing effort. However, an expected effect of gestures on degraded speech processing was not found. This could suggest that children's brains are not yet developed similarly as adults' brains, but since the present study was very small and the expected tendency is visible, it should be replicated on a bigger scale.

1. Introduction

Communication is a multimodal phenomenon. When people try to convey a message, they do not only use speech, but also gestures. Co-speech gestures are hand movements made to support the speech they accompany or to convey additional meaning. Some of those gestures depict their meaning in the way the gestures are formed. This overlap between form and meaning is known as iconicity (Perniss, Thompson & Vigliocco, 2010).

Whether speech and gestures are integrated during processing has been studied a lot (e.g. Beattie & Shovelton, 1999a, 1999b; Broaders & Goldin-Meadow, 2010; Holle & Gunter, 2007; Holler et al., 2014; Kelly, Özyürek, & Maris, 2010; Kelly, Healey, Özyürek, & Holler, 2015; Kelly, Kravitz, & Hopkins, 2004; Özyürek, Willems, Kita, & Hagoort, 2007; Riseborough, 1981; Sekine, Sowden, & Kita, 2015; Stanfield, Williamson, & Özçalişkan, 2014; Wu & Coulson, 2007a). Gesture-speech integration can be described as the unification of information from both the auditory and the visual modality to obtain the speaker's intended message. One form of evidence to prove that speech and gesture are integrated is a phenomenon known as *bimodal enhancement*. Bimodal enhancement means that comprehension of a message is facilitated when the information is offered in two different modalities instead of one (Diederich & Colonius, 2004). This effect is found across different modalities (e.g. Kayser, Petkov, Augath, & Logothetis, 2005; Kelly et al., 2010). In the situation with speech and gestures, this enhancement effect implies that the presence of a visual modality, i.e. co-speech gestures, can positively influence the comprehension of the auditory modality, i.e. speech, resulting in a facilitated understanding of the intended message. Another way in which gesture-speech integration can be proved is with *inverse effectiveness*, which implies that the enhancement effect of one modality increases when comprehension of the other modality gets more difficult (Stein & Meredith, 1993). One example of inverse effectiveness is that the enhancement effect of gesture is stronger in adverse listening conditions.

A lot of research on the processes of gesture-speech integration in adults exists, but whether this works similarly in children is less studied. Especially in adverse listening settings there is a lack of knowledge. Therefore, the present study was conducted.

In the next paragraphs, a short overview is given of the existing research on gesture-speech integration and the reason for conducting the present study. First, it is discussed how speech and gesture comprehension work in adults, and whether these two modalities interact, both behaviorally and neuroscientifically. After a general discussion of gesture-speech integration, the focus will be on this process in the adverse listening condition degraded speech forms. After that, the same aspects will be discussed for children. Finally, the present study will be introduced.

1.1 Speech and gesture comprehension by adults

All healthy hearing adults can understand speech more or less. Also comprehending iconic gestures seems to be a self-evident ability of the average healthy adult. Wu and Coulson (2005), for example, presented adults with cartoon videos followed by a soundless video of a man making an iconic gesture. Participants were asked to judge the relatedness of the meaning of the gestures to the preceding cartoon. A larger N400 was found when the gestures mismatched the cartoons compared to when they matched, suggesting the participants were able to comprehend the meaning of the gestures. The other way around, Wu and Coulson (2007b) first showed participants iconic gestures and then a written probe word. The participants were given the same task as in Wu and Coulson (2005). Again, mismatching gestures and probe words elicited a larger N400 than matching gestures and probe words. This showed a priming effect of iconic gestures on speech comprehension, suggesting that adults can extract the meaning of iconic gestures even when they are not simultaneously accompanied by speech.

1.1.1 Gesture-speech integration in adults

It is now clear that adults understand speech and gestures when they are shown separately. The current view of what happens when they are presented simultaneously is that both modalities are integrated, and thus, that their individual information is unified to form one communicative message. This gesture-speech integration has been studied a lot, both behaviorally and neuroscientifically. Behaviorally, evidence is provided by investigating how gestures affect accuracy scores or reaction times in various tasks (e.g. Beattie & Shovelton, 1999a, 1999b; Holler, Shovelton, & Beattie, 2009; Holler et al., 2014; Kelly et al., 2010; Kelly et al., 2015; Riseborough, 1981).

One approach to study gesture-speech integration is to compare speech only conditions to conditions with speech and gesture. An example of a study using this approach was conducted by Beattie and Shovelton (1999a). In their study some participants were asked to watch cartoons and narrate the content afterwards. The iconic gestures they made during their narrations were filmed and used as stimuli. Other participants were presented with cartoon narrations with either a combination of speech and the iconic gestures, or only speech. After each cartoon they answered questions about different aspects of the cartoon. If gesture and speech are integrated during processing, this would result in a higher accuracy when gestures accompanied the speech to enhance speech comprehension. The findings are in line with this prediction. Participants in the condition with gestures accompanying the speech answered significantly more questions correctly than the participants who only heard speech. Similar results were found in Riseborough (1981).

Another approach to reveal gesture-speech integration was used in Kelly et al. (2015). This approach is based on an assumption that participants who are presented with mismatching speech and gesture would have poorer recognition accuracy than people who are shown matching speech and gesture, assuming gesture-speech integration occurs. To test this, the researchers combined spoken sentences describing everyday actions with iconic gestures that either matched or mismatched the concurrent speech. Before each speech-gesture combination was presented, a written probe word was inserted. Half of the participants had to decide whether the probe word was related to the speech or not, and the other half whether the word was related to the gesture or not. For all participants they found that more errors were made and response times were longer when the gesture mismatched the speech compared to when they matched. This suggests that gesture and speech are integrated during processing.

The previously mentioned studies provide behavioral evidence of gesture-speech integration, but knowledge of the neural processing of gesture and speech is also important to reach full understanding. A good measure to study these neural processes is the N400, a component found in event-related potentials (ERP) measured by means of electroencephalography (EEG) that reflects the effort needed to integrate semantic information. The N400 occurs between 300 and 600 ms after stimulus presentation and the peak is found around 400 ms. The more effort needed for integration, the larger the N400 amplitude (Kutas & Federmeier, 2000, 2011). Regarding speech and gestures, integration of these two modalities could be visualized by N400s in adverse settings. If one sees gestures produced by a speaker while comprehending his or her speech, this listener would show more processing effort when these gestures mismatch the speech than when the two modalities match. In the mismatching condition, this would result in a larger N400 amplitude. Consequently, many studies have used the N400 as a measure of gesture-speech integration (e.g. Habets, Kita, Shao, Özyürek, &

Hagoort, 2011; Holle & Gunter, 2007; Holler et al., 2014; Kelly et al., 2004; Obermeier, Holle, & Gunter, 2011; Özyürek et al., 2007; Wu & Coulson, 2007a).

Wu and Coulson (2007a) studied the neural integration of gesture and speech by presenting participants with videos with speech and a gesture. After each video, a picture probe was shown that was either related to the gesture and the speech, only to the speech, or unrelated to both. During the experiment ERPs were measured. A smaller N400 was found when the picture probe was related to both speech and gesture than when it was only related to the speech and thus mismatched the gesture. This shows that the processing effort was larger when the picture mismatched the gesture, and therefore that gesture is integrated with speech.

Another study using ERP measures was done by Kelly et al. (2004). The participants watched videos in which an actor utters one of four words that correspond to either the glass or the dish on the table in front of him. 'Thin' and 'tall' correspond to the glass, and 'wide' and 'short' to the dish. There were four gesture conditions. One condition was the matching condition, in which the gesture contained the same information as the speech. In the second condition the gesture provided complementary information. For example, the actor uttered the word 'tall' but gestured the thinness of the same object, the thin and tall glass. In the third condition the gesture mismatched the speech. For example, the man said 'tall' but gestured the shortness of the other object, the short dish. The last condition contained no gesture. The results show a larger N400 when the gesture mismatched the speech compared to the matching condition.

A study that also measured neural effects of integration of speech and gestures using matching and mismatching gestures, but in a broader verbal context was done by Özyürek et al. (2007). Participants listened to sentences containing a verb and, simultaneously, a gesture. Both either matched or mismatched the preceding context of the sentence. When the gesture mismatched the rest of the sentence, a larger N400 was found than when they matched. This effect was in all respects similar to the effect of a mismatching verb, suggesting gestures and speech are integrated in the verbal context in a similar way.

Holle and Gunter (2007) studied gesture-speech integration using a disambiguating approach. Participants watched videos with spoken sentences and gestures. The first part of a sentence contained an homonym (e,g. "ball"). This homonym was unbalanced, which means that it had a dominant meaning (e.g. "a round object one can play with"), and a subordinate meaning (e.g. "a formal dance"). Coincident with the homonym, an iconic gesture was made that was related to one of the meanings of the homonym. The second part of the sentence contained a target word which disambiguated the homonym being related to either the dominant or the subordinate meaning of the homonym. This means that some gestures and target words matched and the others mismatched. In one experiment, the participants were asked to judge the congruency between the homonym-gesture combination in the first part of the sentences and the target word in the final part, which means that they had to integrate gesture and speech to complete the task. Another experiment was conducted to check whether participants would still use gestures to disambiguate speech if they did not necessarily have to integrate gesture and speech to do the task. Now they were asked to judge whether they had heard a certain word or seen a certain arm movement after some videos. In both experiments the N400 at target words was larger when the gesture mismatched the target word compared to when they matched, suggesting task-independent gesture-speech integration.

In addition to providing evidence of gesture-speech integration, Kelly at al. (2010) proposed the integrated-systems hypothesis which states two properties explaining how speech and gestures are integrated. The first property is that gesture-speech integration is bidirectional, which means that gestures affect speech processing and speech affects gesture processing. The second property is that the suggested mutual interaction between speech and gesture is obligatory; one is not able to process only one modality when both speech and gesture are presented. By means of two experiments, the researchers attempted to find support for the integrated-systems hypothesis. In the first experiment, participants were presented with a prime video of an action (e.g. someone chopping vegetables). After each prime video they watched a target video with speech and a gesture. Sometimes the speech and the gesture matched, sometimes they were weakly mismatching (e.g. the speech “cut” and the gesture “chop”), and sometimes they strongly mismatched (e.g. the speech “chop” and the gesture “twist”). Participants were asked whether any information in the target videos related to the prime video. They performed better and faster at this task when the gesture and speech in the target videos matched than when they mismatched. Performance decreased more when they strongly mismatched compared to when they weakly mismatched. These results indicate that gesture-speech integration occurs bidirectionally. The second experiment was similarly, except for the task description. This time the participants were asked whether information only available in the speech was related to the prime videos. For this task, the findings were the same as in the first experiment, suggesting that even when the gestures do not require attention, this modality is still integrated with speech. This indicates that gesture-speech integration is an obligatory process. The findings that gesture-speech integration is both bidirectional and obligatory provide evidence of the integrated-systems hypothesis.

Besides EEG studies looking at the N400 component to provide neural evidence of gesture-speech integration, Willems, Özyürek, and Hagoort (2007) looked at gesture-speech integration using fMRI. They assumed that if gesture-speech integration takes place in a brain area, this brain area would be more active in a condition where speech is accompanied by a mismatching gesture, and more inactive in the condition where matching gestures co-occurred with speech. They found that this was the case in the left inferior frontal cortex and the left superior temporal cortex, areas where movement and language processes occur. This suggests that integration of the two modalities also occurs in these areas.

The previously discussed studies provide sufficient evidence that speech and gestures are integrated during processing. In addition, some studies provide evidence that gesture-speech integration is enhanced or constrained under certain circumstances. To understand the full process and the find the optimal circumstances for this integration, four properties are discussed next.

One of those properties that influence the effectiveness of gesture-speech integration is synchrony. Obermeier et al. (2011) tested gesture-speech integration in different synchrony conditions. In the synchronous condition speech and gestures were simultaneously presented and in the asynchronous condition the gesture preceded the speech. When speech and gesture were synchronous, neural effects were found that the participants used the meaning of the gesture for speech comprehension. In the asynchronous condition, the same results were only found when participants were explicitly told to pay attention to the gesture and speech. Participants did not use the meaning of gestures for speech comprehension when the task was implicit. This suggests that synchrony has impact on gesture-speech integration in the meaning that integration only occurs in an asynchronous condition when people pay explicit attention to the gesture-speech combination. Another study that also found an effect of synchrony on gesture-speech integration is Habets et al. (2011). They compared three levels of synchrony, synchronous gesture and speech, 160 ms asynchrony, and 360 ms asynchrony. Participants seemed to integrate gesture and speech in the synchronous and the 160 ms asynchronous conditions, but not in the 360 ms asynchronous condition. These findings suggest there is a timespan in which an effect of gesture on speech comprehension is most effective.

A second property that seems to impact the degree of gesture-speech integration is whether the recipient of the speech is addressed by the speaker or not. To test this, Holler et al. (2014) asked participants to watch short video clips in which an actor utters a sentence about an object. After each video clip they had to choose the object the speaker talked about.

Sometimes the speaker's eye gaze was turned to the participant and sometimes she looked in another direction. In half of the video clips a matching gesture was made and in the other half only speech was used. The researchers found that when participants were unaddressed, their response times were longer in the speech only condition than in the condition with speech and gestures. This was not found for addressed participants. Therefore, it seems like recipients' speech comprehension suffers when they are not addressed, but they extract more information from the co-speech gestures in that situation.

The different characteristics of iconic gestures form a third influencing factor. Beattie and Shovelton (1999b) classified the following properties of gestures: identity, number, description of action, manner, shape, size, movement, direction, rotation, upward movement, speed, relative position, location of action, orientation, and contact. In their experiment, participants were presented with video clips with either only speech, only a gesture, or both speech and gesture. After each video clip, the participants were asked seven questions about the fifteen properties. The amount of information obtained by the participants was significantly larger in the condition with both gesture and speech than in the other conditions with only speech or only gestures, which shows an enhancing effect of gesture. However, this difference between the speech only condition and the gesture and speech combined condition was only significant for two characteristics, size and relative position. This means that these two characteristics provide the most additional information of gestures on speech comprehension.

Lastly, in all discussed studies above providing evidence of gesture-speech integration, videos were used to show the gesture stimuli. This raised the question whether this is representative of communicative setting in reality. Holler et al. (2009) studied whether there were different findings in video and face-to-face conditions. They found that the gestural enhancement effect was at least as large or even larger when gestures were presented face-to-face compared to when videos of gestures were shown. This suggests that using video stimuli in research regarding gesture-speech integration does not lead to better results than a face-to-face setting, and is thus representative of real bimodal situations.

In all discussed studies above providing evidence of gesture-speech integration in adults, the researchers used clear speech. In the next paragraph, research will be discussed in which a different communicative situation, that of degraded speech, was used to test gesture-speech integration.

1.1.2 Gesture enhancement effect on degraded speech in adults

A more specific communicative situation in which gestures have been proved to enhance speech comprehension than in clear speech, is when speech is unclear. For example when people have hearing impairments (Obermeier, Dolk, & Gunter, 2012) or when people are in noisy environments (Kendon, 2004). In other studies, degraded speech settings have been investigated using noise-vocoded speech (e.g. Drijvers & Özyürek, 2017, 2018; Holle, Obleser, Rueschemeyer, & Gunter, 2010). Noise-vocoded speech is speech in which the speech signals have been processed to have degraded fine spectral information, but preserved temporal and amplitude information (Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995). If gesture-speech integration takes place in this communicative context, inverse effectiveness would occur. This means that gestures would have a larger enhancing effect on speech comprehension when speech is more degraded.

To study the gesture enhancement effect on degraded speech comprehension, Drijvers and Özyürek (2017) showed participants video clips of an actor uttering a word. In some of the video clips the speech was clear, in some it was 6-band degraded (moderate degradation), and in some 2-band degraded (severe degradation). To test the enhancement effect of gestures and visible speech (lip movements) on speech comprehension, different conditions of visual cues were created. In some videos there were no visual cues, which means there were no gestures and the actor's lips were blurred. Some other video clips only contained visible speech, and the last ones contained both visible speech and gestures. After watching each video clip, the participants had to type the word they thought the actor tried to convey. The researchers analyzed the effect of visible speech compared to the condition with no visual cues, and they compared the condition with both visible speech and gestures to the one with only visible speech. The last one is used to measure an enhancement effect of gestures. They found that the participants performed worse when the speech was degraded than when it was clear, and that their performance improved the more visual cues were given. Gestural enhancement was found in both degraded speech conditions. This suggests that speech and gestures are semantically integrated during processing, and it provides evidence of inverse effectiveness in this speech setting.

The same researchers also studied gesture-speech integration with degraded speech on a neural level. Here, Drijvers and Özyürek (2018) used video clips (similar to the ones used in Drijvers and Özyürek, 2017) to compare clear and degraded speech, and matching and mismatching gestures in native Dutch listeners and non-native German listeners who were all advanced learners of Dutch. The participants showed a larger N400 when the speech was degraded compared to when it was clear, in both the matching and mismatching conditions. For native listeners, a larger N400 was also found in conditions with mismatching gestures compared to those with matching gestures. Non-native listeners did not show a neural effect between matching and mismatching gestures in the degraded speech conditions, suggesting they need clearer speech to be able to use gestures. The other findings suggest that gesture-speech integration also occurs on a neural level.

1.2 Speech and gesture comprehension by children

Before discussing the development of gesture-speech integration, it is important to know when children can understand speech and iconic gesture separately. Approximately, the children's speech system develops between 0 and 5 years of age (Schaerlaekens, 1977). After this age they are roughly able to produce and comprehend speech.

Comprehension of different types of gestures are developed at different ages. The first gesture type that is comprehended are deictic gestures. Comprehension of these gestures is fully established around 18 months of age (Schulze & Tomasello, 2015). Deictic gestures are gestures in which the referential meaning is not given by the form of the gesture but by the context (Özçalışkan & Goldin-Meadow, 2011), which means children do not need to be able to connect gestures to a referent to comprehend these gestures. Gestures that do require this ability for comprehension are iconic and arbitrary gestures, which is possibly the reason why comprehension of these gestures types is developed at a later age than deictic gesture comprehension (Özçalışkan & Goldin-Meadow, 2011). The final gesture comprehension, of iconic and arbitrary gestures, takes place when children are 26 months old (Namy & Waxman, 1998; Namy, Campbell, & Tomasello, 2004).

1.2.1 Gesture-speech integration in children

Whether gesture integrates with speech has been investigated far less in children than in adults. Some studies, however, did study gesture-speech integration in children and suggest that children develop their ability to integrate gesture and speech around 3 years of age (Broaders & Goldin-Meadow, 2010; Sekine et al., 2015; Stanfield et al., 2014).

Stanfield et al. (2014) studied the age at which children are able to integrate gestures and speech by showing video clips consisting of speech and gesture combinations to 2, 3 and

4-year-old children. The gestures depicted the action and the shape of an object, for example hands moving towards the mouth with hands shaped as if eating a sandwich, and the speech was a description of the action, for example ‘I am eating’. The gesture provided additional information to the speech. After each speech-gesture combination the experimenter placed two pictures on a table. In the eating example one picture depicted a sandwich and the other one a bowl of cereals. The participants had to choose which object the experimenter meant with the speech and gesture. 3 and 4-year-old children were able to choose the correct picture above chance, while 2-year-olds could not. This shows that children start using the meaning of gesture to comprehend a message when they are 3 years old.

Another study showing gesture-speech integration by the age of 3 was done by Sekine et al. (2015). 3-year-old and 5-year-old children and adults were presented with videos that either contained only speech, only gesture or a combination of speech and gesture. The speech consisted of action verbs and the gesture depicted one of the ways the action could be carried out. After each video clip, four pictures were shown of which one matched both the gesture and speech, one only the speech, another one only the gesture, and the last picture mismatched both the gesture and speech. The researchers found that when 5-year old children and adults were presented with both speech and gesture, they were able to choose the correct picture more often than 3-year-old children. However, since children under the age of 3 learn more from live face-to-face situations than from videos (Anderson & Pempek, 2005), Sekine et al. conducted a second experiment. This was almost a replication of the first experiment with the exception that the stimuli were now produced live by the experimenter and only 3-year-olds were tested. In this live setting, the 3-year-old children did perform above chance in the condition with both speech and gestures and this score was also significantly higher than in the video setting in the first experiment.

Evidence of gesture-speech integration in older children was also found in a different setting. Since the way questions are asked in witness interviews can influence the things people think they remember, gesture might also have an effect. Broaders and Goldin-Meadow (2010) found that 5 and 6-year old children were more likely to confirm situations that had not occurred when the speech in the interviews was accompanied by gestures compared to when it was not. This means that 5 and 6-year-old children used information that was only conveyed by gestures and thus that they can integrate gesture and speech.

In addition to behavioral evidence of gesture-speech integration in children, Sekine et al. (under review) and Schoechl (2018) studied the possible neural semantic integration by means of the ERP component N400. They presented 6 and 7-year-old children with short video clips consisting of speech in combination with either a matching or a mismatching gesture while measuring their ERPs. The children showed a larger N400 in the condition with mismatching speech and gesture compared to the condition in which the speech and gesture matched. These findings suggest that gesture-speech integration not only occurs behaviorally in 6 and 7-year-old children, but also neurally.

Lastly, to complete our understanding of gesture-speech integration in children, two neuroimaging studies are discussed. Dick, Goldin-Meadow, Solodkin, and Small (2012) suggest that the brain areas that become active when children see gestures that mismatch the speech and that are inactive when they match seem to be overlapping with the active areas in adults’ brains. These areas are the inferior frontal gyrus and the posterior middle temporal gyrus. However, the researchers found that the pattern of the activity differed between adults and children. Partially similar results were found by Demir-Lira et al. (2018). In conditions where gestures disambiguated speech, the researchers measured increased brain activity in inferior frontal gyri, the left superior temporal gyrus and the right middle temporal gyrus in 8 to 10-year-old children, compared to when gestures accompanied unambiguous speech. These areas also roughly overlap with the ones active in adults' brains (Willems et al., 2007) although it is a broader area in children.

1.2.2 Gesture enhancement effect on degraded speech in children

Whether gestures can enhance comprehension of degraded speech in children has barely been studied. Eisenberg, Shannon, Martinez, Wygonski, and Boothroyd (2000) found that 5 to 7-year-old children were worse at degraded speech comprehension than 10 to 12-year-olds and adults, they needed more spectral resolution to perform at the same level as the older participants. However, it was still unknown what this meant for the effect of gestures on degraded speech comprehension in children. Therefore, Sekine and Özyürek (in prep) investigated whether children would have similar enhancement from gestures on degraded speech comprehension as adults. They presented 6 and 7-year-old children with video clips of an actor uttering a word. In some videos the speech was 4-band noise-vocoded (severe degradation) and in some others 8-band noise-vocoded (moderate degradation). In half of the videos with clear speech and in half of the ones with degraded speech she also made a gesture that matched the speech. The children were asked to repeat the word the actor was trying to say. They found evidence that gesture-speech integration occurs in these children on a behavioral level, since gestures had an enhancing effect on the accuracy scores. Compared to adults, children needed less degradation to show optimal gestural enhancement. They showed the largest enhancement effect in the 8-band degradation while adults used gestures best in the 4-band degradation. However, there is no neural evidence of gesture-speech integration with degraded speech in children.

1.3 Present study

In adults, in both clear speech and degraded speech settings gesture-speech integration has been studied, and this has been done on both the behavioral level (e.g. Drijvers & Özyürek, 2017; Beattie & Shovelton, 1999a, 1999b; Holler et al., 2009; Holler et al., 2014; Kelly et al., 2010; Kelly et al., 2015; Riseborough, 1981) and the neural level (e.g. Drijvers & Özyürek, 2018; Habets et al., 2011; Holle & Gunter, 2007; Holler et al., 2014; Kelly et al., 2004; Obermeier et al., 2011; Özyürek et al., 2007; Willems et al., 2007; Wu & Coulson, 2007a). In children, gesture-speech integration has not been studied a lot. Nevertheless, some studies did provide evidence of this integration in clear speech, both behaviorally (Broaders & Goldin-Meadow, 2010; Sekine et al., 2015; Stanfield et al., 2014) and neuroscientifically (Demir-Lira et al., 2018; Dick et al., 2012; Schoechl, 2018; Sekine et al., under review). In addition, Sekine & Özyürek (in prep) did a behavioral study on gesture-speech integration in a degraded speech setting and provided evidence that gesture-speech integration also takes place in children in this setting. The part that still lacks evidence is whether children show neural processes of gesture-speech integration in a degraded speech setting, which was the reason to conduct the present study.

In the present study, 6 and 7-year-old children were presented with short video clips in which an actor utters a verb. In half of the videos the speech is clear, and in the other half the speech is 8-band degraded. This level of degradation was used because this seems to be the optimal level for children to use gestures for speech comprehension (Eisenberg et al., 2000; Sekine & Özyürek, in prep). In addition, half of the videos in both speech conditions consist of speech only, and the other half contains gestures matching the speech. After each video clip, the children are asked to repeat the word they think the actor uttered. Multiple other studies (e.g. Schoechl, 2018; Sekine et al., under review) used and compared matching and mismatching gestures, but since the ecological validity is very low in this situation due to the lack of mismatching gestures in reality, in this study matching gestures and no gestures are compared. During the experiment, the brains' electrophysiological response is measured. EEG was chosen as the measure of the neural processes of gesture-speech integration, because the online ERP component N400 reflects the neural effort of semantic processing. Since the integration of speech and gestures would cost more effort than processing only one modality, this would be visible in a larger N400 amplitude.

In line with previous research, the behavioral effects of gesture-speech integration are reflected in accuracy scores, in this study the percentage correctly repeated verbs. The children's accuracy scores are predicted to be higher in clear speech conditions than in degraded speech conditions. Furthermore, if the two modalities integrate, the expectation is that the presence of gestures would facilitate speech comprehension when the speech is degraded assuming inverse effectiveness operates. Since clear speech is not an adverse situation, gestures are not expected to benefit speech comprehension.

On a neural level, it is expected to observe a larger N400 in the degraded speech conditions compared to the clear speech conditions because degraded speech comprehension is more effortful than clear speech comprehension. In addition, a larger N400 would be found in the conditions with gestures than in those without gestures, since processing gestures seems to be obligatory and this semantic processing costs more effort. If gesture-speech integration occurs, the difference in N400 between the gesture and no gesture conditions will also be larger in the degraded speech condition than in the clear speech condition. This is because semantically integrating speech and gestures instead of processing only one of the two modalities in an adverse listening condition is more effortful due to inverse effectiveness than in clear speech.

2. Method

2.1 Participants

17 native Dutch speaking children of 6 or 7 years old participated in this study, but due to an excessive amount of artifacts caused by movement, technical problems, and someone who had to use the restroom in the middle of the experiment, the data of 10 participants (3 females, $M_{AGE} = 7{,}0$, $SD_{AGE} = 0.56$) could be used. All children were raised monolingually, and 9 of them were right handed. None of the participants reported having any hearing or speech impairments. To recruit children as participants for this experiment, many elementary schools were contacted, flyers were distributed in supermarkets, in waiting rooms, in the Radboud University, and the Lindenberg. Flyers were also distributed at a child language festival at a library.

2.2 Materials

A part of the video stimuli and verbs were taken from Sekine et al. (under review) and Schoechl (2018). The original list of Dutch action verbs was created by Drijvers & Özyürek (2017), but Sekine et al. and Schoechl chose only a selection of the verbs based on the criterion that at least 80% of 5 and 6-year-old children in the Netherlands are familiar with them (Bacchini, Boland, Hulsbeek, Pot, & Smits, 2005; Schaerlaekens, Kohnstamm, Lejaegere, & Vries, 1999). 160 of these verbs were used in the present study.

In each video clip, a native Dutch female speaker uttered an action verb. In half of the videos, she made a gesture that matched the speech. The video clips in which a gesture was

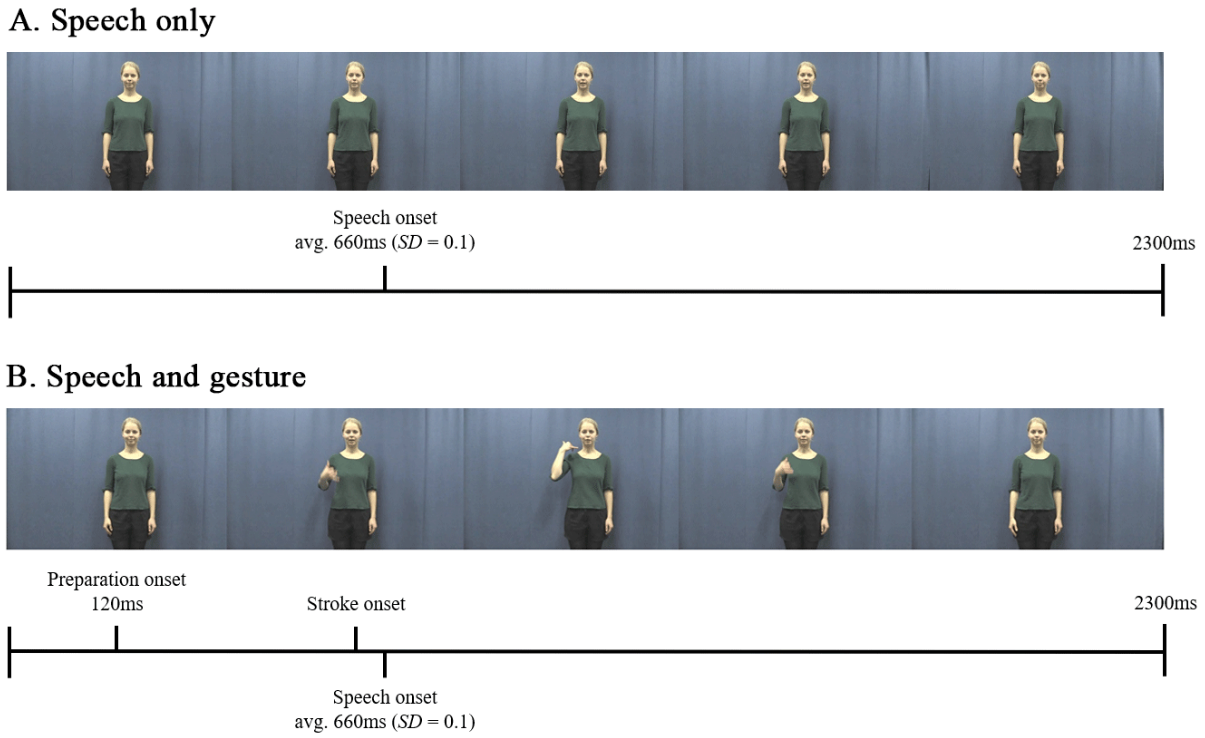


Figure 1. The timeline of the video stimuli. *A. Speech only.* The actor produced speech without gestures. The video clips were 2300 ms long and the speech onset started on average at 660 ms. *B. Speech and gesture.* The actor produced speech with gestures. The video clips were also 2300 ms long and the speech onset started on average at 660 ms. The onset of gesture preparation was at 120 ms, and the stroke onset was located shortly before the speech onset.

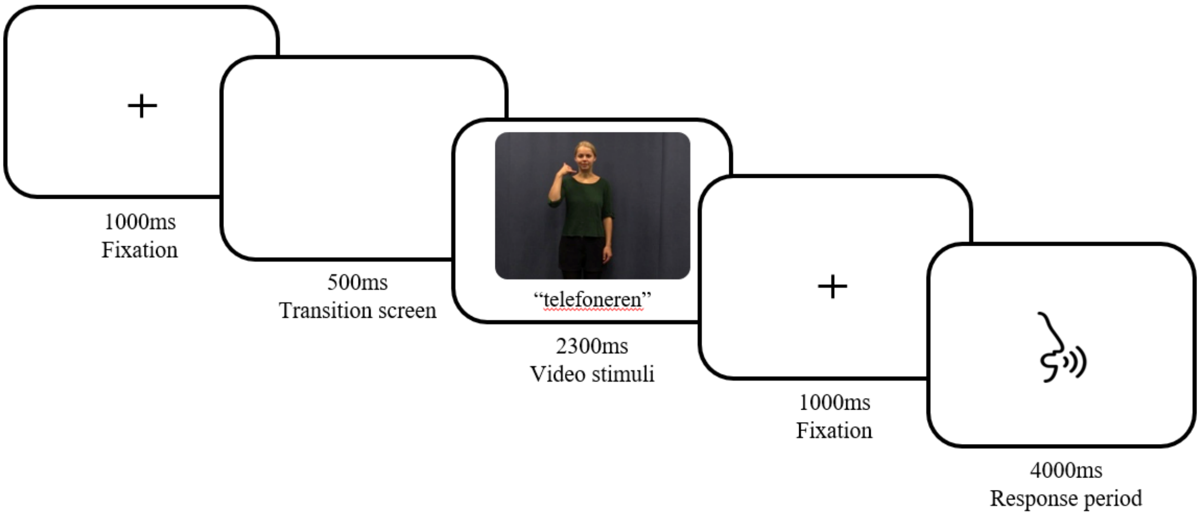


Figure 2. The design of the trials. Each trial started with a fixation cross lasting for 1000 ms, then there was a transition screen for 500 ms, after that a 2300 ms long stimuli was shown, then a fixation cross for 1000 ms, and lastly, the speech icon appeared for 4000 ms.

made were created by Sekine et al. (under review) and Schoechl (2018). The actor was instructed to utter a verb and make a gesture simultaneous with the speech. The gestures were made spontaneously, but the researchers checked whether they were iconic gestures that represented the action the verb implied. The actor was asked to speak in a child-directed voice. In the videos, the actor was visible from her head to her knees. She was asked to have a neutral but friendly facial expression and to keep her arms casually hanging on either side of het body. Her clothing and the background were in a neutral color. The camera Canon XF205 was used to record the video clips, and they were edited to have a length of 2300 ms with ELAN (Version 4.6.1, Lausberg & Sloetjes, 2009). The preparation of each gesture started 120 ms after video onset and the average speech onset was at 660 ms (SD = 0.1). The gesture stroke, which is the part that contains the meaning of an iconic gesture, started before the speech onset. In a similar way, the video clips without gestures were created. The timelines of videos in both conditions are shown in Figure 1.

In both the video clips with gestures and those without gestures, the speech in half of the stimuli was degraded. This was done using noise-band vocoding (Shannon et al., 1995). A degradation of 8-band noise vocoding was chosen because behavioral studies revealed that children benefit most from gestures at this degradation level (Eisenberg et al., 2000; Sekine & Özyürek, in prep).

Stimuli were presented in 4 blocks and each block contained 40 trials. The design of the trials is shown in Figure 2. Each trial started with a black screen with a white fixation cross in the middle for 1000 ms. Then there was a black transition screen for 500 ms to measure the baseline. After that, a 2300 ms long stimulus was shown followed by another fixation cross for 1000 ms. Finally, a speech icon was shown for 4000 ms in which the children had to repeat what they think they heard in the stimulus. There were different lists for each participant in which the order and verb-condition combinations varied.

2.3 Pre-test

To test whether the iconicity of the gestures was sufficient for 6 and 7-year-old children to disambiguate the verbs, Schoechl (2018) conducted a pre-test. 104 children (M_{AGE} = 6.74, SD_{AGE} = 0.64) were divided into eight groups. Each group was presented with a different list of 43 or 44 videos in which an actor tried to communicate a word only with her hands. After that, the researchers named a verb and asked the children to what extent the movement looked like that verb. In half of the items the gestures matched the verb and in the other half they mismatched the verb. The children had to answer the question with stars, 5 stars meaning they looked very much alike, and 1 star meaning the gesture did not look like the verb at all. Four practice items were included to make sure the children understood the rating procedure. Based on the criteria that the mean rating of the matching items was 3 or above (SD ± 1) and the mean rating of mismatching items was 2.6 or below (SD ± 1), 120 verbs were selected.

2.4 Procedure

When a child and his or her caregiver arrived, we asked the child to sit down on a chair. One or two of the experimenters explained that they were going to put a cap on the child's head to measure their brain activity. The child was told that it is safe and would not hurt. When the child agreed, the child's head was measured and the cap was put on. After showing the child that the needles of the syringes with gel were thick and harmless, the experimenter(s) put gel in the cap to make sure that the electrodes would be in contact with the head. In the meantime the child could color lightbulbs on a paper when the lights of the electrodes changed from red to green. While the cap was prepared, another experimenter asked the parent to fill out a consent form. After that, the parents were also asked to fill out a participant information form and the Edinburgh Handedness Inventory (Oldfield, 1971), respectively attached in Appendix A and Appendix B. By means of the participant information form information was obtained about the child's and parent's date of birth, whether the child had knowledge of another language than Dutch, the type of schooling the child received, and the parent's highest level of education. The Edinburgh Handedness Inventory (Oldfield, 1971) was used to find out which hand the child used in different situations.

When the EEG cap was ready, the child was taken into a soundproof booth to play a language game on a computer. The children were shown the brain activity signal measured by the electrodes and the effect of, in particular, eye, head and jaw movements on the brain waves. Then the experiment, presented on in Presentation® software (version 18.0, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com), could start. The children were explained that they were going to watch short videos in which an actor says one word. After each video a speech icon would come up and then the child had to repeat what they think the actor said. Three important aspects were pointed out. First, the child should keep looking at the fixation cross that appeared in the middle of the computer screen before each video and between the videos and speech icons. That way, the cleanest brain waves would be obtained. Second, the child had to wait with repeating the heard word till the speech icon appeared. Lastly, they should try to move and blink as little as possible during the experiment, particularly during the videos. To check whether the children understood the task, eight practice items were used. Then the children were told that there was a camera and a microphone in the booth so their parents and the experimenters could always see and hear them. After that, everyone would leave the booth unless the child was too scared to be left alone. In that case the parent could take a seat in the back of the booth, with the instruction not to distract the child in any way.

After each block, the children had a small break in which they were given a puzzle and they could put a stamp on a treasure map. At the end of each break, the children were told that they had to continue doing the same task until the next break and they were reminded not to move. If they kept moving during a block or touched the electrodes on their face, an experimenter had to interfere by speaking through the microphone, gently asking the child to sit still for a little bit longer.

When the experiment ended and the last puzzle and treasure map were finished, the child was led out of the computer booth and the cap was taken off. Because of the gel left behind on the children's head, an experimenter showed them the shower room where their parent could wash their hair.

Upon return, the children received a certificate of participation with their name on it and they could choose a sticker sheet and a balloon. In the first sessions the parents had to fill out a payment form in order to receive the compensation of 30 euros. In later sessions they received a 30 euros worth VVV gift card instead.

A test session took no more than two hours, including preparation of the cap and washing the child's hair afterwards. The computer task took about 40 minutes, including the three short breaks.

2.5 Behavioral data analysis

The voice data was analyzed using Praat (Boersma & Weenink, 2015). Per repeated action verb, the speech onset was noted down and whether the participant repeated the correct word.
Only children of whom the EEG data was used, were included in the behavioral data analysis.

In this experiment, a within-subject design was used. The dependent variable was the percentage of correctly repeated action verbs per condition. There were two independent variables, both with two levels: whether the video clips contained speech only or speech in combination with a gesture, and whether the speech was clear or degraded. Combined, these variables formed four conditions: clear speech only, degraded speech only, clear speech in combination with a gesture, and degraded speech in combination with a gesture. To analyze these behavioral data, a 2 by 2 (Speech quality: clear vs. degraded speech, and Modality: speech only vs. speech and gesture) repeated-measures analysis of variance (ANOVA) was used.

2.6 EEG data acquisition and analysis

EEG data was measured from 32 Ag-AgCl electrodes. Of these electrodes, 27 were mounted in the cap (actiCap) according to the 10-20 standard system, two were placed on the temples on either side of the head to measure horizontal electrooculograms, and two were placed under and above the left eye to measure vertical electrooculograms. Another electrode was placed on the right mastoid, and the last electrode on the left mastoid for re-referencing. The EEG was filtered through a 0.02 – 100 Hz band-pass filter and on-line digitized with a 500 Hz sampling frequency (BrainVision Recorder, Brain Products, Gilching, Germany).

For the pre-processing of the EEG data, BrainVision Analyzer 2.1 (Version 2.1.1, Brain Products, Gilching, Germany) was used. First, we re-referenced the EEG data offline to the average of the left and right mastoid and filtered the data with a high-pass filter at 0.01 Hz and a low-pass filter at 35 Hz. Then we segmented the data into epochs from 200 ms before video onset to 1900 ms after video onset. After that, a semi-automatically executed analysis was done to remove all movement artifacts, except for eye movement artifacts. The artifacts caused by blinks and eye movements were semi-automatically identified and removed in the next part of the analysis, by means of an Ocular Independent Component Analysis (ICA). In the last step, all remaining artifacts were removed. The mean number of trials that remained was 23 (SD = 4.65) for clear speech trials without gesture, 24 (SD = 3,67) for degraded speech trials without gesture, 23 (SD = 5,14) for clear speech trials with gesture, and 24 (SD = 4,51) for degraded speech trials with gesture. For each participant, the average of the excluded trials was 41% (66/160). Then the output file for the statistical analysis in SPSS was created, with the time-locked averages in the 300-500 ms time window after speech onset of all channels per participant per condition.

In SPSS, the reference channel and the four channels measuring eye movement were removed. Then, the average of all channels was calculated for each participant and condition.

On this data, a 2 by 2 repeated-measures ANOVA was performed, with Speech quality (clear vs. degraded speech) and Modality (speech only vs. speech and gesture) as independent variables and the mean amplitude of the brain activity as the dependent variable.

To check for effects in different parts of the head, averages per condition were also calculated for the left and right hemisphere separately. Similar as for the whole head, these separate brain areas were analyzed by means of a 2 by 2 repeated-measures ANOVA.

3. Results

3.1 Behavioral results

The mean percentages of correct answers and the standard errors for the two speech quality levels, clear and degraded, and the two modality levels, gesture or no gesture, are shown in Table 1. In Figure 3, these mean accuracy scores are presented by means of a boxplot. The accuracy scores seem to be higher in the clear speech conditions than when the speech was degraded. In addition, the percentage of correct answers seems to be higher in the degraded speech condition when the speech was accompanied by a gesture than without a gesture.

The 2 by 2 repeated-measures ANOVA showed a significant main effect of Speech quality on the mean percentage of correct answers, $F(1, 9) = 65.89, p < .001, \eta^2_p = .88$. Children were better at repeating verbs when the speech was clear than when the speech was degraded. There was no significant main effect of Modality on the mean percentage of correct answers, $F(1, 9) = 4.97, p = .05, \eta^2_p = .36$. Speech accompanied by a gesture did not facilitate repeating verbs compared to speech without a gesture. Finally, the analysis showed an interaction effect of Speech quality and Modality on the mean percentage of correct answers, $F(1, 1) = 15.05, p < .01, \eta^2_p = .63$. Bonferroni corrected post-hoc analyses revealed a significant difference

Table 1. Mean accuracy scores in percentages and standard errors between brackets.

		Speech quality	
		Clear	Degraded
Modality	No gesture	96.25 (1.30)	54.25 (4.86)
	Gesture	95.00 (1.29)	64.75 (4.84)

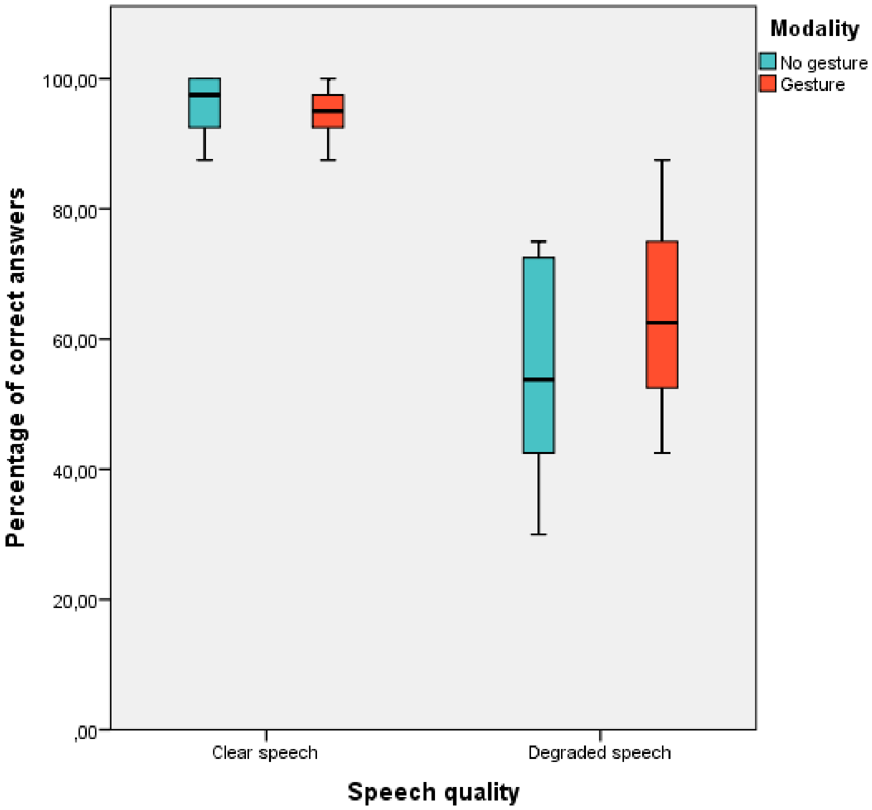


Figure 3. Boxplot of the behavioral data for both speech types and modalities. The horizontal black lines within the boxes depict the median of the percentage of correct answers, the lowest end of the vertical bars shows the minimum score and the highest end the maximum score.

between gestures and no gestures in the degraded speech condition, $t(9) = 3.19$, $p_{bon} = .02$, with a higher mean percentage of correct answers in the gesture condition than in the condition without gestures. This result showed that gesture only affects children's speech comprehension when speech is degraded.

3.2 EEG results

The channels on the whole head were included for the first statistical analysis. In Figure 4, waveforms of the mean amplitude of the brain activity are shown for both speech quality conditions and both modality conditions. Negativity is plotted upward. In the figure, the onset of the videos occurs at 0 ms and speech onset on average at 660 ms. Between 300 and 500 ms after speech onset (indicated by the black box in Figure 4) the mean amplitude was more negative in the conditions with gestures compared to those without gestures, in particular in the condition with degraded speech. Figure 5 shows the topographical plots with the mean amplitude for each speech and gesture condition in the 300-500 ms time window after speech onset. There was barely a difference between clear and degraded speech, but the presence of co-speech gestures created a distinctly different pattern compared to speech without gestures. Gesture caused more negative brain activity in frontal and central brain areas.

The 2 by 2 repeated-measures ANOVA showed no significant main effect of Speech quality on the mean amplitude, $F(1, 9) = .05$, $p = .83$, $\eta^2_p < .01$. So the mean amplitude was not significantly more negative when children heard degraded speech than when they heard clear speech. A significant main effect of Modality on the mean amplitude was found, $F(1, 9) = 9.03$, $p = .02$, $\eta^2_p = .50$. When gestures accompanied the speech, this led to more negative mean amplitude compared to speech without a gesture. Finally, the analysis showed no interaction effect of Speech quality and Modality on the mean amplitude, $F(1, 9) = 3.28$, $p = .10$, $\eta^2_p = .27$.

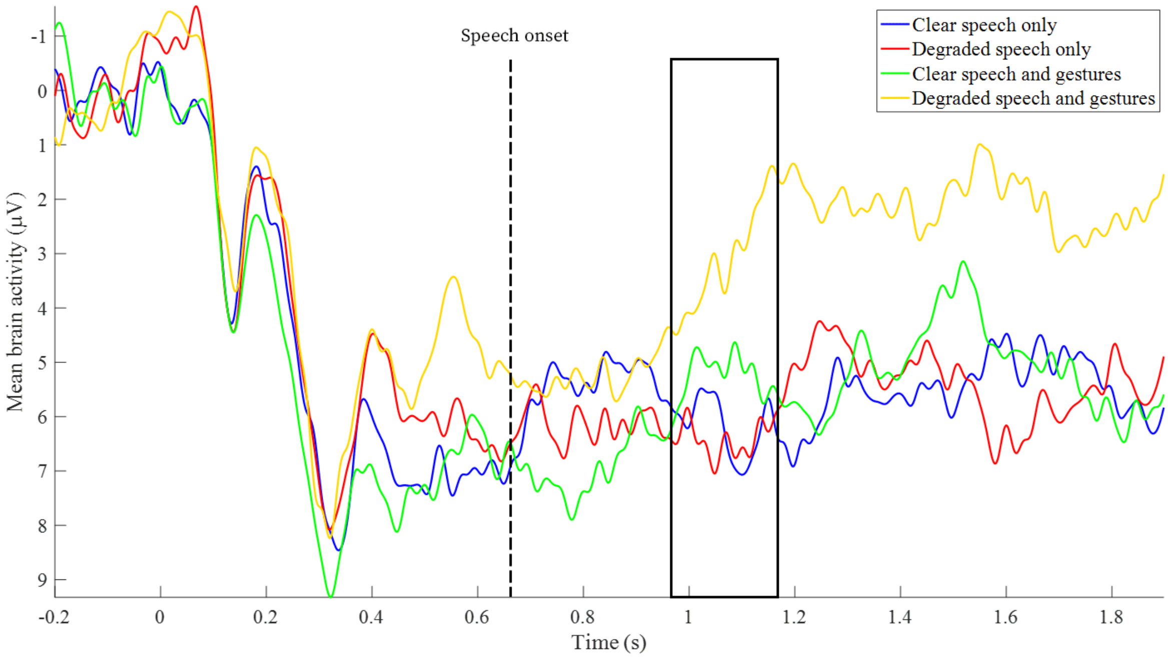
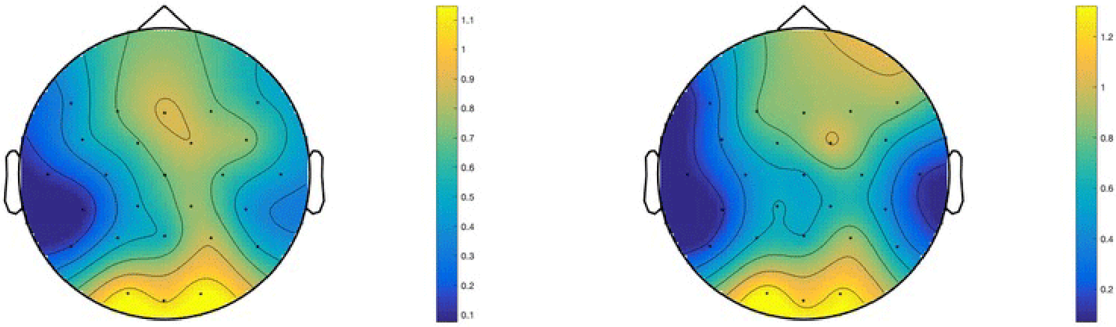


Figure 4. Grand average waveforms for ERPs measured by the channels on the whole head. The mean amplitude of the brain activity is shown for both speech types and gesture conditions. Negativity is plotted upward. The waveforms are time-locked to video onset, speech onset starts on average at 660 ms. The black box shows the 300-500 ms time window after speech onset, between 960 and 1160 ms after video onset.

A. Clear speech only **B. Degraded speech only**



C. Clear speech and gestures **D. Degraded speech and gestures**

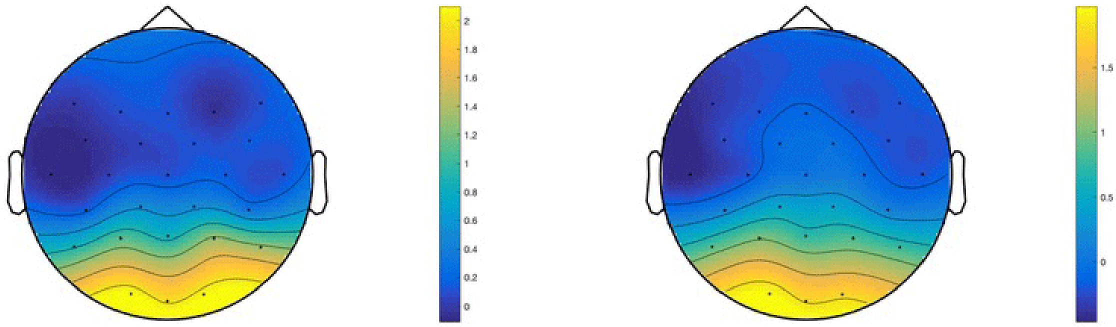


Figure 5. Topographical plots of both speech and gesture conditions in the 300-500 ms time window after speech onset. *A. Clear speech only.* The mean amplitude of the brain activity in the condition with clear speech and no gestures. *B. Degraded speech only.* The mean brain amplitude in the condition with degraded speech and no gestures. *C. Clear speech and gestures.* The mean amplitude in the condition with clear speech and gestures. *D. Degraded speech and gestures.* The mean amplitude in the condition with degraded speech and gestures.

Two more 2 by 2 repeated-measures ANOVAs with the same variables were conducted to find possible effects in the left and right hemisphere separately. This analysis regarding the left hemisphere showed no main effect of Speech quality, $F(1, 9) = .14$, $p = .72$, $\eta^2_p = .02$. There was a significant main effect of Modality, $F(1, 9) = 8.07$, $p = .02$, $\eta^2_p = .47$, showing a more negative amplitude of the brain activity in conditions with gestures than in conditions without gestures. No interaction effect of Speech quality and Modality was found, $F(1, 9) = 2.39$, $p = .16$, $\eta^2_p = .21$. Also regarding the right hemisphere, the analysis showed no main effect of Speech quality, $F(1, 9) < .01$, $p = .93$, $\eta^2_p < .01$. A significant main effect of Modality was found, $F(1, 9) = 9.43$, $p = .01$, $\eta^2_p = .51$, again with a more negative amplitude in the gesture conditions compared to the conditions without gestures. There was no interaction effect, $F(1, 9) = 3.90$, $p = .08$, $\eta^2_p = .30$.

4. Discussion

The present study was conducted to find neural evidence of gesture-speech integration in 6 and 7-year-old children. To test this, the principle of bimodal enhancement and inverse effectiveness were used. Bimodal enhancement with gesture and speech means that speech comprehension can be facilitated by the use of accompanying matching gestures, and inverse effectiveness suggests that this enhancement of gestures on speech comprehension is larger in more difficult settings. In this study the auditory modality was made more difficult by degrading speech. The participants were presented with video clips in which an actor uttered a verb. In half of the videos the speech was clear and in the other half the speech was degraded. In addition, half of the videos in both speech conditions did not contain a gesture, and in the other half the actor made a matching gesture during her speech. After each video clip, the children were asked to repeat the verb they think the actor tried to utter. If gesture-speech integration would occur in 6 and 7-year-old children, this would be visible in behavioral and neural processes. Behaviorally, they would show higher accuracy scores on the repeating task when the speech is clear compared to when it is degraded. In addition, the children would perform better at degraded speech comprehension when this speech was accompanied by a gesture compared to when they heard degraded speech only. On a neural level, semantic integration of speech and gestures is more effortful than the processing of only one modality which suggests that gesture-speech integration would lead to a larger N400 when degraded speech is combined with gestures than in the degraded speech condition without gestures.

As expected on the basis of previous behavioral research (Sekine & Özyürek, in prep) the mean percentage of correct answers was higher in the clear speech condition compared to the degraded speech condition. This finding suggests that degraded speech comprehension is more difficult than clear speech comprehension. Also no overall enhancing effect of gestures was found, but this was expected because clear speech comprehension is a simple task in which gestures are not needed. Lastly, an interaction effect showing that gesture facilitates degraded speech comprehension was found as predicted. These behavioral results suggest that 6 and 7-year-old children integrate speech and gestures during processing.

On a neural level measured on the whole brain, it was expected to find a larger N400 for degraded speech than for clear speech because degraded speech comprehension requires more effort. However, this was not found. There was a significant difference in brain activity between the gesture and the no gesture conditions, suggesting that the processing of two modalities needs more effort to process semantic information. In contradiction to the expectations, no interaction effects of the different speech and gesture types were found. This finding that processing gestures does not show more processing effort in degraded speech comprehension compared to no gestures and degraded speech was unexpected because this would suggest that inverse effectiveness does not occur. And, therefore, that gesture-speech integration does not exist on a neural level in these children. To check for possible effects in specific areas in the brain, analyses were also done for the left and right hemisphere separately. Similar effects as for the whole brain were found for both hemispheres.

The unexpected results of this study are interesting. It would have been very likely that gesture-speech integration occurs on a neural level in children, since neural effects were found in children of the same age by Schoechl (2018) and Sekine et al. (under review) between matching and mismatching gesture conditions. It was also expected because adults showed neural effects of gesture-speech integration in the degraded speech setting (Drijvers & Özyürek, 2018) and on the behavioral level children seemed to experience gestural enhancement similarly as adults in the same speech setting (Sekine & Özyürek, in prep). The fact that some expected neural effects were absent could suggest that children's brains are not yet developed similarly as adults' brains for the ability to neurally integrate gestures and speech. However, a tendency of the expected results is visible in Figure 4. The effects were probably not all significant because of a large spreading of the brain activity which could be improved by testing more participants. This makes it more likely that the unexpected findings are caused by the small number of participants in the present study. This makes it very hard to draw conclusions from the results. For that reason, it is very important to replicate this study with a highly increased number of participants. Another aspect that could have had an effect on the different findings in this study compared to those in Sekine and Özyürek (in prep) is the stimuli design. In the present study, the children had to wait till the speech icon appeared to repeat the verb while the children in Sekine and Özyürek's study could response immediately. The fact that they had to wait could have resulted in frustration and more interfering movements causing extra artifacts. This design could be changed in replication studies.

If replication studies are conducted and indicate that 6 and 7-year-old children can integrate speech and gesture, which tendency is visible in the present data and plots (see Table 1, Figure 3, and Figure 4), it would also be interesting to conduct a similar study with younger children, and if no gesture-speech integration is found, with older children. Also comparisons with gesture-speech integration in adults can be useful.

In the present study only the enhancement effect of gestures was taken into account, but visible speech (lip movements during speech) seemed to be able to enhance speech comprehension as well in adults (Drijvers & Özyürek, 2017). It could still be tested whether this has a similar effect in children or whether they are less successful at making use of lip movements.

There are many more aspects of gesture-speech integration in children that could be studied in the future. For example the properties that can influence integration of gesture-speech in adults found in previous research, synchrony of gesture and speech (Habets et al., 2011; Obermeier et al., 2011), whether or not the speaker addresses the listener (Holler et al., 2014), properties of the gestures like size and relative position (Beattie & Shovelton, 1999b), and whether the communicative setting is face-to-face or videos (Holler et al., 2009). If all of these properties would also be tested for children, the processes of gesture-speech integration would become much clearer. That way, the optimal conditions for this integration can be used in later research.

Another relevant characteristic of gesture-speech integration is the integrated-systems hypothesis proposed by Kelly et al. (2010). This hypothesis explains how the integration processes of the two modalities work. Again, this has only been studied in adults. It would be interesting to find out if gesture-speech integration in children is also bidirectional and obligatory.

5. Conclusion

On the basis of previous research, children seem to be able to integrate speech and gestures in clear speech settings on both the behavioral and neural level. Also evidence of gesture-speech integration in a degraded speech setting was provided on a behavioral level. The present study investigated the effects of gesture on degraded speech comprehension on a neural level. The behavioral data were in line with previous findings, but the EEG data showed some unexpected neural findings. This result suggests that 6 and 7-year-old children are not as developed as adults on a neural level. However, due to the small number of participants is it very difficult to make conclusions about the present data. In addition, differences in the stimuli design compared to previous studies can have affected the present findings. Since the present data seem to show a tendency of the expected enhancement effects, it would be very interesting to repeat this study with a higher number of participants and an optimized stimuli design. Additionally, replication studies are also very important to provide more evidence of the presence of gesture-speech integration in a degraded speech setting in children.

Acknowledgements

During my thesis I had a lot of support from my supervisors. First I want to thank Prof. Dr. Asli Özyürek for making time to be my supervisor. When I had to think of a research topic for my thesis, I could not easily come up with good ideas. When I asked her, she immediately offered me a place in your research group to work on this subject. I appreciate that a lot.

I also really want to thank Dr. Kazuki Sekine for being my supervisor. Since pretty much everything I had to do during my thesis was new for me (EEG, testing children, the programs I used, the working place, etc.), I was really happy he taught me all of it patiently in the first months, and that I could always ask questions if I needed help. Even when he was back in Japan, he was always ready to help me during Skype meetings.

Last but not least, I am also thankful to Dr. Joost Rommers, for taking the time to be my second evaluator.

References

Anderson, D. R., & Pempek, T. A. (2005). Television and very young children. *American Behavioral Scientist, 48*(5), 505-522. https://doi.org/10.1177/0002764204271506

Bacchini, S., Boland, T., Hulsbeek, M., Pot, H., & Smits, M. (2005). Een gefundeerde woordenlijst gefaseerd naar verwerving Duizend-en-een-woorden.

Beattie, G., & Shovelton, H. (1999a). Do iconic hand gestures really contribute anything to the semantic information conveyed by speech? An experimental investigation. *Semiotica, 123*(1/2), 1-30. https://doi.org/10.1515/semi.1999.123.1-2.1

Beattie, G., & Shovelton, H. (1999b). Mapping the range of information contained in the iconic hand gestures that accompany spontaneous speech. *Journal of Language and Social Psychology, 18*(4), 438-462. https://doi.org/10.1177/0261927X99018004005

Boersma, P., & Weenink, D. (2015). Praat: Doing phonetics by computer.

Broaders, S. C., & Goldin-Meadow, S. (2010). Truth is at hand: How gesture adds information during investigative interviews. *Psychological Science, 21*(5), 623-628. https://doi.org/10.1177/0956797610366082

Demir-Lira, Ö. E., Asaridou, S. S., Beharelle, A. R., Holt, A. E., Goldin-Meadow, S., & Small, S. L. (2018). Functional neuroanatomy of gesture-speech integration in children varies with individual differences in gesture processing. *Developmental Science, 21*(5), e12648. https://doi.org/10.1111/desc.12648

Dick, A. S., Goldin-Meadow, S., Solodkin, A., & Small, S. L. (2012). Gesture in the developing brain. *Developmental Science, 15*(2), 165-180. https://doi.org/10.1111/j.1467-7687.2011.01100.x

Diederich, A., & Colonius, H. (2004). Bimodal and trimodal multisensory enhancement: Effects of stimulus onset and intensity on reaction time. *Perception & Psychophysics, 66*(8), 1388-1404. https://doi.org/10.3758/BF03195006

Drijvers, L., & Özyürek, A. (2017). Visual context enhanced: The joint contribution of iconic gestures and visible speech to degraded speech comprehension. *Journal of Speech, Language, and Hearing Research, 60*, 212-222. https://doi.org/10.1044/2016_JSLHR-H-16-0101

Drijvers, L., & Özyürek, A. (2018). Native language status of the listener modulates the neural integration of speech and iconic gestures in clear and adverse listening conditions. *Brain and Language*, *177–178*, 7-17. https://doi.org/10.1016/j.bandl.2018.01.003

Eisenberg, L. S., Shannon, R. V., Martinez, A. S., Wygonski, J., & Boothroyd, A. (2000). Speech recognition with reduced spectral cues as a function of age. *The Journal of the Acoustical Society of America, 107*(5), Pt. 1, 2704-2710. https://doi.org/10.1121/1.428656

Habets, B., Kita, S., Shao, Z., Özyürek, A., & Hagoort, P. (2011). The role of synchrony and ambiguity in speech-gesture integration during comprehension. *Journal of Cognitive*

Neuroscience, 23(8), 1845-1854. https://doi-org.ru.idm.oclc.org/10.1162/jocn.2010.21462

Holle, H., & Gunter, T. C. (2007). The role of iconic gestures in speech disambiguation: ERP evidence. *Journal of Cognitive Neuroscience, 19*(7), 1175-1192. https://doi-org.ru.idm.oclc.org/10.1162/jocn.2007.19.7.1175

Holle, H., Obleser, J., Rueschemeyer, S. A., & Gunter, T. C. (2010). Integration of iconic gestures and speech in left superior temporal areas boosts speech comprehension under adverse listening conditions. *NeuroImage, 49*(1), 875-884. https://doi.org/10.1016/j.neuroimage.2009.08.058

Holler, J., Schubotz, L., Kelly, S., Hagoort, P., Schuetze, M., & Özyürek, A. (2014). Social eye gaze modulates processing of speech and co-speech gesture. *Cognition, 133*(3), 692-697. https://doi.org/10.1016/j.cognition.2014.08.008

Holler, J., Shovelton, H., & Beattie, G. (2009). Do iconic hand gestures really contribute to the communication of semantic information in a face-to-face context? *Journal of Nonverbal Behavior, 33*, 73-88. https://doi.org/10.1007/s10919-008-0063-9

Kayser, C., Petkov, C. I., Augath, M., & Logothetis, N. K. (2005). Integration of touch and sound in auditory cortex. *Neuron, 48*(2), 373-384. https://doi.org/10.1016/j.neuron.2005.09.018

Kelly, S.D., Healey, M., Özyürek, A., & Holler, J. (2015). The processing of speech, gesture, and action during language comprehension. *Psychonomic Bulletin & Review, 22*(2), 517-523. https://doi.org/10.3758/s13423-014-0681-7

Kelly, S. D., Kravitz, C., & Hopkins, M. (2004). Neural correlates of bimodal speech and gesture comprehension. *Brain and Language, 89*, 253-260. https://doi.org/10.1016/S0093-934X(03)00335-3

Kelly, S. D., Özyürek, A., & Maris, E. (2010). Two sides of the same coin: Speech and gesture mutually interact to enhance comprehension. *Psychological Science, 21*(2), 260-267. https://doi.org/10.1177/0956797609357327

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

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences, 4*(12), 463-470. http://dx.doi.org/10.1016/S1364-6613(00)01560-6

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual Review of Psychology, 62*, 621-647. https://doi.org/10.1146/annurev.psych.093008.131123

Lausberg, H., & Sloetjes, H. (2009). Coding gestural behavior with the NEUROGES-ELAN system. *Behavior Research Methods, 4*(3), 841-849. Retrieved from http://tla.mpi.nl/tools/tla-tools/elan/

Namy, L. L., Campbell, A. L., & Tomasello, M. (2004). The changing role of iconicity in non-verbal symbol learning: A U-shaped trajectory in the acquisition of arbitrary gestures. *Journal of Cognition and Development, 5*(1), 37-57. https://doi.org/10.1207/s15327647jcd0501_3

Namy, L. L., & Waxman, S. R. (1998). Words and gestures: Infants' interpretations of different forms of symbolic reference. *Child Development, 69*(2), 295-308. https://doi.org/10.2307/1132165

Obermeier, C., Dolk, T., & Gunter, T. C. (2012). The benefit of gestures during communication: Evidence from hearing and hearing-impaired individuals. *Cortex, 48*(7), 857-870. http://dx.doi.org/10.1016/j.cortex.2011.02.007

Obermeier, C., Holle, H., & Gunter, T. C. (2011). What iconic gesture fragments reveal about gesture-speech integration: When synchrony is lost, memory can help. *Journal of Cognitive Neuroscience, 23*(7), 1648-1663. https://doi-org.ru.idm.oclc.org/10.1162/jocn.2010.21498

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*(1), 97-113. https://doi.org/10.1016/0028-3932(71)90067-4

Özçalışkan, Ş., & Goldin-Meadow, S. (2011). Chapter 12. Is there an iconic gesture spurt at 26 months? In G. Stam & M. Ishino (Eds.), *Integrating Gestures: The Interdisciplinary Nature of Gesture* (pp. 163-174). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Özyürek, A., Willems, R. M., Kita, S., & Hagoort, P. (2007). On-line integration of semantic information from speech and gesture: Insights from event-related brain potentials. *Journal of Cognitive Neuroscience, 19*(4), 605-616. https://doi.org/10.1162/jocn.2007.19.4.605

Perniss, P., Thompson, R. L., & Vigliocco, G. (2010). Iconicity as a general property of language: evidence from spoken and signed languages. *Frontiers in Psychology, 1*, E227. https://doi.org/10.3389/fpsyg.2010.00227

Riseborough, M. G. (1981). Physiographic gestures as decoding facilitators: Three experiments exploring a neglected facet of communication. *Journal of Nonverbal Behavior, 5*(3), 172-183. https://doi.org/10.1007/BF00986134

Schaerlaekens, A. M. (1977). *De taalontwikkeling van het kind. Een oriëntatie in het Nederlandstalig onderzoek.* Groningen, Netherlands: Wolters-Noordhoff.

Schaerlaekens, A. M., Kohnstamm, G. A., Lejaegere, M., & Vries, A. K. (1999). *Streeflijst woordenschat voor zesjarigen: gebaseerd op nieuw onderzoek in Nederland en België.* Swets & Zeitlinger.

Schoechl, C. (2017). Children's integration of action gestures and verbs during online language comprehension: An event-related potentials study (Master's thesis). Retrieved from https://www.ru.nl/master/cns/journal/archive/volume-13-issue-1-2018 /print-edition/

Schulze, C., & Tomasello, M. (2015). 18-month-olds comprehend indirect communicative acts. *Cognition, 136*, 91-98. https://doi.org/10.1016/j.cognition.2014.11.036

Sekine, K., Schoechl, C., Mulder, K., Holler, J., Kelly, S., Furman, R., & Özyürek, A. (under review). Evidence for children's online integration of simultaneous information from speech and iconic gestures: An ERP study.

Sekine, K., Sowden, H., & Kita, S. (2015). The development of the ability to semantically integrate information in speech and iconic gesture in comprehension. *Cognitive Science, 39*, 1855-1880. https://doi.org/10.1111/cogs.12221

Sekine, K., & Özyürek, A. (in prep). How do gestures contribute to comprehension of degraded speech in children?

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science, 270*(5234), 303-304. https://doi.org/10.1126/science.270.5234.303

Stanfield, C., Williamson, R., & Özçalişkan, Ş. (2014). How early do children understand gesture-speech combinations with iconic gestures? *Journal of Child Language, 41*(2), 462-471. https://doi.org/10.1017/S0305000913000019

Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses.* Cambridge, MA: The MIT Press.

Willems, R. M., Özyürek, A., & Hagoort, P. (2007). When language meets action: The neural integration of gesture and speech. *Cerebral Cortex, 17*(10), 2322-2333. https://doi.org/10.1093/cercor/bhl141

Wu, Y. C., & Coulson, S. (2005). Meaningful gestures: Electrophysiological indices of iconic gesture comprehension. *Psychophysiology, 42*, 654-667. https://doi.org/10.1111/j.1469-8986.2005.00356.x

Wu, Y. C., & Coulson, S. (2007a). How iconic gestures enhance communication: An ERP study. *Brain and Language, 101*, 234-245. https://doi.org/10.1016/j.bandl.2006.12.003

Wu, Y. C., & Coulson, S. (2007b). Iconic gestures prime related concepts: An ERP study. *Psychonomic Bulletin & Review, 14*(1), 57-63. https://doi.org/10.3758/BF03194028

Appendix A. The participant information form

Radboud Universiteit

CLS | Centre for Language Studies

Deelnemersinformatie

Hartelijk dank voor uw deelname aan onze studie. We zouden graag wat informatie van u willen hebben voor onze database. Alle informatie en data worden anoniem verwerkt en individuen kunnen niet worden geïdentificeerd. Alle data wordt beschermd en opgeslagen volgens ethische richtlijnen van de Radboud Universiteit en alleen de onderzoekers hebben hier toegang toe.

Uw geboortedatum (Dag/Maand/Jaar)

____

De geboortedatum van uw kind (Dag/Maand/Jaar)

____

Als uw kind naast Nederlands nog andere talen spreekt, noem hier dan alle talen die uw kind kan spreken.

____

Hoe oud was uw kind wanneer hij/zij deze ta(a)l(en) heeft verworven?

____

____

Welk soort onderwijs krijgt uw kind?

- ○ Normaal onderwijs
- ○ Thuisonderwijs
- ○ Anders, namelijk ____

Wat is uw hoogst genoten onderwijs?

- ○ Geen diploma
- ○ LBO/VMBO
- ○ Havo/VWO
- ○ MBO
- ○ HBO
- ○ WO
- ○ (Post)-Doctoraat

Uw contactinformatie (e-mailadres of mobiel)

Als u interesse heeft in de resultaten van dit onderzoek en/of graag meer informatie wilt ontvangen over deelname in andere studies, vul dan hier uw contactinformatie in. Geef aan welke opties toepasbaar zijn.

○ Ja, stuur mij informatie over de resultaten van deze studie

○ Ja, informeer mij over deelname in andere studies

Uw naam: ____

Contactinformatie: ____

Bedankt voor uw medewerking

Kazuki Sekine en Suzanne de Jong

Appendix B. The Edinburgh Handedness Inventory

Radboud Universiteit

CLS | Centre for Language Studies

De Edinburgh Handvoorkeur Vragenlijst

Geef a.u.b. in onderstaande activiteiten aan welke handvoorkeur uw kind heeft door de betreffende voorkeur te omcirkelen. Kies voor "beide" als uw kind geen voorkeur heeft.

1. Met welke hand schrijft uw kind over het algemeen?	links	beide	rechts
2. Met welke hand tekent uw kind over het algemeen?	links	beide	rechts
3. Welke hand zou uw kind gebruiken bij het raken van een doel met een bal?	links	beide	rechts
4. Met welke hand houdt uw kind zijn/haar tandenborstel vast?	links	beide	rechts
5. Welke hand gebruikt uw kind bij het gebruiken van een mes (wanneer hij/zij geen vork vast heeft)?	links	beide	rechts
6. Wanneer uw kind een lucifer afstrijkt, in welke hand houdt hij/zij de lucifer dan?	links	beide	rechts
7. Wanneer uw kind een doos opent, welke hand houdt dan de deksel vast?	links	beide	rechts
8. Welke hand houdt de lepel vast bij het eten van soep?	links	beide	rechts
9. Met welke hand gebruikt uw kind een schaar?	links	beide	rechts
10. Wanneer uw kind een bezem gebruikt, welke hand is dan de bovenste op de bezem?	links	beide	rechts