Radboud Universiteit

IN·DEI·NOMINE·FELICITER

Radboud University Nijmegen

The effect of eyebrow movements on question identification

Thesis BSc General Linguistics

Vere Jacobs
1083289
13/6/22
Communication of Social Interaction
Naomi Nota and Judith Holler

Content

Abstract

In conversation, many speakers use social actions. It is important for the listener to quickly recognize the social action of the speaker in order to understand what the speaker is trying to convey. Questions frequently occur in conversations, and require immediate action from the listener. Several studies have demonstrated that eyebrow frowns and raises facilitate question identification. However, few studies have investigated the effect of eyebrow movements as they naturally occur in communication on accuracy and response times. Therefore, the current study investigated the effect of eyebrow movements, which are based on natural intensities, on question identification. To investigate this, we did an online experiment. In this experiment, participants did a forced choice two-alternative behavioural task, where participants had to judge whether the utterance pronounced was a question or a statement. The utterances were pronounced by an avatar, whose behaviour was based on the natural behaviour shown in a multimodal Dutch face-to-face corpus. The participants' task was to identify whether the utterance of the avatar was a question or a response as quickly and accurately as possible. The aim of this experiment was to determine whether specific visual signals, such as eyebrow frowns or raises, play a role in the perception of questions. Gorilla, an online testing platform, was used to collect data. Results showed similar accuracy scores for the presence (frowns, raises) and absence of eyebrow movement. Additionally, participants were significantly faster in identifying questions when an utterance was paired with a frown in comparison to a raise or the absence of eyebrow movement. Thus, the presence of frowns results in a quicker identification of questions.

1. Introduction

Language is multimodal: messages can be conveyed using visual and auditory signals (Holler & Levinson, 2019). In conversations, speakers use many visual and auditory signals to communicate information to the listener, that support the understanding of utterances (Perniss, 2018).

It is important for the listener to quickly recognize the social action of the speaker in order to understand what the speaker is trying to convey (Gísladóttir et al., 2012). A common example of a social action are questions, which are fundamental in conversations (Enfield et al., 2010). Social actions are useful to determine the goal of an utterance (Trujillo & Holler, 2021). Questions differ from other social actions, such as statements, as they require immediate action from the listener (Kendrick & Torreira, 2015).

Languages have different approaches to mark questions. One approach is by using different question-structures, for example polar questions (e.g. '*Is hij twintig jaar oud?*', 'Is he twenty years old?'), content questions (e.g. '*Wat is jouw favoriete film?*', 'What is your favourite movie?') and alternative questions (e.g. '*Welke soort films kijk jij liever: horror, of drama?*', 'What kind of movies would you rather watch: horror or drama?'). In Dutch, polar questions are the most frequent question type, as found by Englert (2010). Dutch polar questions (e.g. '*Heeft hij een pet op?'*, 'Is he wearing a hat?') usually contain the inversion of subjects and verbs and do not make use of an auxiliary verb (Borràs-Comes et al., 2014). Another strategy to mark questions is by prosody. Different question types are related to different voice and pitch, for example in wh-questions (e.g. 'What are you going to do today?') and rhetorical questions (e.g. 'Can I ask a question?') (House, 2002). As stated by Englert (2010), intonation also plays a crucial role in questions in Dutch. In particular, it was found that rising intonation is crucial for interpreting declarative sentences (e.g. "*Ze is jouw dochter?',* 'She is your daughter?') as questions. This was also found by Borràs-Comes et al. (2014), who stated that rising intonation is used for questions in general. Thus, there are several strategies to mark questions relating to the prosody and word order of a language.

Previous studies have shown an association between questions and visual signals of the face, henceforth referred to as facial signals (e.g. Bavelas et al., 2014; Chovil, 1991; Granström et al., 2002; Krahmer & Swerts, 2007; Srinivasan & Massaro, 2003). Chovil (1991) looked at which facial signals were associated with questions. To research this, videos of participants discussing topics with each other were recorded. The facial signals that were used by participants when discussing the topics were analysed. For example, eyebrow movement, together with the twisting of the mouth to one side or the pulling back of one corner of the mouth occurred when a participant was remembering or thinking about an incident. Speakers, as well as listeners, were found to utilize different facial signals at varying times. Results showed that facial signals are narrowly related to social actions. Whereas Chovil (1991) looked at facial signals that related to a variety of social actions, Srinivasan and Massaro (2003) investigated the facial signals that were associated specifically with statements and questions. It was found that questions co-occurred with eyebrow raise as well as a head tilt. Therefore, it seems that specific facial signals correspond with specific social actions. In addition to the evidence that facial signals were associated with questions, there are also other studies showing that visual signals play a role in the communication of specific social actions (Chovil, 1991; Ekman, 1979; Flecha-García, 2010; Nota et al., 2021). These studies have found that eyebrow raises and frowns are associated with questioning. Chovil (1991) specifically looked how facial signals accompanied speech by analysing how these signals conveyed information in natural dyadic conversations. It was found that eyebrow frowns and raises often accompanied questions. A study that supports these findings is a study by Nota et al. (2021), in which the association of several facial signals with social actions were investigated in a dyadic Dutch face-to-face conversational corpus. Specifically, they studied facial signals that occurred during questions and responses. Results showed that questions and responses were characterized by different distributions of facial signals. For example, it was found that eyebrow movements often occurred in questions. Importantly, eyebrow frowns were found to be among the strongest visual signals to mark questions. From these previous studies, which looked at facial signals that occurred with specific social actions, it seems that eyebrow raises and frowns are a strategy to mark questions.

While the above-mentioned studies addressed which facial signals co-occurred during the production of questions, other studies investigated the effect that visual signals had on language perception (Bavelas et al., 2014; Chovil, 1991; Granström et al., 2002; Srinivasan & Massaro, 2003). An example of a study that investigated this was a study by Granström et al. (2002). In this study, a talking head was used, which is an animated face that is based on the visual and auditory cues that are used naturally by humans. This talking head conveyed negative expressions or affirmative expressions. Participants had to listen to a conversation between a talking head and a human. They were asked to determine whether the avatar accepts the human utterance or is uncertain about it. This was done by using response values.

The results showed that subjects are sensitive to both acoustic and visual parameters of the talking head in conversation. Although this study showed that participants are perceptive to visual and auditory information using a talking head, it has not investigated whether participants react to these parameters separately. This was done by Srinivasan and Massaro (2003), who used talking heads to investigate enhanced eyebrow signals and to determine whether visual signals had a bigger influence on question identification as opposed to auditory signals by measuring responses. This study looked at artificially enhanced eyebrow movements and focused on distinguishing the visual signals from the auditory signals. Despite enhancing the visual signals, the study revealed that the visual signals did not contribute to question identification more than auditory signals. Different methods are used in the studies previously mentioned to investigate which facial signals co-occurred in questions. However, response times of listeners were never researched. This is a crucial part of rapid turn-taking during conversation, as listeners are expected to respond to the speaker as quickly as possible to avoid a gap. Facial signals may help anticipate a speaker's social action and can therefore help identify the speaker's social action as quickly as possible in order to provide a rapid response. Therefore, response times of listeners require more investigation.

Whereas the studies mentioned previously serve as examples of studies using talking heads, there may be methods that are more representative of the way humans communicate face-to-face. In contrast to the methods that the above-mentioned studies have used, Nota et al. (in prep) used avatars to investigate whether eyebrow movements had an effect on question identification. It may be useful to conduct studies with avatars, as avatars display behaviour that is more representative of natural communication, and are therefore more human-like in comparison to talking heads. The avatars used in this study are also more modern, and therefore more naturalistic virtual agents in comparison to the talking head that was used by Granström et al. (2002). Importantly, Nota et al. (in prep) observed higher accuracy scores and faster response times on trials where the avatars performed a question with an eyebrow frown compared to no eyebrow movement. This suggests that the presence of eyebrow movements facilitate detection of questions. However, this study looked at eyebrow movements with fixed intensities. Thus, the signals could potentially be enhanced when actual intensities of the eyebrow movements as they occurred in the original conversations were lower. Therefore, it is currently unknown whether eyebrow movements based on natural intensities have an effect on question identification. This would be informative to gain more insights into the role that specific facial signals play in the perception of fundamental social actions in conversation, like questions. Therefore, the current study investigated the effect of eyebrow movements that are based on natural intensities on question identification, with stimuli based on the facial and speech behaviour of that participants displayed, which was derived from a corpus, to determine whether specific visual signals such as eyebrow frowns or raises play a role on the perception of questions. Because the stimuli are based on natural behaviour, it can be a step towards representative language use.

2. Research Question and Hypothesis

The research question was: “Do eyebrow movements based on natural intensities facilitate question identification?”. Previous research has shown a facilitatory effect of eyebrow frowns on question identification (Nota et al., in prep). We therefore hypothesized that eyebrow movements with natural intensities would have a similar facilitatory effect on the participants’ responses, by resulting in higher accuracy scores and faster response times when identifying questions, in comparison to the accuracy scores and response times found by studies that used eyebrow movements with fixed intensities. Since no facilitation effect was observed for eyebrow raises in Nota et al. (2021, in prep), and because Nota et al. (2021) observed that eyebrow movements were among the strongest visual markers for questions in comparison to other visual signals, we further hypothesized that the effect of higher accuracy scores and faster response times would be most visible for eyebrow frowns.

This study supports and expands previous research on the importance of visual signals in question identification. It also gives insights into the importance of naturally enhanced eyebrow raises and frowns, in question identification. The findings of this study may be especially relevant for the development of educational video resources as well as the development of social robots.

3. Method

3.1 Participants

The study was approved by the Ethics Committee of the Social Sciences department of the Radboud University Nijmegen (ethic approval code ECSW 2018-135).

Participants were volunteers recruited amongst acquaintances (*n = 15*) and SONA (*n* = 1), an online platform to recruit participants. To take part in the experiment, participants had to be between 18 and 45 years old and had to have no motoric, hearing or language issues. The participant’s native language had to be Dutch. Before taking part in the study, participants had to fill in a questionnaire to decide whether they fit the criteria of the study. Of the 23 participants that took part in the experiment, five participants were excluded from the analysis, as they did not fit the requirements of the study. Furthermore, two participants failed to complete the experiment, and were therefore excluded from the analysis. This resulted in a final sample of 16 participants (15 female, 1 male). The participants' age ranged from 21 to 29 years old (M = 23.1, SD = 1.4).

3.2 Stimuli

3.2.1 Corpus coding

The experiment makes use of avatars, based on the work from Nota et al. (in prep). The avatars performed speech and visual behaviour that was based on speakers in a Dutch dyadic corpus of spontaneous conversations (CoAct corpus, ERC project #773079 led by JH). The visual and auditory behaviour that participants showed in the videos were coded by making use of ELAN (5.5; Sloetjes & Wittenburg, 2008).

First, questions (and responses, which are not part of the current study) were annotated by two Dutch native speakers. Any visual physical signals, context, phrase, tone and behaviour of the listener were taken into account. All spoken responses to a question, including standard interjections such as "uh" or "hmm" were transcribed. Nonverbal sounds, such as laughter and sighs, were not included in the transcription. For more information on the creation of the corpus, see Nota et al. (2021). For 12% of all data (4 dyads, all parts), interrater reliability was assessed between the two coders with raw agreement and a modified Cohen's kappa (Cohen, 1960; Landis & Koch, 1977) using EasyDIAg (Holle & Rein, 2015). EasyDIAg is an open-source tool that has become the standard for calculating a modified Cohen's kappa. It is determined by the degree of temporal overlap between transcriptions, categorization of values and behaviour segmentation. A standard overlap criterion of 60% was used. Calculating the interrater reliability between the two coders resulted in a raw agreement of 75% and k = 0.74 for questions. This indicated substantial agreement.

Second, social actions were annotated. The questions that were coded were divided into sub-categories of questions. The current study focuses on Information Requests, which were among five other categories: Understanding Checks, Self-Directed questions, Stance or Sentiment questions, Other-initiated repair questions and Active participation questions, which we are not using in this study. For more information on the coding process, see Trujillo and Holler (2021). Using the same procedure as for the previous annotations, the interrater reliability was calculated for the social action categories. This was done for 10% of the total amount of question annotations. Calculating this interrater reliability resulted in a raw agreement of 76% and k = 0.70, indicating substantial agreement once again.

Third, facial signals were also annotated in ELAN (5.5; Sloetjes & Wittenburg, 2008), based on the frontal view videos from the CoAct corpus. Facial signals were only annotated if they carried some sort of communicative meaning that was related to the questions. Movements like swallowing, inhaling, laughter, or articulation were left out of the transcription. Facial signals consisted of eyebrow movements such as frowns, raises, eye widenings, blinks, squints (among other signals). The signals were annotated from the first hint of movement until the speaker returned to neutral position. Because visual signals and speech are formed in genuine conversation, visual behaviour can start prior to or extend past the actual verbal message. Therefore, facial signals were coded from beginning to end, except when speech not constituting part of the question or response in question began, or when laughter (without voice) occurred. In those instances, the annotation continued until the first evidence of speech unrelated to the questions or laughter, or began after the last evidence. For a more concise view on how accurate coders were and the interrater reliability score for each unique facial signal, see Nota et al. (2021). The paired comparisons of the facial signals showed an average raw agreement of 76% (*min* = 70%, *max* = 82%) and an average kappa of 0.96 (*min* = 0.04, *max* = 0.97). This indicated an almost perfect agreement for all other facial signals.

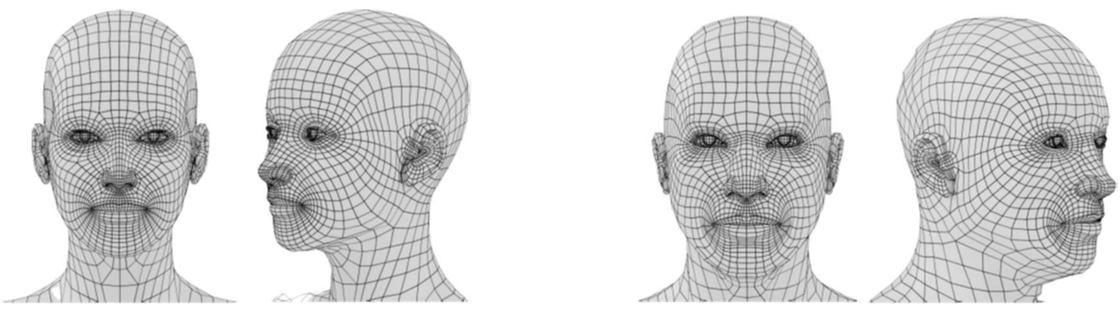
For the present experiment, we looked specifically at the eyebrow raises and frowns that the participants used in the corpus. The eyebrow movements from the avatars were manipulated to match these natural intensities. The movement intensities of the eyebrow movements in the CoAct corpus were annotated by two native speakers of Dutch. They rated the intensities on a scale of one to five, one being barely any movement and five being an extremely high movement intensity. The first rater coded 100% of the data, and trained the second rater on 10%. The second rater then coded 90% of the data independently, to enable computation of the reliability between raters on the intensity scores. The same procedure was used to calculate interrater reliability between coders as for the questions, social actions, and facial signals. The paired comparisons showed an average raw agreement of 81%, and k = 0.76, indicating an agreement for the intensities.

3.2.2 Avatars

Using ffmpeg (4.3.1;Tomar, 2006) and custom scripts in Bash (3.2.51-1) and Ruby, the related video clips of the stimuli were extracted from the corpus by converting the mp4 recordings to mp4 iframes, trimming them, and exporting the audio (wav) (2.3.0). The avatars were made in Blender (version 2.83.12; Community, 2018) with the open-source MB Lab (version 1.7.6, Bastoni, 2021) plug-in for parametric 3D modeling of realistic humanoid beings. In terms of skin, hair, and clothing, the avatar was designed to look like the original corpus speaker. The action units, gaze, and position of the corpus video clips obtained by OpenFace were used to detect faces (version 2.2.0; Baltrusaitis et al., 2018). The FACSvatar-Blender plug-in (version 0.4.0) was used to import this into Blender using the open-source add-on FACSvatar (version 0.3.4-alpha; van der Struijk et al., 2018), which is a framework for real-time facial animation using the Facial Action Coding System (FACS) (Ekman & Friesen, 1978). The avatars also demonstrated gaze and blinks, which were inserted manually in addition to the facial signals in the stimulus. The timing of all facial signals was derived from the corpus's facial signal transcriptions, with the exception of gaze and blinks (for more details, see Nota et al., 2021). The standard intensities obtained by FACSvatar were used to determine the intensities of gaze and blinks. The intensities were further modified to ensure that they were suitable for the avatar models and were consistent throughout situations. The intensities were also matched amongst female and male avatar models to make them look similar (Figure 1).

Lastly, using a unique Python (version 3.7) script with phoneme annotations, the avatars' lip motions were coordinated with the audio files. Using the automatic segmentation program WebMAUS Basic (version 3.3; Kisler et al., 2017) and manual transcription, the precise beginnings and endings of the phonemes were segmented in Praat (5.1, Broersma & Weenink, 2021). These transcriptions were imported into the ELAN (5.5; Sloetjes & Wittenburg, 2008) file and generated text files.

Figure 1. Female (left) and male (right) avatar models (Nota et al., In prep).



3.2.3 Utterances
The avatar performed 40 polar questions and 40 content questions. There were two conditions in the experiment. In the presence-condition, the questions contained eyebrow movement. In the absence-condition, the questions did not contain eyebrow movement. See Figure 2 for an

example of these conditions. The polar and content questions contained eyebrow raises and eyebrow frowns in the presence-condition. All eyebrow movements were removed from the avatar to represent the absence-condition.

To distract from the main research question, fillers were added. These consisted of 80 statements to balance out the requests for information and to make sure the participants do not identify our research question. Twenty statements had eyebrow raises, twenty statements had frowns, and forty statements did not have any eyebrow movement.

The other fillers were statements and questions that contained squints and eye widenings. These visual signals were assigned a random intensity to represent the natural variation in movement intensities. The ranges of these intensities of these fillers are shown in Table 1. Blinks were added throughout.

3.2.4 Facial signals intensities

The eyebrow movements were given intensity scores as they occurred naturally in the CoAct corpus. The avatars were designed to match these intensity scores. Table 1 shows the facial signal intensities that the avatar models adhered to, for eyebrow signals as well as the fillers. The intensities are shown in Figure 2 and 3.

Table 1. Facial signal intensities per model and component. This table shows the minimum and maximum facial shapes that the avatars adhered to. The minimum and maximum can be viewed as opposites, which is why *eyeClosed_max* refers to closed eyes and *eyeClosed_min* refers to widened eyes. Some facial shapes, such as *eyeSquint_max*, *eyeClosed_min* and *eyeClosedPressure_min* did not have a need for an opposite intensity, which is why it has been left out of the table.

Facial signal component	**Female avatar**	**Male avatar**
Eyebrow raise		
browOutVert(LR)_max	.90	.90
browSqueeze(LR)_min	1.00	1.00
browsMidVert_max	1.00	1.00
Eyebrow frown		
browOutVert(LR)_min	.90	.90
browSqueeze(LR)_max	1.00	1.00
browsMidVert_min	0.04	0.08
Squint		
eyeSquint(LR)_max	.50	.325
Eye widening		
eyeClosed(LR)_min	.45	.60
eyeClosedPressure(LR)_min	.15	.20

The proportions from Table 2 were divided into five categories to match the five intensities that the naturally occurring eyebrow movements had in the corpus. This was done by splitting the proportions from Table 2 into five proportions to match the five intensities assigned to the eyebrow movements made by participants in the corpus. This resulted in the proportions for raises and frowns, as shown in Table 3. Eyes widenings and squints were assigned a random intensity, as they were fillers. In order to keep these facial signals as natural as possible, the percentages from Table 2 were also split into five proportions for these signals.

Table 2. Facial signal intensities per component, divided into five proportions.

Intensity					
	1	2	3	4	5
Signal					
Raise	0.16	0.33	0.52	0.74	1.00
Frown	0.16	0.33	0.52	0.74	1.00
Eyeswidening	0.20	0.40	0.60	0.80	1.00
Squint	0.20	0.40	0.60	0.80	1.00

Figure 2. The intensity scores from lowest to highest for frowns for the male avatar.

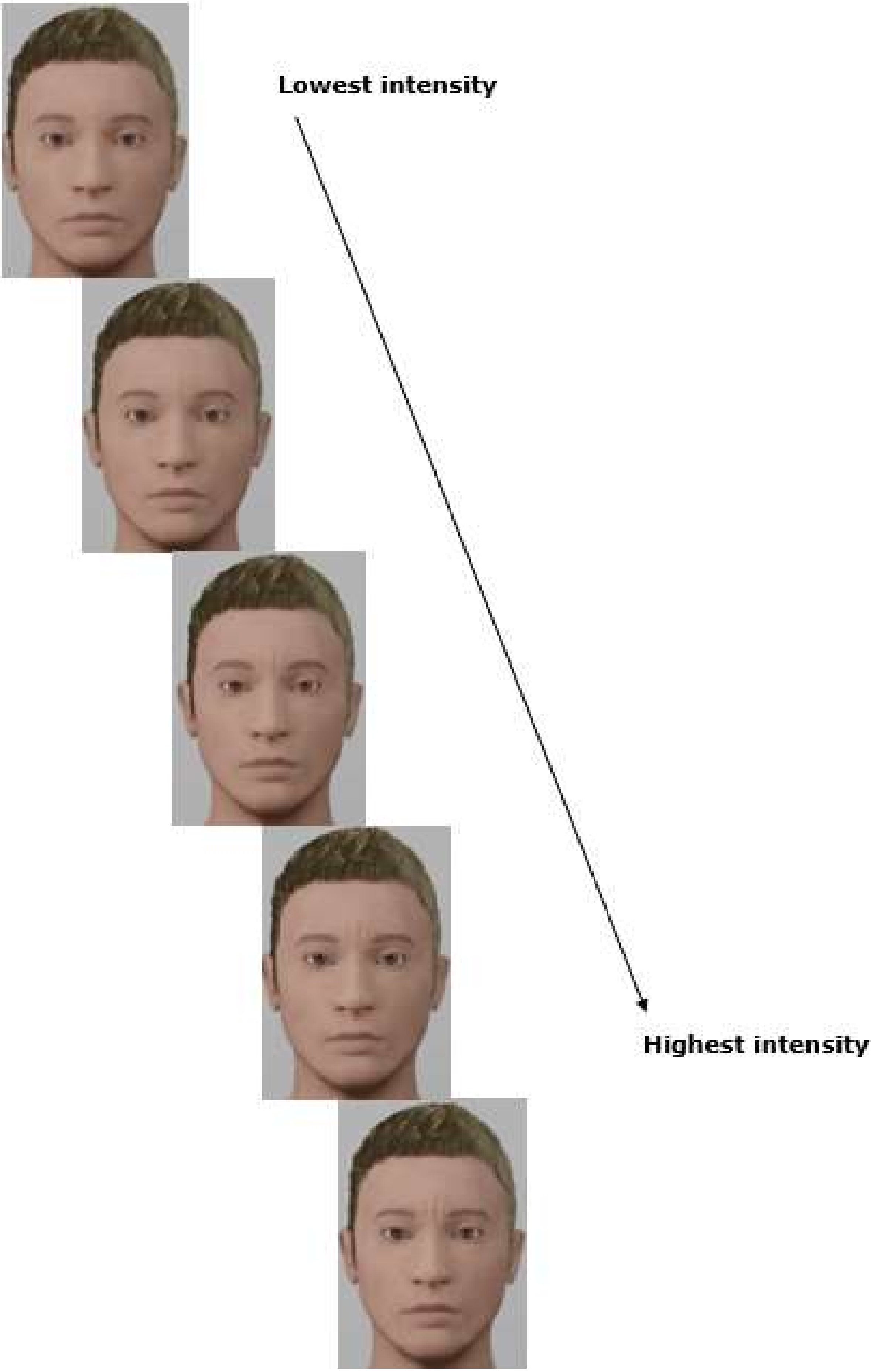
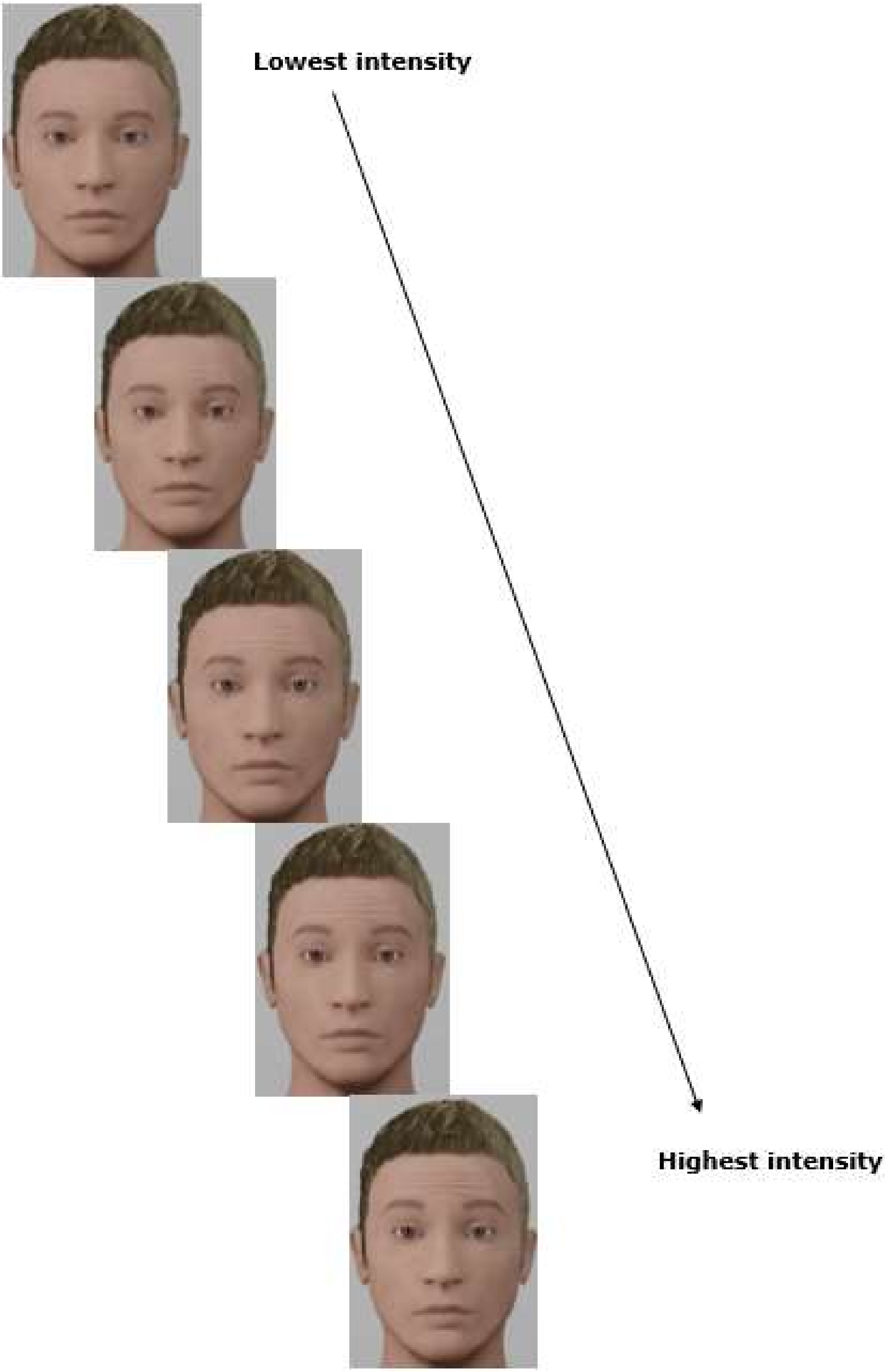

Lowest intensity
Highest intensity

Figure 3. The intensity scores from lowest to highest for raises for the male avatar.



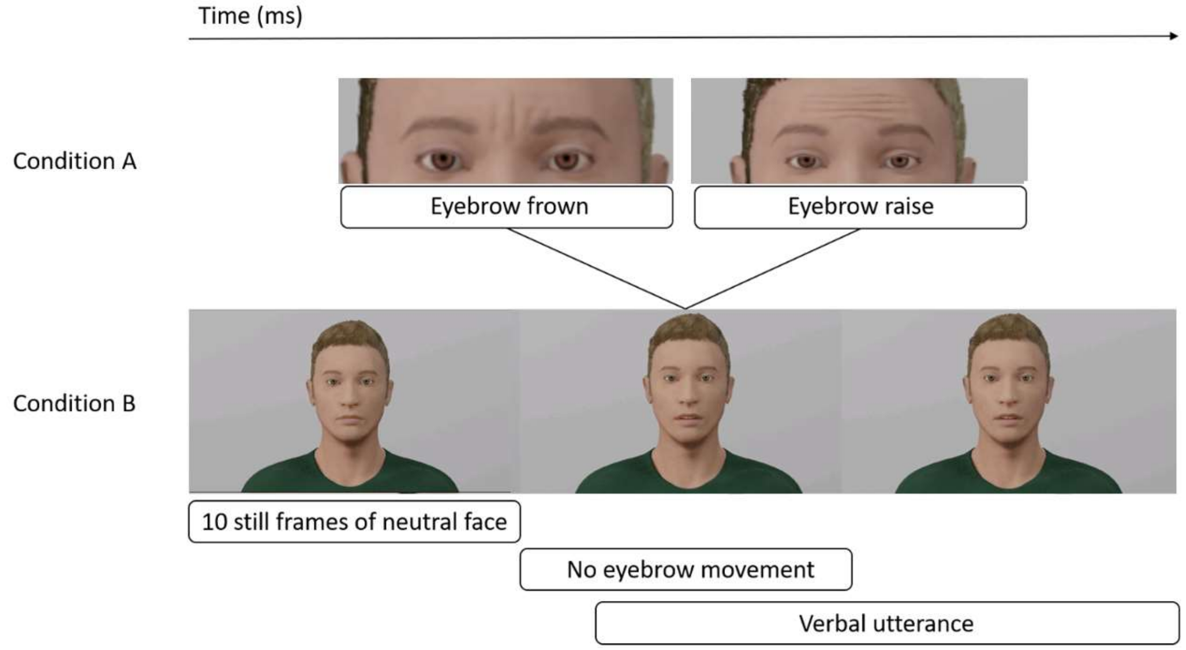
The timing of all facial signals, except for gaze and blinks, were based on the movements as they occurred naturally in the corpus. The intensities of these facial signals were consistent between conditions and were matched between the female and male versions of the avatar models. This was done to make them look as similar as possible.

3.3 Design

All items were randomized. The target stimuli were presented in four blocks, of which the order was randomized. The order of the blocks was also randomized. Breaks were added between each block.

The participants were shown one video at a time (for an example, see Figure 4). This video showed the avatar, and participants heard the auditory stimuli. Participants then had to judge whether the utterances are statements or questions. They were instructed to judge the utterances as fast and as accurately as possible.

Figure 4. Example of a video clip. The video clip contains an utterance with eyebrow movement (condition A) and without eyebrow movement (condition B). The duration of the video clip (in ms) is the same for both conditions.



Every video was consistent in the way they were shown to participants. The video clip started off with a neutral face, meaning without any facial signals. Then, the verbal utterance and corresponding facial signals began. These facial signals were based on ELAN coding, as previously described. The facial signals appeared gradually, to avoid sudden and unnatural onset. Questions in the absence-condition and statements without eyebrow movement contained eyebrow movements that were kept consistent at the original onset movement level as they appeared in the corpus.

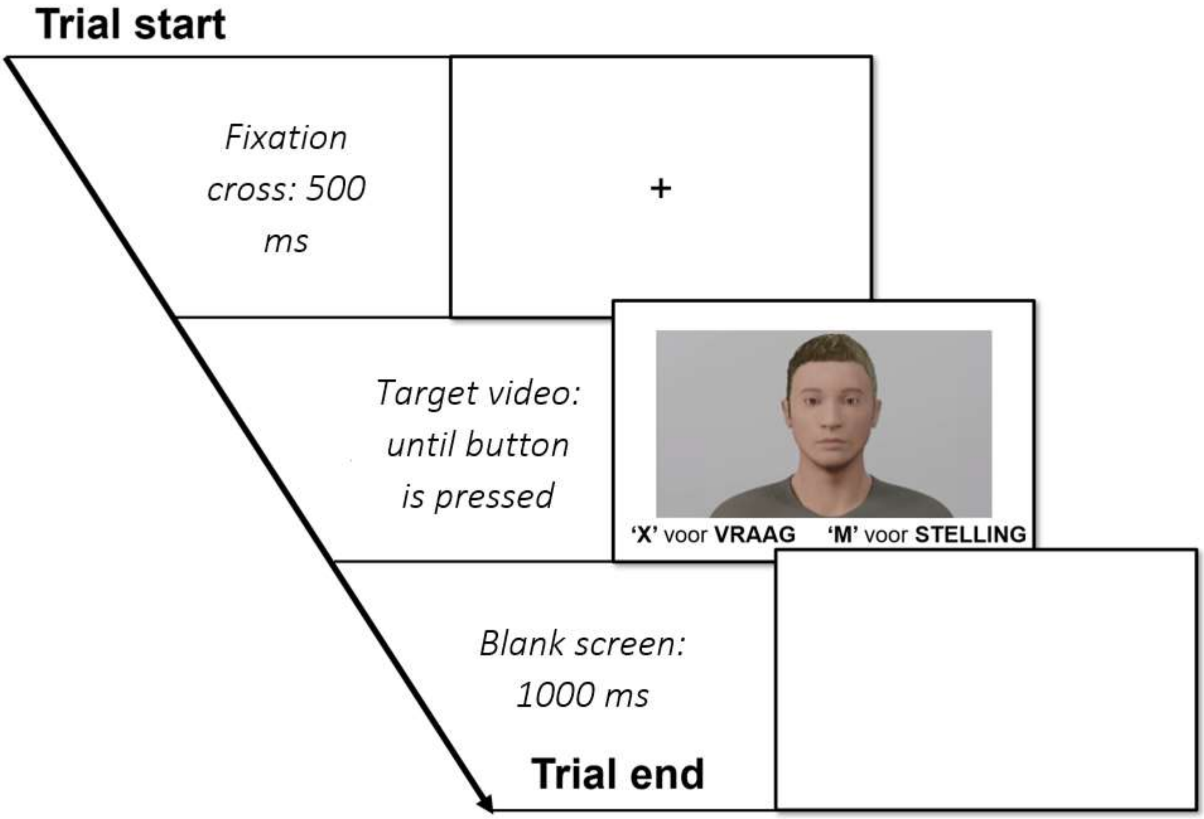
3.4 Procedure

Before taking part in the experiment, participants had to fill in several questionnaires about general demographics and language background. The experiment was performed in Gorilla, an online platform to perform experiments (Anwyl-Irvine et al., 2020). Participants were instructed to use the most recent version of Google Chrome and were required to have a Windows operating system. They were also asked to use earphones or headphones that were wired.

Prior to starting the experiment, participants were instructed to take part in the experiment in a quiet room to make sure there were no distractions. Then, a sound was played to test the volume and to ensure that videos would automatically start playing. Following this test, the experiment launched into a full-screen modus. The goal of the experiment was to measure the response times and accuracy scores per participant for trial. Therefore, participants had to decide whether the utterance of the avatar was a question or a statement as quickly and accurately as possible. They did this by pressing the key ‘X’ when they thought the utterance was a question, and ‘M’ when they thought the utterance was a statement.

Participants first judged 16 practice items, followed by 180 trials that consisted of the clips of the avatars. The participants did not receive feedback on their accuracy scores for the practice items. Before each trial, a fixation cross appeared in the middle of the screen. This cross appeared for 500ms, after which the video clip of the avatar was played. As soon as participants pressed the button, a blank screen was shown for 1000ms. In the event that no button was pressed, the entire video clip played. Then, the next trial began. Figure 5 shows an example of the trial.

Figure 5. Example of a trial.



After completing the experiment, participants filled in the Empathy Quotient questionnaire (Baron-Cohen & Wheelwright, 2004), in which participants indicated whether they agree with statements relating to empathy, and the Actions and Feelings Questionnaire (van der Meer et al., 2021; Williams & Cameron, 2017), in which participants indicated whether they agree with statements relating to self-awareness of actions feelings-related actions. Then, participants were asked to evaluate the avatars on likability and ease of understanding and humanness. Lastly, participants filled in a questionnaire that assesses awareness of the aim of the experiment. Upon completion of the experiment and filling in the questionnaires, participants were thanked for their participation and were informed about the aim of the experiment.

3.4 Analysis

To investigate the effect of the presence of eyebrow movement on accuracy scores and response times, we first checked for normality of the residuals for both the data for accuracy scores and the data for response times, using a Shapiro-Wilck and Levene's Test. If the data was normally distributed, we performed a one-way ANOVA using *R* and *R studio* (version 1.4.1717; RStudio Team, 2021) with eyebrow movement (frowns and raises) and the absence of eyebrow movement as independent variables. The accuracy scores and response times were the dependent variables. In the analysis, we specifically looked at the effects of eyebrow frowns and raises in questions and compared them to the absence of such visual signals, and compared the means of these two groups. Then, we were able to conclude whether there was an effect of the presence of eyebrow movement on response times and accuracy scores.

4. Results

4.1 Accuracy

The Shapiro-Wilk normality test indicated that the accuracy scores were normally distributed ($W = 0.45$, $p = <.001$). Levene's Test showed that the variances for accuracy were equal. Therefore, we performed a one-way ANOVA. Results showed an insignificant effect of presence of eyebrow movement on accuracy ($F(2,1176) = 0.49$, $p = 0.62$). This means that participants were not better at identifying questions with eyebrow movements (frowns $M = 0.86$, raise $M = 0.85$) compared to questions without eyebrow movements ($M = 0.84$) (see Figure 6 and 7).

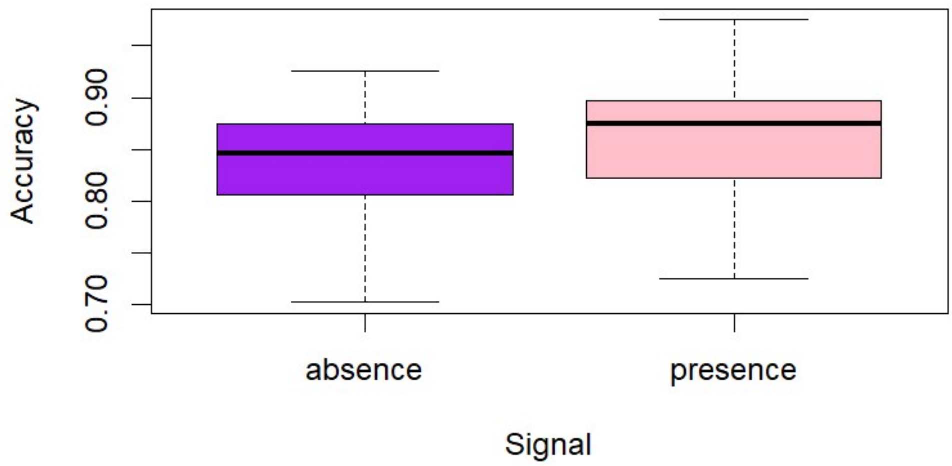


Figure 6. Boxplots showing the distribution of accuracy for absence for eyebrow movement and presence of eyebrow movement. The x-axis displays the absence or presence of eyebrow movement, the y-axis shows the accuracy scores of the participants. The black bar indicates the median of the accuracy scores. The lower box indicates the 25th percentile, the upper box the 75th. The lower whisker indicates the minimum accuracy score, and the upper whisker indicates the maximum accuracy score.

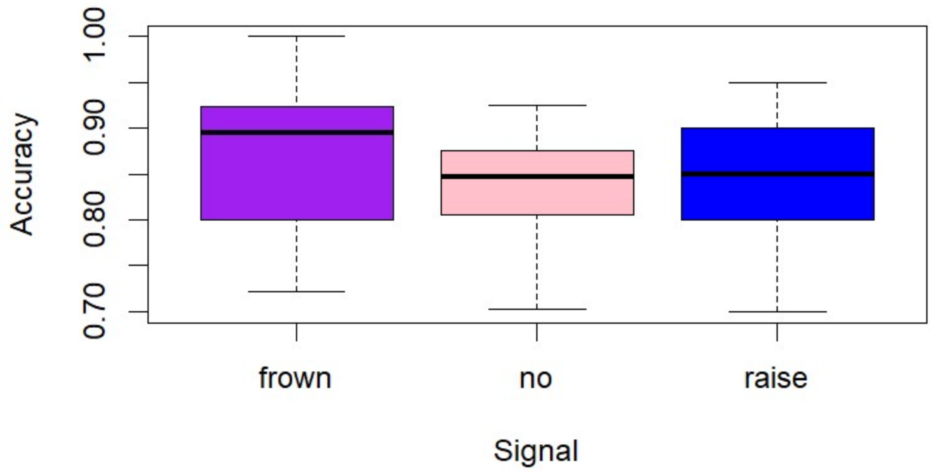


Figure 7. Boxplots showing the distribution of accuracy for frowns, raises and the absence of eyebrow movement ('no'). The x-axis displays the eyebrow signal type, the y-axis shows the accuracy scores of the participants. The black bar indicates the median of the accuracy scores. The lower box indicates the 25th percentile, the upper box the 75th. The lower whisker indicates the minimum accuracy score, and the upper whisker indicates the maximum accuracy score.

4.2 Response Times

The Shapiro-Wilk normality test indicated that the response times were normally distributed, W = 0.99, p = < 0.001. However, Levene's Test showed that the variances for accuracy were not equal (F(2,992) = 3.75, p = 0.02). Therefore, we performed a Welch test. The Welch test showed a significant result (F(2, 623.07) = 12.57, p = < .001). A post-hoc test showed that responses to questions accompanied by the presence of frowns (M = 1005, SD = 511) differed significantly (p = < .001) from responses to questions accompanied by the absence of eyebrow movements (M = 1121, SD = 532). It also differed significantly (p = < .001) from the responses to questions accompanied by the presence of raises (M = 1254, SD = 589). The responses to questions accompanied by the absence of eyebrow movements also differed significantly from the responses to questions accompanied by the presence of raises (p = <.001).

These results indicate that participants were significantly faster to recognize questions when these questions contained frowns in comparison to the absence of eyebrow movement or raises. Participants were also faster in recognizing questions when these did not contain any eyebrow movement in comparison to when they contained raises (see Figure 8 and 9).

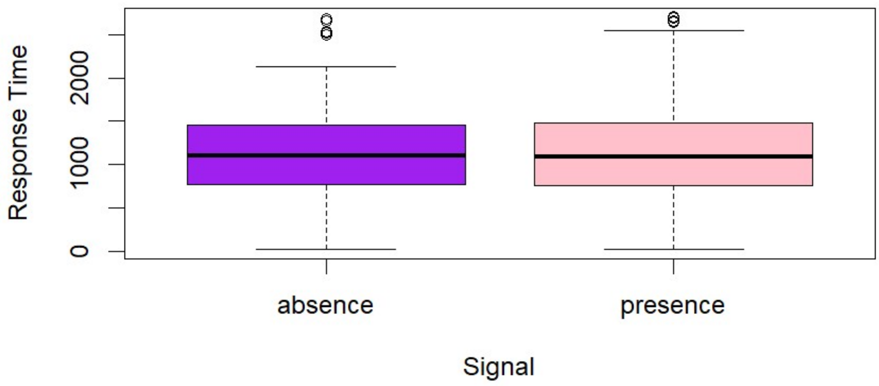


Figure 8. Boxplots showing the distribution of response times for absence and presence of eyebrow movement. The x-axis displays the absence or presence of eyebrow movement, the y-axis displays the response times of the participants. The black bar indicates the median of the response times. The lower box indicates the 25th percentile, the upper box the 75th. The lower whisker indicates the minimum response time, and the upper whisker indicates the maximum response time. The dots represent the outliers.

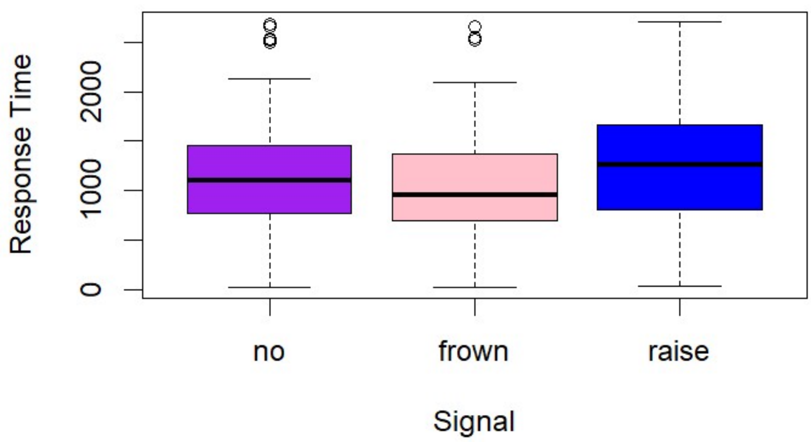


Figure 9. Boxplots showing the distribution of response times for absence for eyebrow movement, presence of eyebrow movement with frown and presence of eyebrow movement with raise. The x-axis displays the type of eyebrow movement, the y-axis displays the response times of the participants. The black bar indicates the median of the response times. The lower box indicates the 25th percentile, the upper box the 75th. The lower whisker indicates the minimum response time, and the upper whisker indicates the maximum response time. The dots represent the outliers.

5. Discussion

In the present study, we investigated whether eyebrow movements (frown or raise) facilitate question identification in comparison to the absence of eyebrow movements. To investigate this, we used avatars. The visual and auditory behaviour of these avatars was based on a Dutch dyadic conversational corpus. These avatars were used in a forced choice two-alternative experiment, where participants had to decide whether the utterance was a question or a statement. Specifically, we looked at the accuracy scores and response times of the participants.

Results showed similar accuracy scores for the presence or absence of eyebrow movement. Interestingly, no significant difference was found between the type of eyebrow movement (frowns, raises, absence) either. As for response times, participants were significantly faster to identify questions when an utterance was paired with a frown in comparison to a raise or the absence of eyebrow movement. Moreover, participants were significantly quicker to identify a question with no eyebrow movement in comparison to a raise.

The findings that the presence of eyebrow movements help to identify questions faster is in line with previous research that looked into the association between question identification and eyebrow movements (Bavelas et al., 2014; Chovil, 1991; Flecha-García, 2010; Granström et al., 1999; Nota et al., 2021, in prep; Srinivasan & Massaro, 2003). This finding supports the idea that eyebrow movements play a role in question identification. This finding also goes beyond what is already known from previous studies, as the effect of eyebrow movements on response times for question identification has hardly been investigated.

The results showing that frowns specifically play a role in the identification of questions is in line with previous research. This result is similar to that found by Srinivasan & Massaro (2003). The effect that frowns play a bigger role in question identification in comparison to frowns is demonstrated further by Nota et al. (in prep). It could be that frowns play an essential role in the response times for question identification, whereas eyebrow raises do not. We speculate that it might be that eyebrow raises are a weaker signal for question identification, as they occur with many other social actions, and are therefore not perceived as facial signals that mark questions specifically.

The findings that the presence of frowns and raises did not make a difference in accuracy scores in comparison to the absence of eyebrow movement is not in line with what we expected based on previous research. We expected that frowns specifically would result in higher accuracy scores. Previous research by Nota et al. (in prep) has shown that frowns help identify questions more accurately. However, these previous findings were based on enhanced eyebrow movements. Therefore, it may be that eyebrow movements based on natural intensities do not play a significant role in the accuracy scores for question identification, but only extreme eyebrow movements do.

There are some limitations to this study. Due to the low number of participants, it may be that results would differ significantly from the current study, should this study be repeated on a larger scale. Research on a larger sample is necessary to state precisely whether eyebrow movements based on natural intensities have an effect on accuracy scores and response times. We speculate that the results for a larger sample might include more variation in the results. However, further research is needed to confirm our suspicions.

Furthermore, it might be the case in this study that participants were more attentive and more careful when judging the utterances, as they were doing the experiment in favour of a friend. As they were judging the utterances more carefully, this may have had an impact on the response times and accuracy scores. We speculate that the results may differ when the experiment is done with participants that do not have any relation to the researchers. However, further research is needed to confirm whether this speculation is correct.

Another limitation is that no exploratory analysis was done for the effect of intensity scores on accuracy scores. Therefore, we were unable to find out whether the different intensities of the eyebrow movements played a role in question identification. It may be that only the highest intensity has an effect on question identification, as this is in line with research by Nota et al. (in prep).

Future studies may find it interesting to distinguish between the intensity scores, to identify whether the results differ per intensity. From that, one might be able to conclude whether only extreme eyebrow movements have an effect on response times and accuracy scores or not.

6. Conclusion

Overall, the results showed that frowns have an effect on response times. However, eyebrow movements did not have an effect on accuracy scores. This suggests that frowns play a role in the communication of questions. This study provides insight into the effect of the manipulation of eyebrow movements on the response times on question identification. Our study shows the important role of investigating visual signals used in communication, and may be informative for the development of educational video resources as well as the development of social robots.

7. References

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x

Baltrusaitis, T., Zadeh, A., Lim, Y. C., & Morency, L.-P. (2018). OpenFace 2.0: Facial Behavior Analysis Toolkit. *2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018)*, 59–66. https://doi.org/10.1109/FG.2018.00019

Baron-Cohen, S., & Wheelwright, S. (2004). The Empathy Quotient: An Investigation of Adults with Asperger Syndrome or High Functioning Autism, and Normal Sex Differences. *Journal of Autism and Developmental Disorders*, *34*(2), 163–175. https://doi.org/10.1023/B:JADD.0000022607.19833.00

Bastoni, M. (2021). *MB-Lab (1.7.6) [Python]*. https://github.com/animate1978/MB-Lab

Bavelas, J., Gerwing, J., & Healing, S. (2014). Hand and facial gestures in conversational interaction. In *The Oxford handbook of language and social psychology* (pp. 111–130). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199838639.013.008

Broersma, P., & Weenink, D. (2021). *Doing phonetics by computer* (6.1.40) [Computer software]. http://www.praat.org/

Chovil, N. (1991). Discourse-Oriented Facial Displays in Conversation. *Research on Language & Social Interaction*, *25*, 163–194. https://doi.org/10.1080/08351819109389361

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, *20*(1), 37–46. https://doi.org/10.1177/001316446002000104

Community, B. O. (2018). *Blender—A 3D modelling and rendering package.* Blender Foundation. https://www.blender.org/

Ekman, P., & Friesen, W. V. (1978). Manual of the facial action coding system (FACS). *Trans. ed. Vol. Consulting Psychologists Press*.

Enfield, N. J., Stivers, T., & Levinson, S. C. (2010). Question–response sequences in conversation across ten languages: An introduction. *Journal of Pragmatics*, *42*(10), 2615–2619. https://doi.org/10.1016/j.pragma.2010.04.001

Englert, C. (2010). Questions and responses in Dutch conversations. *Journal of Pragmatics*, *42*, 2666–2684. http://dx.doi.org/10.1016/j.pragma.2010.04.005

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. https://doi.org/10.1016/j.specom.2009.12.003

Gísladóttir, R. S., Chwilla, D. J., Schriefers, H. J., & Levinson, S. C. (2012). Speech act recognition in conversation: Experimental evidence. *Http://Pubman.Mpdl.Mpg.de/Pubman/Item/Escidoc:1340580:11*. https://repository.ubn.ru.nl/handle/2066/102009

Granström, B., House, D., & Lundeberg, M. (1999). *Prosodic Cues in Multimodal Speech Perception*. https://www.semanticscholar.org/paper/PROSODIC-CUES-IN-MULTIMODAL-SPEECH-PERCEPTION-Granstr%C3%B6m-House/31f81f6cd554566e7ec131d2111a766d0355a722

Granström, B., House, D., & Swerts, M. (2002). *Multimodal Feedback Cues in Human-Machine Interaction*. https://www.isca-speech.org/archive_open/sp2002/sp02_347.html

Holle, H., & Rein, R. (2015). EasyDIAg: A tool for easy determination of interrater agreement. *Behavior Research Methods*, *47*(3), 837–847. https://doi.org/10.3758/s13428-014-0506-7

Holler, J., & Levinson, S. C. (2019). Multimodal Language Processing in Human Communication. *Trends in Cognitive Sciences*, *23*(8), 639–652. https://doi.org/10.1016/j.tics.2019.05.006

House, D. (2002). *Intonational and visual cues in the perception of interrogative mode in Swedish*.

Kendrick, K., & Torreira, F. (2015). The Timing and Construction of Preference: A Quantitative Study. *Discourse Processes*. https://doi.org/10.1080/0163853X.2014.955997

Kisler, T., Reichel, U. D., & Schiel, F. (2017). Multilingual processing of speech via web services. *Computer Speech & Language*, *45*, 326–347.

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*(3), 396–414. https://doi.org/10.1016/j.jml.2007.06.005

Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, *33*(1), 159–174. https://doi.org/10.2307/2529310

Nota, N., Trujillo, J. P., & Holler, J. (2021). Facial Signals and Social Actions in Multimodal Face-to-Face Interaction. *Brain Sciences*, *11*(8), 1017. https://doi.org/10.3390/brainsci11081017

Nota, N., Trujillo, J. P., & Holler, J. (in prep). *Conversational eyebrow frowns facilitate question identification: An online VR study*.

Nota, N., Trujillo, J. P., & Holler, J. (In prep). *Female (left) and male (right) avatar models* [Photograph].

Perniss, P. (2018). Why We Should Study Multimodal Language. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.01109

RStudio Team. (2021). *RStudio: Integrated Development Environment for R* (1.4.1717) [Computer software]. RStudio PBC. http://www.rstudio.com/

Sloetjes, H., & Wittenburg, P. (2008). *Annotation by category: Elan and iso dcr*.

Srinivasan, R. J., & Massaro, D. W. (2003). Perceiving prosody from the face and voice: Distinguishing statements from echoic questions in English. *Language and Speech*, *46*(1), 1–22.

Tomar, S. (2006). Converting video formats with FFmpeg. *Linux Journal*, *2006*(146), 10.

Trujillo, J. P., & Holler, J. (2021). The Kinematics of Social Action: Visual Signals Provide Cues for What Interlocutors Do in Conversation. *Brain Sciences*, *11*(8), 996. https://doi.org/10.3390/brainsci11080996

van der Meer, H. A., Sheftel-Simanova, I., Kan, C. C., & Trujillo, J. P. (2021). Translation, Cross-Cultural Adaptation, and Validation of a Dutch Version of the Actions and Feelings Questionnaire in Autistic and Neurotypical Adults. *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-021-05082-w

van der Struijk, S., Huang, H.-H., Mirzaei, M. S., & Nishida, T. (2018). FACSvatar: An Open Source Modular Framework for Real-Time FACS based Facial Animation. *Proceedings of the 18th International Conference on Intelligent Virtual Agents*, 159–164. https://doi.org/10.1145/3267851.3267918

Williams, J., & Cameron, I. (2017). The Actions and Feelings Questionnaire in Autism and Typically Developed Adults. *Journal of Autism and Developmental Disorders*, *47*. https://doi.org/10.1007/s10803-017-3244-8