Foreigner directed gesture: larger, faster, longer

Valentijn Prové[1], Bert Oben[1], and Julien Perrez[2]
[1]*Department of Linguistics, University of Leuven (KU Leuven), Belgium*
[2]*Department of Modern Languages, University of Liege (ULiege), Belgium*

Abstract

This paper addresses the issue of foreigner talk from the perspective of multimodal communication. More specifically, it aims at analysing to what extent native speakers (NS) adapt their non-verbal behaviour (in this case manual co-verbal gestures) when interacting with non-native speakers (NNS). Our study is based on a sample of NS-NNS and NS-NS task-based interactions that were analysed using OpenPose estimations to derive kinematic data on the spatial and temporal dimensions of gesture production. On the basis of the existing literature, it was hypothesized that non-verbal foreigner talk would be characterized by an increase in gesture rate, gesture duration, hold rate and duration, gesture size, gesture trajectory and gesture velocity. Whereas our findings do not support the hypothesis that NS produce more gestures when interacting with NNS, they show that NS use gestures that are larger in size (mainly in the horizontal dimension), that are more diverse in the vertical dimension, that are performed faster and that cover a larger trajectory. These findings confirm that foreigner talk manifests itself in formal features of non-verbal communication and call for more research on the exact role of multimodal foreigner talk in multilingual communication as well as on its potential effects on non-native speakers' attitudes and language proficiency.

Keywords: foreigner talk, gesture studies, foreign language learning, multimodal communication, foreigner directed gestures

1 Introduction

One of the basic tenets in sociolinguistics, cognitive linguistics and increasingly also in psycholinguistics and the broader domain of cognitive science, is the assumption that language (use) should not be studied in a social vacuum. In other words, linguistic choices are not the exclusive result of purely cognitive processes, but also dependent on context. This context, crucially, often involves other speakers, and as a consequence, language use of one speaker is related to the language use of the speakers in his/her vicinity. The fact that speakers adapt their language use to that of their interlocutors has been approached through many different theoretical frameworks, including Communicative Strategies (Faerch & Kasper, 1983 ; Canale & Swain, 1980 ; Poulisse, 1990 ; Yule & Tarone, 1990 ; Dörnyei & Scott, 1997 ; Gullberg, 2011), Recipient Design

(Sacks et al., 1974; Depperman, 2015), Audience Design (Bell, 1984, 2002 ; Clark & Murphy, 1982) and Communication Accommodation Theory (CAT, Coupland and Giles, 1988; Gasiorek et al., 2015).

In an increasingly connected world, conversations between interlocutors with different proficiency levels of the language used, occur ever more frequently: teachers and pupils do not always share the same mother tongue, in a globalized business environment the workplace has become a multilingual stage, and also in the public domain interactions between natives and non-natives are the order of the day. To achieve the conversational goal at hand, interlocutors develop conversational strategies to accommodate to one another. For verbal behaviour this type of adaptation is well documented (a.o. Ferguson, 1971; Horton & Gerrig, 2016; Long, 1981; Milk, 1990; Roche, 1998; Rodriguez-Cuadrado, Baus & Costa, 2018; Rogers et al., 2013; Wooldridge, 2001). To date, an elaborate account of non-verbal adaptations is still underexplored territory. In this study we will focus on non-verbal adaptations during dyadic interactions made by native speakers for the benefit of non-native speakers' understanding. More specifically, we will compare the formal properties of hand gestures (e.g. size, frequency, velocity) performed by native speakers towards non-native interlocutors, with gestures performed towards native interlocutors.

1.1 Foreigner talk as static register vs. dynamic coordination

Native speakers (NS)[1] adapt their speech to the language proficiency of their interlocutors. For example, they talk more slowly, reduce syntactic complexity or use more high-frequent words to accommodate less proficient interlocutors. When NS interact with non-native speakers (NNS) this type of adaptation is referred to as *foreigner talk* (FT). Ferguson (1971), who coined the term, describes it as "more-or-less conventionalised varieties of language used by members of a speech community to address people whose knowledge of the language of the community is felt to be less than normal" (Ferguson & DeBose, 1977:100). In his view, FT is a simplified register that shares features with ways of talking to older people, infants and pets. To investigate FT as a variety of English, Ferguson (1975:3) asked participants to provide equivalents of sentences that NS would use to communicate with "uneducated non-Europeans" [sic] and in doing so he reports recurrent features such as grammatical omissions, expansions and reduplications. He furthermore suggests that this variety is learned and passed on by the media, evoking *'me Tarzan, you Jane'* as a typical example.

The very static view on FT as a fixed register that can be either applied or not applied, has recently been challenged by researchers who focus on the flexibility in adaptations towards NNS (Margic, 2017). In

---

[1] We acknowledge the fact that '(non-)native speaker' is a questionable term. In an increasingly mobile and international world, place of birth and linguistic experience can combine in multiple ways. We use the terms 'native' and 'non-native' to intuitively and loosely refer to a speaker's first language (which again can be problematic in cases of (near-)bilingualism or highly proficient second language speakers). See our methods section for a further specification of 'nativeness' in the present dataset.

this more recent take on FT, the focus lies on the interaction between NS and NNS, and the ad hoc solutions they find for overcoming communicative issues. NS appear to modify their speech more extremely if they are only able to rely on stereotyped expectations about proficiency (Scarborough et al., 2007). In real interaction, attributions made from top-down expectations are influenced by addressee feedback (Warren-Leubecker and Bohannon III, 1982). In other words, NS might start out with a model of the NNS's proficiency, but during interaction this model gets updated, NS and NNS build common ground (cf. Clark, 1996), and NS (but also NNS) adapt their language use accordingly.

In addition, FT adaptation interacts with other temporal phenomena. Rodriguez-Cuadrado et al. (2018) study how words get phonetically reduced if they recur more frequently in FT. This phenomenon is known as *word reduction* and occurs frequently in 'normal' speech among NS. The occurrence of word reduction in interactions between NS and NNS suggests that FT strategies change in the course of the interaction. This is at odds with Ferguson's research, which focused on the NS only, and tested strategies of FT in imagined rather than real interactions. Interesting in this respect is the work by Scarborough and colleagues (2007) who show that NS perform more simplifications when they are asked to talk to an *imagined* NNS compared to when they actually talk to NNS. Even though the authors do not explicitly draw this conclusion, their results are indicative of the dynamic character of FT we wish to advocate: NS update their model of the NNS's proficiency on the fly and adapt their speech accordingly.

A specific context for interaction between interlocutors with a large difference in language proficiency is the classroom. The adaptations by teachers are often referred to as *teacher talk* (see for example Edwards & Westgate, 1994; Kim & Elder, 2005; Lew, 2016; Lee, 2007; Walsh, 2002; Yanfen & Yuqin, 2010). Research on second language acquisition has indicated that interaction is indispensable for language learning (Gass, Mackey & Pica, 1998; Loewen & Sato, 2018; Long, 2014) and simplification strategies in classroom interactions do occur, but there is conflicting evidence on whether or not language learners benefit from teacher talk (cf. Maleki & Pazhakh, 2012; Issidorides & Hulstijn, 1992). Even though the topic of teacher talk is related to foreigner talk, we want to clarify that we are not considering classroom interaction in the present study, in which we focus on non-educational, interactional settings only.

1.2 Foreigner talk at different levels

Depending on the theoretical framework and on the exact definition of the phenomenon, foreigner talk has also been labelled as *xenolect* (Roche, 1989), *foreigner-directed speech* (Fischer 2016), *adapted foreigner-directed communication* (Bergmann et al., 2018) or *native/non-native interaction* (Rodriguez-Cuadrado et al., 2018). Notwithstanding the (sometimes subtle) differences in explaining the processes underpinning FT, research on the topic converges in demonstrating how FT occurs at many different linguistic levels. At the phonological level, FT appears to be characterized by longer silence durations in between utterances, but also in between words and syllables (Osada, 2003; Papousek & Hwang, 1991). In addition, NS avoid clitics (for example avoid “d’you think” or “he’ll go” for their full forms “do you think” and “he will go”) and display a wider range in intonation (Roche, 1998; Smith, 2008). Both at a morphological and syntactic level, NS apply strategies of simplification. These strategies include a preference for the canonical word order, explicitly marking syntactic relations (e.g. by preferring “the book of John” over “John’s book”) or choosing coordinated over subordinated constructions (Dashti, 2013; Renkema, 1987; Woolridge, 2001). At a more pragmatic level, NS produce more polar questions relative to WH-questions, reduce the overall number of idioms, and more often use repetitions and paraphrases (Bortfeld & Brennan, 1997; Seidlhofer, 2009). From a larger discourse-structure perspective, there is evidence that NS more explicitly express relations between utterances or larger discourse structures, more often ask for feedback from the NNS or allow for more abrupt topic changes than they would in interactions among natives only (Smith et al., 1991; Gass, 2003).

So far, the studies presented above only take into account speech properties of FT. Research in the domain of gesture studies has convincingly demonstrated how speakers flexibly adapt their bodily behaviour to meet the needs of their conversational partners. For example, interlocutors copy each other’s hand gestures to facilitate mutual understanding (Holler & Wilkin, 2011; Kimbara, 2008; Mol et al., 2012; Authors, 2016), alter their body orientation or gaze direction to indicate viewpoint shifts (Debras, 2015; Stec, 2012), or alter the direction of their hand gestures to maximally exploit the resources of the shared space between conversational partners (Özyürek, 2002). Gullberg and Kita (2009) propose that gestures in peripheral gesture space may guide the focus of attention and highlight the relevance of gestures in interaction. Moreover, gestures are made smaller and less precise in repeated gestural referrals (Gerwing & Bavelas, 2004; Namboodiripad, Lenzen, Lepic & Verhoef, 2016; Holler & Bavelas, 2017) to mark information as given, which is directly analogous to word reduction. These are all examples of adaptation strategies in NS-NS interactions, in which speakers adapt their bodily behaviour to what they believe to be beneficial for their conversational partners.

Recently, however, also scholars involved in research on foreigner talk, are starting to look into other semiotic levels and find that simplification strategies can be found in hyperforms of bodily behaviour, such as big or slow gestures and sustained eye gaze. Gullberg (2011:143) argues that gesture and space provide powerful means for disambiguation. More specifically, gestures meant for NNS are articulated differently, with locations associated with entities being spatially more differentiated. In the context of interactional learning in classroom settings, Tellier and Stam (2010) and Azaoui (2013) have found a significantly increased gesture rate and gesture size. The same results have been reported in a word explanation task (Bergmann et al. 2018) and in communication with foreign virtual agents (Lugrin et al., 2018). Concerning gesture frequency, Adams (1998) reports that the fact that the gesture rate increased in the NNS condition of his storytelling experiment was mainly due to an increase in the rate of deictic gestures.

Also for the related phenomenon of *motherese* (i.e. adaptations made by caregivers to the benefit of the children they are interacting with) hand gestures appear to be larger and more iconic (Brand et al., 2002; Campisi & Özyürek, 2013).

These results for adaptations in gesture use by NS, albeit still very scarce and often resulting from pilot studies, tie in with results from acoustic hyperarticulation features of FT outlined above. For instance, given that pitch, vowel space and vowel duration increase in speech directed to NNS (Uther et al. 2007), we could expect gestures to be larger and performed more slowly as well. As for the velocity, however, Trujillo et al. (2018) have found that speakers produce faster gestures in settings that are communicatively more demanding. Hence, gesture velocity will remain one of the more explorative parts of our analysis. Moreover, we can expect gestural adaptation to be very variable in nature. Knoll et al. (2015) observe that clarity-enhancing strategies are more variable in FT compared to interactions with other 'less proficient' addressees because the communicative needs of NNS are less univocal. Infants, for instance, clearly benefit from hyperarticulation (or larger gestures) for developing phonetic discrimination skills and older people may have hearing (or viewing) difficulties, which elicits increased loudness (or gesture size) from the speaker. However, NNS are confronted with more variable communicative challenges during interaction. As a consequence, NS have to take into account more than physical or developmental factors in crafting their utterances during FT.

Even though some attention has been paid to the formal characteristics of hand gesture in mothesere (mainly gesture size: Brand et al., 2002; Campisi & Özyürek, 2013), there remains a lot of work to be done in accounting for the characteristics of foreigner-directed gesture (FDG). In the first place, the existing studies quite coarsely annotate physical properties of FDG, for example by having raters judge gesture size and complexity on 4- or 5-point scales (Brand et al., 2002; Campisi & Özyürek, 2013) or by using McNeil's (1992) gesture space diagram as a proxy for gesture size (Bergman et al., 2018). Second, FDG, so far, has been studied in very different conversational settings with NNS either being physically present or imagined by the NS, and with NS asked to either demonstrate an action or describe short words/sentences. Validating findings (e.g. of increased gesture size in FDG) under different conversational circumstances (e.g. bidirectional communication between NS and NNS or longer stretches of interaction) is, given the limited number of studies on the topic, an effort worthwhile.

1.3 The present study

From the literature review above, we can tentatively conclude that NS produce more and more expressive gestures in NS-NNS interactions, compared to NS-NS interactions. However, the evidence is still scarce and fragmented. With the present study, we aim to replicate those results (i) in a more interactional type of conversation and (ii) with more kinematic detail in both spatial and temporal dimensions. More specifically, we hypothesize that NS adapt their manual gestures when interacting with NNS by:

1) Increasing gesture rate
2) Increasing gesture duration
3) Increasing hold rate and duration
4) Increasing gesture size
5) Increasing gesture trajectory
6) Increasing the maximum velocity of their gestures

How we tested the hypotheses outlined above, will be made clear in the methods section below.

2 Method and materials

2.1 Participants

Forty-four students from undergraduate courses volunteered for this study. Ten native speakers of French (mean age 21.8) were recruited at ULiège to create a group with native-native (NS) dyadic interactions in French. At the Antwerp Campus of KU Leuven (Belgium), a Dutch-speaking counterpart was composed by 12 native speakers of Dutch (mean age 19.2). At KU Leuven (Leuven Campus), 11 native speakers of Dutch formed dyads with 11 native speakers of French (mean age 22.4) so as to allow for native-non-native interactions. As a result, we disposed of a NS control condition and a NNS test condition in two languages. This design enabled us to investigate NS and NNS interactions in both Dutch and French and to use the language spoken as a control variable.

As already discussed in the introduction, the categories *native* and *non-native* are not unproblematic. We want to stress here that we acknowledge the categories to be very coarse-grained. A lot of inter-individual differences underlie the two distinct groups of NS and more particularly NNS. NNS in our study have very different proficiency levels ranging from near-native (e.g. university students studying the given language or participants with a lot of relatives speaking the given language) to non-fluent (e.g. students that have learned the given language at secondary school but hardly ever speak it outside that educational context). What all NNS have in common is that they (i) grew up in and attended school in a French/Dutch-speaking community and (ii) had sufficient language skills of the other language to engage in a rudimentary conversation.

2.2 Procedure

Per group, the subjects were allocated a partner and each pair engaged in a picture description task. The researcher requested them to sit on chairs that were oriented towards each other at a distance that was well-suited for conversation. Next to each of the chairs, a camera (Sony HDR-CX160E) was mounted on a tripod to film the participants from a frontal perspective. The subjects were requested to look at and memorize a picture, which was presented to them on a 13-inch laptop screen and describe it to their partners from memory in such a way that they would be able to distinguish it from a similar picture that differed in a few details. As such, the task was conceived as a variant of a 'spot the difference' game with the roles of a director and a matcher. While the subjects were describing the item, the stimulus was removed. There was no time limit set for either the memorization or the description part and the addressees were allowed to ask questions. In total, there were four items to describe ('forest', 'gym', 'lake' and 'playground'), which are provided in the Appendix (Figure 7-14). A subject would describe two items and their partner would describe the other two. The turns were alternated.

Our speech elicitation method is inspired by the Diapix task which has been used for sociophonetic research (Van Engen et al., 2010; Authors, 2016; Hall-Lew & Boyd, 2020). However, whereas in the original Diapix task both participants have their own version of the picture which they need to discuss with each other to spot the differences, the present task was more asymmetric in nature (director and matcher roles) and there was a memory component involved. Hence, it joins in with the term explanation paradigm used in previous studies on FDG (Tellier & Stam, 2010; Bergmann et al., 2018), but it is different in that the stimuli are visual. In order to describe the pictures in detail, the subjects needed to memorize multiple entities together with the spatial relations between them. In line with theories of embodied cognition (see Hostetter & Alibali, 2019 for a review), we expected spatial and visual imagery to activate motor actions and to elicit more gestures.

2.3 Pose estimation and gesture coding

We exported the parts of the recordings during which the stimuli were being described as separate .mp4 clips (resolution = 720 × 480 px, frame rate = 25 fps). Subsequently, we implemented OpenPose (Cao et al. 2017) to automatically detect 25 body parts on each of the frames of the clips.[2] OpenPose is a popular computer vision framework that is well-suited for semi-automated gesture segmentation (De Beugher et al. 2018; Ripperda et al., 2020). In experimental settings, this method allows for a minimally distracting recording set-up with camcorders or even smartphones. Specialized motion tracking devices are more accurate, but they require markers to be attached to the body, a calibration procedure and a fair amount of practical expertise. The only reliable device that is currently available for unobtrusive experimental recordings is the Microsoft Kinect (see Pouw et al., 2020 for a review). Although the temporal resolution of the Kinect (30 Hz) does not significantly differ from the usual frame rate of video recordings as input for

---

[2] In order to enhance the accuracy of the estimations, we enabled the '--tracking' feature and we set the maximum number of people detected ('--number_people_max') to one. For a list of the body parts, see https://github.com/CMU-Perceptual-Computing-Lab/openpose.

computer vision methods (25–30 Hz), it is often used for studying gesture kinematics (e.g. Trujillo et al., 2018) because it supports three-dimensional tracking, thus enhancing spatial resolution.

Since OpenPose is a video-based person tracking software, the output consisted of two-dimensional coordinates of each body part. We imported the estimations as continuous time series data in R (R Core team 2021)[3] and we pre-processed them in two steps. First, we selected the left and right wrist estimations and we scaled and centred them to the neck and hip coordinates so as to minimize differences between the subjects due to body size and camera position. Mathematically, we divided the difference between the wrist and the neck coordinates by the difference between the neck and the hip coordinates for each frame in the video. As a result, we had a new coordinate system that was adjusted for each subject, with the neck as the origin (0,0) and the distance to the hip (0,-1) as a unit to calibrate the axes. Second, we applied a Kolmogorov-Zurbenko low-pass linear filter (from R-package ‘kza’, span = 5, order = 3) to remove the noise due to incorrect estimations. We checked the result of this procedure by exporting animated plots of the adjusted and smoothed wrist trajectories as .mp4 clips and comparing them to the original video frame by frame in ELAN (Wittenburg et al., 2006). Figure 1 illustrates how the unit of the vertical (y) and horizontal (x) position should be interpreted. Zooming in on the y-position, i.e. the vertical movements, we see that the speaker raises both hands in panel 2 and then continues raising the right hand even further in panel 3. This increase in vertical hand position can also be read from the plot below, with both hands starting at roughly -0.3 in panel 1 and ending at roughly at 0 in panel 2, which corresponds to the position of the neck (cf. blue dot). In panel 3 the right wrist moves above the neck, reaching a positive value, roughly at +0.5. The same rationale applies to the horizontal movements: whereas the right hand is further away from the centre of the body (i.e. the vertical line that runs through the blue dot representing the neck) than the left hand in panels 1 and 2, the right hand moves closer to the centre of the body (i.e. closer to 0 in the plot) in panel 3. In other words, the newly created coordinate system should be interpreted literally since it corresponds to anatomical parts.

---

[3] All scripts and datasets used for this study can be found in the Zenodo repository, https://zenodo.org/badge/latestdoi/421998192.

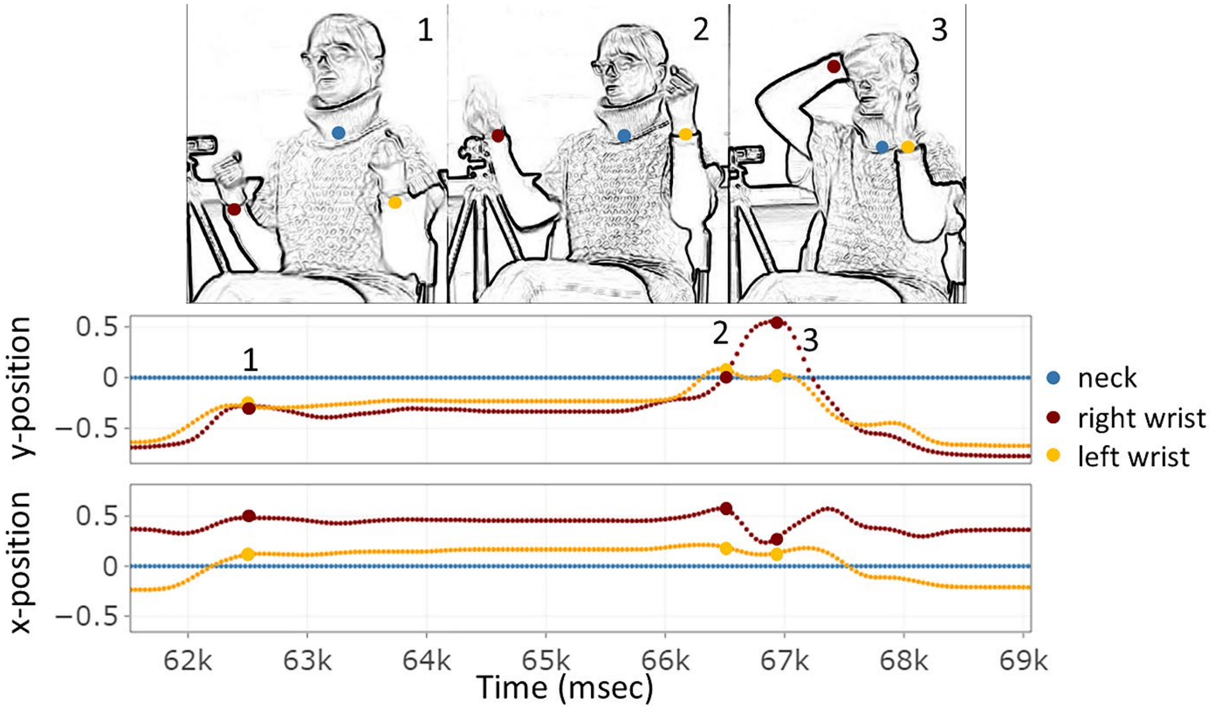

Figure 1. X- and y-coordinates of the right and the left wrist during a gesture unit of ca. 0.7 seconds, scaled and centred to the subject's neck and hip.

Apart from the OpenPose based annotation described above, we also used ELAN to code gesture units, which we define as segments of movement in the upper limbs that contain precisely one stroke, one hold or both. We side with Bressem and Ladewig (2011) in defining a stroke as a change in the position of the hand with respect to time, where the tenseness of the muscles that are involved in this movement remains constant. This means that the configuration of the hand does not change. Hence, a stroke boundary occurs when this articulatory tension is loosened, for instance when the hand finds a rest position or pose, or when the configuration of the hand changes. Moreover, the stroke is regarded as the meaningful part of the gestural movement and, as a consequence, each stroke should represent one functional unit. A hold differs from a stroke by the single criterion that it has no dynamic motion. In other words, a hold means that the hand is actively held in a single configuration, with a constant articulatory tension, but the muscle contraction does not lead to movement. We also coded preparation and retraction phases as part of the gesture units, which are straight or curved movements with a less variable flow of movement that either involve increasing (preparation) or decreasing (retraction) articulatory tension.

2.4 Kinematic measures and variables

In order to provide a kinematic account of FDG, we calculated different measures on the basis of the data described above. In the case of a two-handed gesture, we selected the maximum value of the two hands.

Gesture rate is a relative measure for frequency and was calculated as the number of gesture units per minute. Whereas this measure captures how many strokes a subject makes, our *proportion of gesture time* variable represents the total gesture time relative to the total description time. In other words, it is a measure of how many percent of the time a subject spends gesturing. Similar to gesture rate, we also calculated the *hold rate* per minute.

Amplitude is a one-dimensional measure for gesture size. It was calculated as the maximum value on both the vertical and the horizontal axis. As described in the previous section, the gesture trajectories were scaled with respect to the position of the subject's neck and hip so that the amplitudes can be interpreted in an anatomical or literal way. For instance, a vertical amplitude of 0 means that the wrist is at exactly the height of the neck. *Gesture space* refers to the two-dimensional area that is outlined by a gesture trajectory. It was calculated as the product of the range, i.e. the difference between the maximum and the minimum amplitude within one gesture unit, of the vertical and the horizontal axis. To avoid data skewness, we rescaled the measurements for gesture space using a logarithmic transformation.

Figure 2 illustrates in more visual detail how to interpret the amplitude and gesture space measures. Every dot in the figure corresponds to the coordinates of every video frame for one gesture by the respondent. The dots within the dashed box (in orange) indicate right-hand movement, dots outside of the box left-hand movement. The red crosshairs display the vertical and horizontal amplitude of the right-hand gesture (with the bright red dot highlighting the position of the neck, which is the point of reference in our coordinate system, cf. supra). The dashed box (in orange) constitutes the gesture spaces of the right-hand gesture.

Gesture duration refers to the duration in milliseconds of a gesture unit. *Gesture trajectory* was calculated as the sum of the consecutive Euclidean distances from frame to frame within one gesture unit. Hence, it is a measure for the overall change of the position of the hands. *Peak velocity* is the maximum velocity within one gesture unit as calculated by dividing the distances by the time.

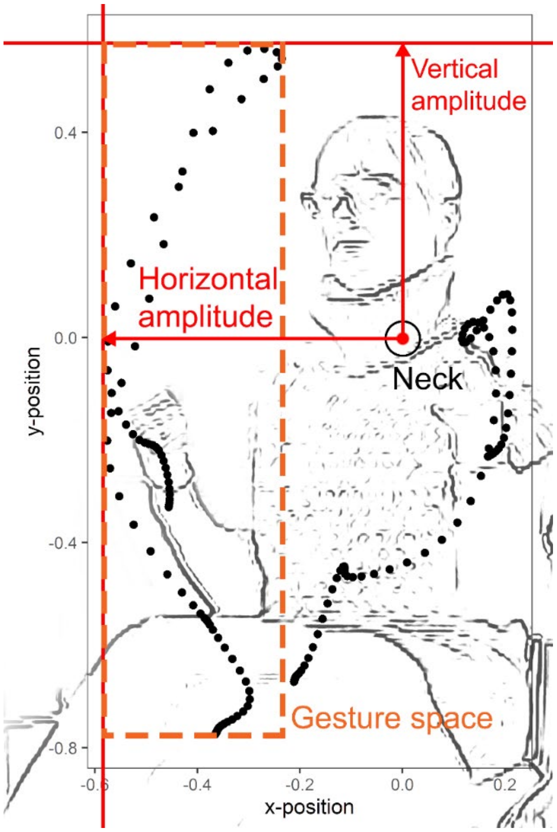

Figure 2. Illustration of vertical/horizontal amplitude and gesture space as measures for gesture size.

3 Results

For the statistical analysis, we structured our dataset in two ways. On the one hand, we aggregated our gesture frequency data per description so that we could conduct ANOVA tests with the gesture rate, hold rate and proportion of gesture time as response variables. On the other hand, we kept the kinematic observations per gesture unit (size, amplitude, trajectory, duration and peak velocity) so as to test mixed effects linear regression models with a random intercept for each subject and item. We removed three participants from the datasets because the video quality was too poor for OpenPose to work, which leaves forty-one participants.

3.1 Gesture rate

In total, we annotated 1,476 gesture segments, with a mean frequency of 17.2 (SD = 8.7) per item description. The participants in the NNS condition produced more gesture segments per description (M = 20.8, SD = 8.7) compared to the NS condition (M = 13.4, SD = 7.0). However, this variation is explained by the fact that the speakers in the NNS took more time in describing the pictures (93.0s on average, SD = 38.4) compared to the NS condition (65.4s on average, SD = 17.3), which resulted in very small differences between the mean gesture rate per minute in the NNS (M = 14.1, SD = 5.0) and the NS (M = 12.5, SD =

6.0) conditions. Unsurprisingly, our ANOVA model showed no main effect of condition for gesture rate, $F_{(1.77)} = 2.76$, p=ns. In addition, there was no interaction of gesture rate between condition and gender, nor between condition and language. What we did observe was a main effect of language: the French-speaking subjects performed significantly more gestures per minute (M = 14.6, SD = 4.6) compared to the Dutch (M = 11.9, SD = 5.6), $F_{(1.77)} = 4.66$, $p < 0.05$.

We built an analogous model for the hold rate and we found that the subjects in the NNS condition did not significantly produce more holds per minute (M = 5.1, SD = 6.1) compared to the NS condition (M = 3.9, SD = 4.6), $F_{(1.77)} = 1.07$, p = ns. Similar to our model for gesture rate, there was no main effect of gender, but the French-speaking subjects significantly produced more holds per minute (M = 6.1, SD = 6.3) than the Dutch-speaking subjects (M = 3.1, SD = 4.1).

The proportional gesture time did not differ between the NNS (M = 0.59, SD = 0.20) and the NS (M = 0.53, SD = 0.26) conditions. However, it should be noted that the mean proportions are relatively high in both conditions, given they are above 50% of the time.

In sum, speakers produced more gestures in the NNS condition in absolute terms (because they talked more), but not in relative terms (because they did not perform more gestures per minute and they also did not spend more time gesturing).

3.2 Gesture size

We first explored the gesture size data by plotting all gesture trajectories in a two-dimensional graph and colouring them by condition (Figure 3). For both the left and the right hands, it is clear that the subjects in the NNS condition produced more gestures in peripheral gesture space, compared to the NS condition. The boxplots in Figure 4 show that this observation is more outspoken on the horizontal axis, that is the subjects in the NNS condition moved their hand more away from the body midline, rather than producing gestures in higher regions of the gesture space.

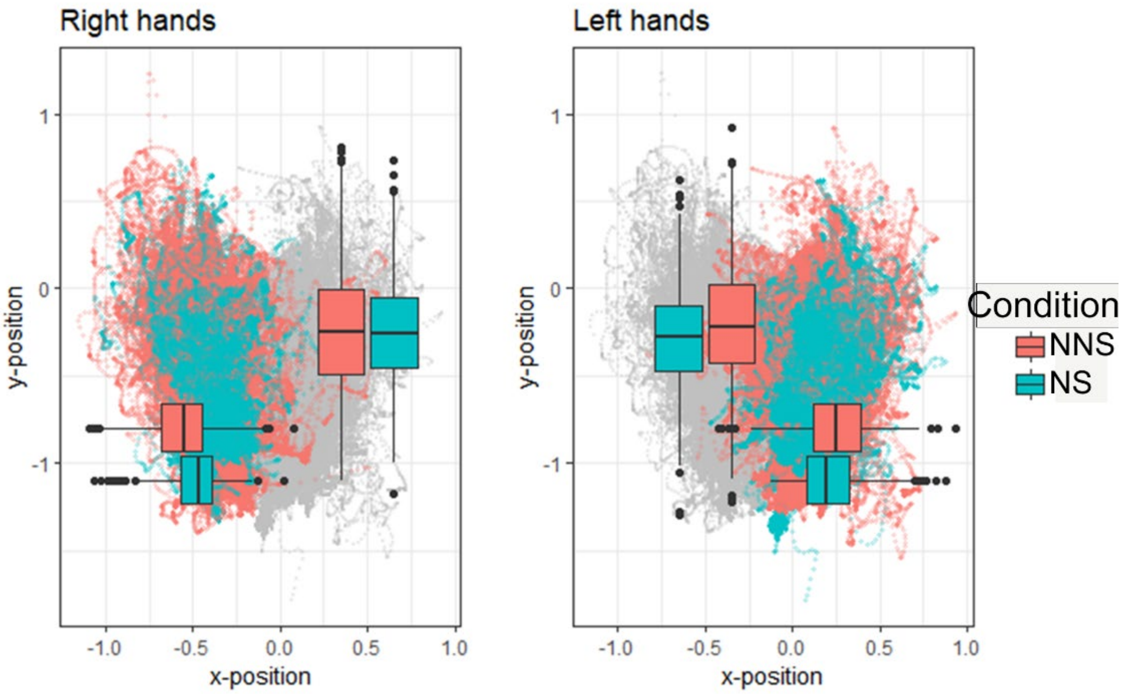

Figure 3. Scatterplot of all gesture trajectories of the right and left hands included in the dataset, colour per condition. The boxplots represent the vertical and horizontal amplitudes.

To statistically test this observation, we computed mixed effects linear models for gesture space, vertical amplitude and horizontal amplitude that allowed a random intercept for item and subject (random effects), with the help of the 'lme4' package in R (Bates et al., 2015). As fixed effects, we tested the condition, gender and language variables. For horizontal amplitude, we also included a hand variable that specified whether the gesture was produced with the left of the right hand or with both hands, because the camera position was slightly shifted to the right and we expected the use of the right hand to cause larger amplitudes. The p-values for statistical significance were obtained using the Likelihood Ratio test as prescribed in Winter (2013).

Concerning the random effects, we used the standard deviation as a measure of how much variability in gesture size there was due to items and subjects. Overall, the standard deviation for item (0.18) was considerably smaller than for subject (0.52). This means that the individual styles of gesturing caused more variability in gesture size than the description items did. As demonstrated in Figure 4, this variability is reflected in the different baselines (intercepts) that the items and the subjects have in our model. The dots represent the items or subjects and the horizontal lines represent the 95% prediction intervals. There are 14 out of 41 participants whose prediction interval does not overlap with zero, i.e. those speakers produce notably larger or smaller gestures.

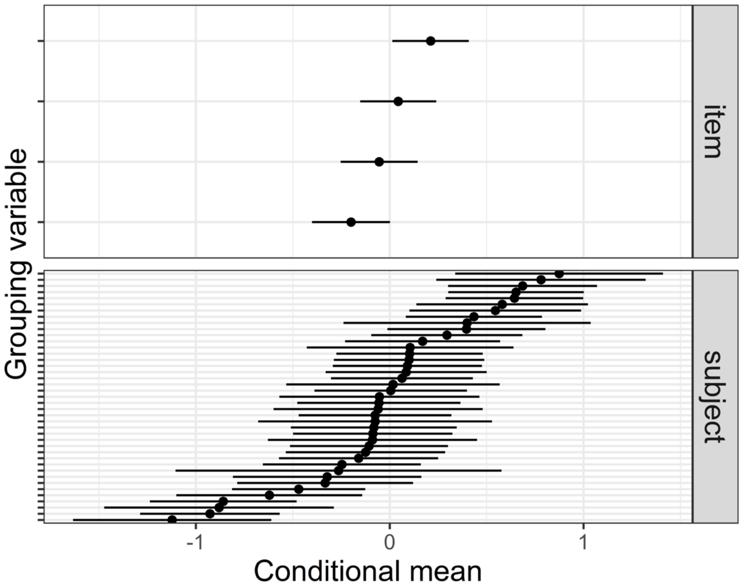

Figure 4. Conditional means and prediction intervals of the random effect grouping variables in the regression model for gesture size.

As for the fixed effects, the NNS condition affected gesture space ($\chi^2_{(1)}$ = 10.80, p <0.01), increasing the intercept (NS condition) of 4.02 (95 % CI [3.70, 4.33]) by 0.62 units (standard error = 0.18), resulting in 4.64 (95 % CI [4.34, 4.93]). Hence, despite the dispersion in the different individual behaviours discussed in the previous paragraph, the explanatory power of the condition (NNS versus NS) remained solid. By contrast, gender and language did not improve the model. In Figure 5, the black line illustrates the regression coefficient of condition. The red error bars represent the upper and the lower limits of the 95% confidence intervals (CI). Put more colloquially, there is a 95% chance that the estimates would fall within this interval if we repeated the present study over and over. The jitter represents the sizes of each gesture, coloured and grouped per subject. From a visual inspection of the plot, it is clear that most subjects, in both conditions, produce gestures with a wide range of different sizes. Conversely, there are very few subjects that produce gestures of the same size all the time. The plot also appears to suggest that the condition effect (i.e. larger gestures in the NNS condition) is not due to the absence of small gestures in the NNS condition, but rather the absence of large gestures in the NS condition.

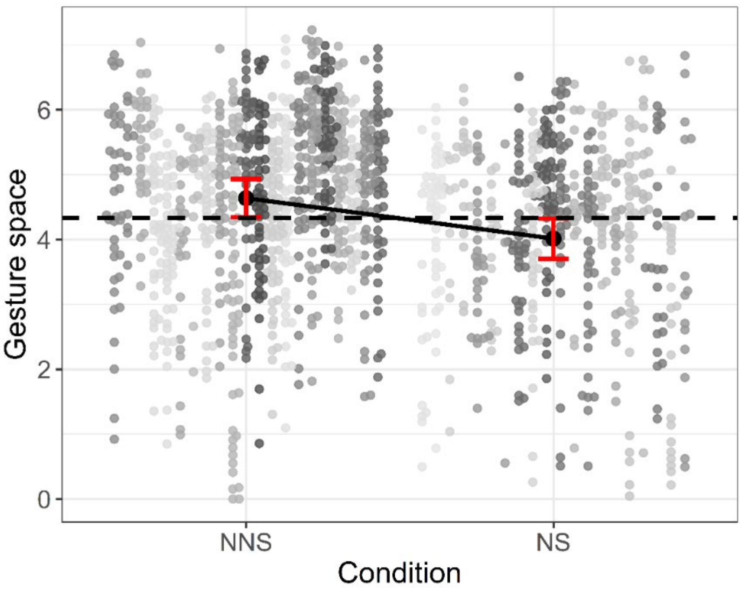

Figure 5. Plot of the estimates for gesture size in the NNS and NS conditions. The red error bars represent 95% CI. The points represent the individual gestures grouped per subject. The dashed line illustrates that the CI's of the two conditions do not intersect.

To provide a more fine-grained insight into the gesture space effect, we have tested whether the larger gestures in the NNS condition are due to larger horizontal hand movements, larger vertical movements, or both. What we found was an effect of the horizontal amplitude: the NNS condition increased by 0.08 units (standard error = 0.03) after improving the null model with the random effects ($\chi^2_{(1)} = 7.15$, $p < 0.01$). The gender and language factors did not improve the model. Given that the unit here ('1') corresponds to the distance between the subject's neck and hip, the actual difference in horizontal amplitude expected on the basis of the model is relatively small.

As could be guessed from a visual exploration of the gesture trajectories in Figure 3, the NNS condition did not significantly affect the vertical amplitude ($\chi^2_{(1)} = 0.66$, p = ns). Although the regression coefficient (0.7) is comparable to the coefficient of horizontal amplitude (0.8), the standard error is considerably higher (0.07, compared to 0.03), which means that the data points are more spread out. Even though there is no significant effect of vertical amplitude in our mixed effects model, Figure 6 does yield some relevant insights. In this figure, subjects are ranked according to the range of their vertical hand movements, with large ranges to the left (i.e. subjects that produce gestures that range from very low to very high up) and small ranges to the right (i.e. subjects for whom the highest and lowest gestures are not that far apart). It shows that subjects in the NS condition have a more restricted vertical range than subjects in the NSS condition. A t-test with condition as grouping variable and the span in vertical amplitude as dependent variable confirms that participants in the NS condition produce gestures with a significantly smaller vertical span (M = 1.02, SD = 0.30) than participants in the NSS condition (M = 1.36, SD = 0.25), $t_{(39)} = 3.89$, p <0.001. In addition, in the NS condition, there are 7 participants that produce all of their gestures below the level of the neck (indicated with large, unfilled triangles), whereas in the NS condition only 2 participants do so. This observation feeds the conclusion that gestures in the NNS condition are not produced 'higher up' per se, but that subjects produce gestures with a more varied vertical range, combining gestures (high) above the neck with low gestures around hip height.

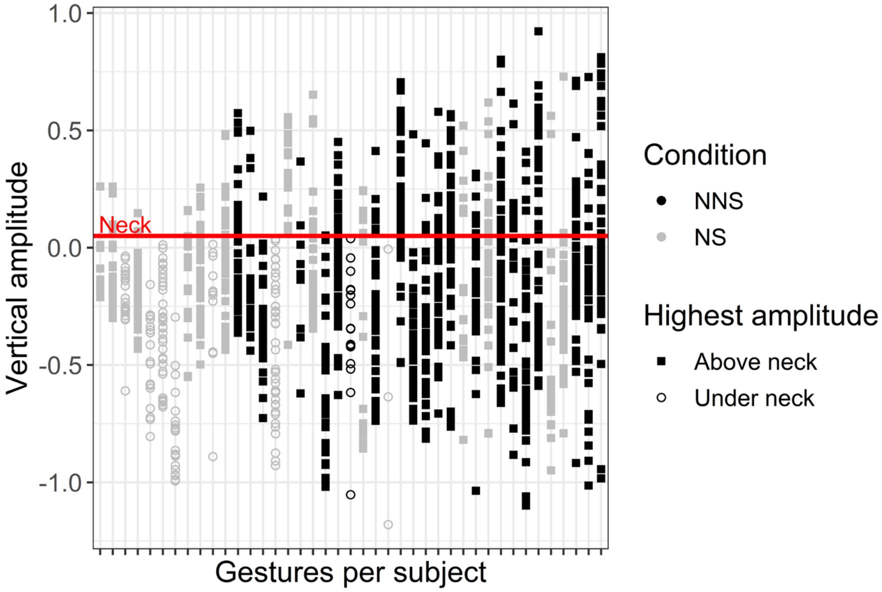

Figure 6. Distribution of the subjects according to the range of their vertical hand movements. Large ranges to the left, small ranges to the right. The unfilled circles represent subjects who only produced gestures under the neck, the filled squares indicate subjects who also produced gestures above the neck.

3.3 Duration, trajectory and peak velocity

The NNS condition affected the trajectory length of the gestures ($\chi^2_{(1)} = 9.15$, p <0.01), increasing the intercept (NS condition) of 0.96 (95 % CI [0.82, 1.10]) by 0.31 units (standard error = 0.10), resulting in 1.27 (95 % CI [1.14, 1.40), but not the gesture duration ($\chi^2_{(1)} = 0.29$, p = ns), with an increase of only 0.11 seconds (standard error = 0.19) resulting in an estimate of 2.59 seconds for the NNS condition compared to 2.48 for the NS condition. Correspondingly, the peak velocity increased in the NNS condition ($\chi^2_{(1)} = 10.41$, p <0.01) by 0.30 units per second (standard error = 0.09). The gender and language factors did not significantly improve the models. We also tested a model for the duration of the hold phase and we did not observe an effect for condition, gender or language. In other words, the instruction givers in the NNS condition performed longer trajectories within the same amount of time, resulting in higher peak velocities.

If the duration of the total gesture units was not affected by the NNS condition, it is not surprising that the same applied to the hold durations. It should be noted, however, that the mean durations of the gesture units were relatively high (M = 2.603 sec, SD = 1.956 for the NNS condition and M = 2.505 sec, SD = 1.679 for the NS condition).

4 Discussion and future directions

In this study we aimed at replicating and refining findings on foreigner-directed gestures. Concerning our first hypothesis, we did not find the expected increase in gesture rate in the non-native condition. NS do produce more gestures when interacting with NNS (compared to when interacting with other NS), but this larger number of gestures can be integrally ascribed to the fact that NS take more time to perform the description task. When calculating the gesture rate in terms of gestures per minute, we observe no difference between conditions. This ties in with the results in Tellier and Stam (2010); however, it is at odds with what Bergmann et al. (2018) and Azaoui (2013) found, viz. a higher gesture rate when NS interact with NNS. A potential reason for the differences in the outcome might reside in the differences in experimental set-up. Azaoui (2013) starts from observations in classroom interactions, which is quite different from our set-up of dyads engaged in task-based interaction. In addition, and also in contrast to our study, both Azaoui (2013) and Bergmann et al. (2018) do not create conditions in which NS either talk to other NS or to NNS. Rather, their conditions concern a difference in proficiency levels of the NNS: a *very low* proficiency level of German (level A1) versus an *intermediate* level (level B1) in the case of Bergmann and colleagues; and L1 learners of French versus L2 learners in the case of Azaoui. In addition, the studies by Bergmann et al. (2018) and Azaoui (2013) cover the gestural behaviour of only a few participants (resp. three and one participants), which given the large inter-individual differences in gestural behaviour (see Figure 4, for example), might considerably impede a fair comparison with the present study. Given the large methodological differences, the small number of studies and the low numbers of participants observed, conclusions on whether (let alone how and why) foreigner talk is characterized by a higher gesture rate have to remain tentative. From the little evidence there is, an increased gesture rate does not seem to be a systematic part of the foreigner talk repertoire. As suggested by the results from Adams (1998), who did not show an overall increase in the NNS-condition, but an increase for deictic gestures only, in future research it is probably worthwhile to not only study formal characteristics of gesture (such as frequency and size) but also take functional aspects (such as gesture types) into account. Depending on the type of interaction (task-based, teaching, chit-chatting, instruction-giving, etc.) foreigner directed gestures might display frequency differences for specific types of gestures (referential, deictic, pragmatic, etc.) only.

Turning to our second hypothesis, we did not find the expected lengthening of the duration of strokes and holds in the NNS condition. Our non-effect does not add to the results from Tellier and Stam (2010) who found gesture duration to be twice as long in the NNS condition than in the NS condition. Overall, participants appear to produce longer gestures in the present study, compared to Tellier and Stam (2010). This could be explained in terms of the Gesture as Simulated Action (GSA) framework proposed by Hostetter and Alibali (2008, 2019). From the framework, it can be predicted that speakers produce more and longer gestures when they activate visuospatial or motor simulations (for example in retelling a visual experience) and when they perform tasks that are more cognitively demanding. Those predictions fit in the comparison between the present study and Tellier and Stam's (2010): that gesture duration is longer in our study can be related to the fact that our task (i.c. describing complex visual scenes from memory) was more demanding and more explicitly tapping into visuospatial simulation than the task in Tellier and Stam (2010, i.c. explaining single words written on a piece of paper). In addition, and from a methodological point of view, differences in gesture duration might be the residue of annotation differences. For example, in our study gestures can get very long when they are repetitive, i.e. when the articulatory tension in the hands persist during a circular movement in which the hands go over and over the same trajectory. From the description of the gesture annotation procedure in Tellier and Stam (2010), it is unclear how cases such as these repetitive gestures are segmented and dealt with.

The overall proportion of gesture time in our experiment was well above 50%. This means that our participants were producing manual gestures more than half of the interactional time. Compared to other experiments in which participants had to describe words, objects or pictures from memory 50% is quite a lot. For example, Kamermans et al. (2019:136) report between 10% and 22% and Krauss (1998:57) reports around 30%. In experimental tasks in which visuospatial simulations are even more challenged, for example jointly designing the floor plan of an apartment, the proportion of gesture time can even amount to levels well above half of the time (cf. Bavelas and colleagues (2011:54) who report a soaring average of 82% of gesture time). When speakers are engaged in a task that already evokes such a high proportion of gesture time, they might have reached a gesture 'ceiling', making it difficult to even increase the gesture time to accommodate to non-native speakers. However, this conclusion remains very tentative and more research is needed to disentangle the influence of experimental task, non-native language proficiency, and other factors in accounting for the proportion of gesture time in FDG.

Our method of body pose estimation using OpenPose most prominently provided return on methodological investment in testing hypotheses 3 to 6 on gesture size, trajectory and speed. Regarding gesture size we found that horizontal but not vertical amplitude was larger in the NNS condition. In other words, NS perform gestures that are 'further out' but not 'higher up' when interacting with NNS (relative to when interacting with NS). We did, however, observe the vertical dimension to be relevant as well, in the sense that interacting with NNS appears to result in a larger vertical span (cf. Figure 6 that graphically highlights how gestures in the NS condition display a smaller vertical range than gestures in the NSS condition). These results tie in with what Bergmann et al. (2018) and Tellier & Stam (2010) found. The former show that when talking to less proficient interlocutors, participants produced more gestures that span multiple spaces (in terms of McNeill's gesture space) and the latter that participants perform more gestures in the periphery and extreme periphery. In the present study, we also observe an increase in gesture space in the NNS condition, but our measuring technique allowed us to break down gesture size into a horizontal and vertical dimension. It would be interesting to uncover whether these observations can be replicated and also hold for other experimental tasks or set-ups.

Bergmann et al. (2018) point out that their significant effect of gesture size was mainly due to the fact that participants produce more tiny gestures (i.c. gestures that remain within the same McNeillean gesture space) in the high proficiency group, and not because they produce more sizable gestures in the low proficiency group. This seems to be at odds with our observation that the significant effect of condition (i.e. larger gestures when interacting with NNS) does not occur because small gestures are abundantly present in the NS condition, but instead because large gestures are absent in the NS condition. This could be related to a difference in experimental task, with retelling visual scenes in our study versus explaining words in Bergmann et al. (2018) driving this subtle difference in results. Foreigner talk strategies involving FDG might, of course, depend on the task at hand and for explaining complex visual scenes (as was the case in the present study) participants might typically resort to occupying more gestural space. Again, these types of speculative interpretations require follow-up research.

We were able to confirm our hypotheses 5 and 6 and found that participants significantly increase the trajectory path of their gestures, and because the gesture duration did not differ between conditions, also the maximum velocity. Because to the best of our knowledge these features of gesture had not yet been studied in research on FDG, these results maintain to be replicated by others. The results do add to the general picture of our study and help to show the different gesture strategies speakers use when explaining complex images to NS compared to NNS. Overall, participants in the NNS condition do not use more gestures than in the NS condition. However they do use more expressive gestures, i.e. gestures that are larger in size (mainly in the horizontal dimension), that are more diverse (range from very low down to very high up) in the vertical dimension, that are performed faster and that cover a larger distance/trajectory.

An interpretation that more expressive gestures are performed as a means to provide less proficient speakers with extra handles to understand the message is definitely plausible, but as for now, it cannot be ruled out that the increased expressiveness is underpinned by or is a residue of verbal processes (cf. Morsella & Krauss (2004) or Suppes et al. (2015) who demonstrate that iconic-deictic gestures do not appear to facilitate comprehension in description). In this line of reasoning, more expressive gestures could be the result of an increase in cognitive load or speech planning, for example because speakers are considering multiple lexical alternatives or more consciously self-monitoring clause or sentence length (cf. Kita, 2000).

Whether the adaptations in the NNS condition are 'for the hearer' or (in part also) 'for the speaker' remains unclear from the existing literature. Because NS speakers adapt their verbal behaviour to NNS, the change in cognitive load that goes with it, might cause the observed changes in gestural behaviour (making them in part or in toto 'for the speaker'). On the one hand, more psycholinguistic research could shed light into this issue, on the other hand, it would be very useful to go beyond observing the behaviour of the NS and link that NS behaviour to attitudes and performance on the part of the NNS. As a consequence, in follow-up research we aim at linking the gestural behaviour of the NS to task performance of the NNS and to attitudes of the NNS towards the NS.

5 Conclusion

Research on verbal strategies of foreigner talk have drawn quite some scientific attention. The topic of foreigner-directed gesture is still very much unexplored territory. With the present paper, and relative to the limited amount of existing research on the topic, we have added kinematic detail in the analysis and extended the conversational context to task-based dyadic interaction. We found that native speakers do not display an overall higher gesture rate when talking to non-natives, compared to talking to other natives. They do produce gestures that are larger on the horizontal axis, more diverse in size on the vertical axis, cover a longer distance in the gesture space and (as a consequence of not having a longer duration) are performed with a higher peak velocity. Taken together, this paints a picture of foreigner-directed gestures as being more expressive. Given that related studies on gestural accommodation strategies were either based on rather idiosyncratic gestural behaviour (with only a few participants), based on different conversational settings (e.g. classroom interaction) or based on interactions in which non-natives possess different levels of proficiency in the language used, we remain in dear need of a lot more converging evidence to make robust claims on how speakers adapt their gesture use to conversational partners with which they do not share a native language. With the present study we contribute to scratching the surface of answering the question 'what do native speakers do when interacting with non-natives?', but nearly all of the work remains to be done to answer the even more interesting and relevant questions of what the underlying mechanisms are that lead to the accommodation behaviour we observe, and whether non-natives actually benefit from those foreigner-directed gestures. Questions we commit to tackling in future research.

Acknowledgements

We are very grateful to Sophie Fafchamps who was involved in recording the data. We also gratefully acknowledge the financial support received for this research by FWO-Flanders: the doctoral fellowship n° 1119321N granted to Valentijn Provė.

Declarations

Informed consent was obtained from all individual participants included in the study. This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the SMEC ethics committee at KU Leuven (15-04-2021/ G-2020-1622).

Data availability statement

The datasets generated during and/or analysed during the current study are available in the Zenodo repository, https://zenodo.org/badge/latestdoi/421998192.

References

Adams, T. (1998). *Gesture in foreigner talk.* [Unpublished doctoral dissertation]. University of Pennsylvania.

Azaoui, B. (2013, June 19–21). One Teacher, Two Instructional Contexts. Same Teaching Gestures? *Tiger: Tilburg Gesture Research Meeting*, 1–4, Tilburg, Netherlands.

Bavelas, J., Gerwing, J., Allison, M., & Sutton, C. (2011). Dyadic evidence for grounding with abstract deictic gestures. In G. Stam & M. Ishino (Eds.), *Integrating Gestures. The interdisciplinary nature of gesture* (pp. 49–60). John Benjamins Publishing Company. https://doi.org/10.1075/gs.4.05bav

Bell, A. (1984). Language style as audience design. *Language in Society*, *13*(2), 145–204. https://doi.org/10.1017/S004740450001037X

Bell, A. (2002). Back in style: Reworking audience design. In P. Eckert & J. Rickford (Eds.), *Style and Sociolinguistic Variation* (pp. 139–169). Cambridge University Press. https://doi.org/10.1017/CBO9780511613258.010

Bergmann, K., del Mar Cordero Rull, M., & Lugrin, B. (2018). A Pilot Study on Adaptive Gesture Use in Interaction with Non-native Listeners. *Proceedings of the 4th Workshop on Multimodal Analyses Enabling Artificial Agents in Human-Machine Interaction*, 26–32. https://doi.org/10.1145/3279972.3279978

Brand, R. J., Baldwin, D. A., & Ashburn, L. A. (2002). Evidence for “motionese”: Modifications in mothers’ infant-directed action. *Developmental Science*, *5*(1), 72–83. https://doi.org/10.1111/1467-7687.00211

Bressem, J., & Ladewig, S. H. (2011). Rethinking gesture phases: Articulatory features of gestural movement? *Semiotica*, 2011 (184), 53–91. https://doi.org/10.1515/semi.2011.022

Campisi, E., & Özyürek, A. (2013). Iconicity as a communicative strategy: Recipient design in multimodal demonstrations for adults and children. *Journal of Pragmatics*, *47*(1), 14–27. https://doi.org/10.1016/j.pragma.2012.12.007

Cao, Z., Simon, T., Wei, S. E., & Sheikh, Y. (2017). Realtime multi-person 2D pose estimation using part affinity fields. *Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition*, 1302–1310. https://doi.org/10.1109/CVPR.2017.143

Clark, H. H., Murphy, G. L. (1982). Audience design in meaning and reference. *Advances in Psychology*, *9*, 287–299. https://doi.org/10.1016/S0166-4115 (09)60059-5

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

Dashti, A. (2013). Interacting with domestic workers in Kuwait: Grammatical features of foreigner talk. A case study. *International Journal of the Sociology of Language*, *2013* (224), 63–84. https://doi.org/10.1515/ijsl-2013-0056

Debras, C. (2015, September). Stance-taking functions of multimodal constructed dialogue during spoken interaction. *Actes du Colloque Gesture and Speech in Interaction 4*. Nantes, France.

De Beugher, S., Brône, G. & Goedemé, T. (2018). A semi-automatic annotation tool for unobtrusive gesture analysis. *Language Resources and Evaluation*, *52* (2), 433–460. https://doi.org/10.1007/s10579-017-9404-9

Deppermann, A. (2015). When recipient design fails: Egocentric turn-design of instructions in driving school lessons leading to breakdowns of intersubjectivity. *Gesprächsforschung* 16, 63-101. http://www.gespraechsforschung-ozs.de/2015.html

Dornyei, Z., & Scott, M. L. (1997). Communication Strategies in a Second Language: Definitions and Taxonomies. *Language Learning*, *47*, 173–210. http://dx.doi.org/10.1111/0023-8333.51997005

Margić, B. (2017). Communication courtesy or condescension? Linguistic accommodation of native to non-native speakers of English. *Journal of English as a Lingua Franca*, *6*(1), 29–55. https://doi.org/10.1515/jelf-2017-0006

Faerch, C., & Kasper, G. (1983). *Strategies in Interlanguage Communication.* London: Longman.

Ferguson, C. (1971). Absence of copula and the notion of simplicity: a study of normal speech, baby talk, foreigner talk and pidgins. In: D. Hymes (Ed.), *Pidginization and Creolization in Language* (pp. 117–140) Cambridge University Press.

Ferguson, C. A. (1975). Towards a characterization of English foreigner talk. *Anthropological Linguistics*, *14*(1), 1–14.

Ferguson, C.A., & DeBose, C. E. (1977). Simplified Registers, Broken Language, and Pidginization. In A. Valdman (Ed.), *Pidgin and Creole Linguistics. Bloomington* (pp. 99–125). Indiana University Press.

Fischer, K. (2016). *Designing speech for a recipient.* John Benjamins Publishing Company.

Galati, A., Brennan, S. E., & Galati, A. (2014). Speakers adapt gestures to addressees' knowledge: Implications for models of co-speech gesture. *Language, Cognition and Neuroscience*, *29*(4), 435–451. https://doi.org/10.1080/01690965.2013.796397

Gasiorek, J. (2015). Perspective-Taking, Inferred Motive, and Perceived Accommodation in Nonaccommodative Conversations. *Journal of Language and Social Psychology*, *34*(5), 577–586. https://doi.org/10.1177/0261927X15584681

Gass, S., Mackey, A., & Pica, T. (1998). The Role of Input and Interaction in Second Language Acquisition Introduction to the Special Issue. *The Modern Language Journa*l, *82*(3), 299–307. https://doi.org/10.1111/j.1540-4781.1998.tb01206.x

Gass, S.M. (2003). Input and Interaction. In C.J. Doughty and M.H. Long (Eds.), *The Handbook of Second Language Acquisition* (pp. 224–255). Blackwell Publishing. https://doi.org/10.1002/9780470756492.ch9

Gerwing, J., & Bavelas, J. (2004). Linguistic influences on gesture's form. *Gesture*, *4*(2), 157–195. https://doi.org/10.1075/gest.4.2.04ger

Gullberg, M. (2011). Multilingual multimodality: communicative difficulties and their solutions in second language use. In J. Streeck, C. Goodwin, and C. LeBaron (Eds.), *Embodied Interaction: Language and Body in the Material World* (pp. 137–151). Cambridge University Press.

Gullberg, M., & Kita, S. (2009). Attention to speech-accompanying gestures: Eye movements and information uptake. *Journal of Nonverbal Behavior*, *33*(4), 251–277. https://doi.org/10.1007/s10919-009-0073-2

Hall-Lew, L., & Boyd, Z. (2020). Sociophonetic perspectives on stylistic diversity in speech research. *Linguistics Vanguard*, *6*(1), 1–11. https://doi.org/10.1515/lingvan-2018-0063

Holler, J., & Bavelas, J. (2017). Multi-modal communication of common ground. In R. B. Church, M. W. Alibali & S. D. Kelly (Eds.), *Why Gesture? How the hands function in speaking, thinking and communicating* (pp. 213–240). John Benjamins Publishing Company https://doi.org/10.1075/gs.7.11hol

Holler, J., & Wilkin, K. (2011). An experimental investigation of how addressee feedback affects co-speech gestures accompanying speakers' responses. *Journal of Pragmatics*, *43*(14), 3522–3536. https://doi.org/10.1016/j.pragma.2011.08.002

Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin and Review*, *15*(3), 495–514. https://doi.org/10.3758/PBR.15.3.495

Hostetter, A. B., & Alibali, M. W. (2019). Gesture as simulated action: Revisiting the framework. *Psychonomic Bulletin and Review*, *26*(3), 721–752. https://doi.org/10.3758/s13423-018-1548-0

Horton, W., & Gerrig, R. (2016). Revisiting the Memory-Based Processing Approach to Common Ground. *Topics in Cognitive Science*, *8*(4), 780–795. https://doi.org/10.1111/tops.12216

Issidorides, D., & Hulstijn, J. (1992). Comprehension of grammatically modified and nonmodified sentences by second language learners. *Applied Psycholinguistics*, *13*(2), 147–171. doi:10.1017/S0142716400005543

Kamermans, K. L., Pouw, W., Fassi, L., Aslanidou, A., Paas, F., & Hostetter, A. B. (2019). The role of gesture as simulated action in reinterpretation of mental imagery. *Acta Psychologica*, *197*, 131–142. https://doi.org/10.1016/j.actpsy.2019.05.004

Kim, S. H., & Elder, C. (2005). Language choices and pedagogic functions in the foreign language classroom: A cross-linguistic functional analysis of teacher talk. *Language Teaching Research*, *9*, 355–380. https://doi.org/10.1191/1362168805lr173oa

Kimbara, I. (2008) Gesture form convergence in joint description. *Journal of Nonverbal Behavior*, *32*(2), pp. 123–131, 10.1007/s10919-007-0044-4

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

Knoll, M. A., Johnstone, M., & Blakely, C. (2015). Can you hear me? Acoustic modifications in speech directed to foreigners and hearing-impaired people. *Proceedings of the Annual Conference of the International Speech Communication Association*, 2987–2990.

Krauss, R. M. (1998). Why do we gesture when we speak? *Current Directions in Psychological Science*, *7*(2), 54. https://doi.org/10.1111/1467-8721.ep13175642

Lee, Y. (2007). Third turn position in teacher talk: Contingency and the work of teaching. *Journal of Pragmatics*, *39*(6), 1203–1230. https://doi.org/10.1016/j.pragma.2006.11.003

Lew, A. W. M. (2016). Input for the second language classroom: Some innovations and insights. *The Forum*, *15*(2), 54–56. https://doi.org/10.7916/D84N0GHB

Loewen, S., Sato, M. (2018). Metacognitive Instruction Enhances the Effectiveness of Corrective Feedback: Variable Effects of Feedback Types and Linguistic Targets. *Language Learning*, *68*(2), 507–545. https://doi.org/10.1177/13621688211004639

Long, M.H. (1981). Input, interaction and second-language acquisition. *Annals of the New York Academy of Sciences*, *379*, 259–278. https://doi.org/10.1111/j.1749-6632.1981.tb42014.x

Lugrin, B., Bergmann, K., Eckstein, B., & Heindl, C. (2018). Adapted Foreigner-directed Communication towards Virtual Agents. *Proceedings of the 18th International Conference on Intelligent Virtual Agents*, 59–64. https://doi.org/10.1145/3267851.3267859

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

Mol, L., Krahmer, E., Maes, A., & Swerts, M. (2012). Adaptation in gesture: Converging hands or converging minds? *Journal of Memory and Language*, *66*(1), 249–264. https://doi.org/10.1016/j.jml.2011.07.004

Morsella, E., & Krauss, R. M. (2004). The role of gestures in spatial working memory and speech. *American Journal of Psychology*, *117*, 411–424.

Namboodiripad, S., Lenzen, D., Lepic, R., & Verhoef, T. (2016). Measuring conventionalization in the manual modality. *Journal of Language Evolution*, *1*(2), 109–118. https://doi.org/10.1093/jole/lzw005

Osada, N. (2003). Analysis of pause occurrence in three kinds of modified speech: Public address, caretaker talk, and foreigner talk. *Pan-Pacific Association of Applied Linguistics*, 7 (1), 77–123.

Ozyurek, A. (2002). Do Speakers Design Their Cospeech Gestures for Their Addressees? The Effects of Addressee Location on Representational Gestures. *Journal of Memory and Language*, *46*(4), 688–704. https://doi.org/10.1006/jmla.2001.2826

Papoušek, M., & Hwang, S. C. (1991). Tone and intonation in Mandarin babytalk to presyllabic infants: Comparison with registers of adult conversation and foreign language instruction. *Applied Psycholinguistics*, *12*(4), 481–504. https://doi.org/10.1017/S0142716400005889

Pouw, W., Trujillo, J.P., & Dixon, J.A. (2020). The quantification of gesture-speech synchrony: An overview and validation of video-based motion tracking. *Behavior Research Methods*, *52*, 723–740. https://doi.org/10.3758/s13428-019-01271-9

Poulisse, N. (1990). *The Use of Compensatory Strategies by Dutch Learners of English*. Foris Publications.

R Core Team (2020). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. https://www.R-project.org/.

Roche, J. (1998). Variation in Xenolects (Foreigner talk). *Sociolinguistica*, *12*, 117–139.

Ripperda, J., Drijvers, L., & Holler, J. (2020). Speeding up the detection of non-iconic and iconic gestures (SPUDNIG): A toolkit for the automatic detection of hand movements and gestures in video data. *Behavior Research Methods*, *52*(4), 1783–1794. https://doi.org/10.3758/s13428-020-01350-2

Rodriguez-Cuadrado, S., Baus, C., & Costa, A. (2018). Foreigner talk through word reduction in native/non-native spoken interactions. *Bilingualism*, *21*(2), 419–426. https://doi.org/10.1017/S1366728917000402

Rogers, S. L., Fay, N., & Maybery, M. (2013). Audience Design through Social Interaction during Group Discussion. *PLoS ONE*, *8*(2). https://doi.org/10.1371/journal.pone.0057211

Rogerson-Revell, P. (2010). "Can you spell that for us nonnative speakers?": Accommodation strategies in international business meetings. *Journal of Business Communication*, *47*(4), 432–454. https://doi.org/10.1177/0021943610377304

Rothermich, K., Harris, H. L., Sewell, K., & Bobb, S. C. (2019). Listener impressions of foreigner-directed speech: A systematic review. *Speech Communication*, *112*, 22–29. https://doi.org/10.1016/j.specom.2019.07.002

Sacks, H., Schegloff, E.A., & Jefferson, G. (1974). A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language*, *50*(4), 696–735. https://doi.org/10.1016/B978-0-12-623550-0.50008-2

Scarborough, R., Dmitrieva, O., Hall-Lew, L., Zhao, Y., & Brenier, J. (2007). An acoustic study of real and imagined foreigner-directed speech. *The Journal of the Acoustical Society of America*, *121* (5), 3044–3044. https://doi.org/10.1121/1.4781735

Seidlhofer, B. (2009). Common ground and different realities: World Englishes and English as a lingua franca. *World Englishes*, *28*(2), 236–245. https://doi.org/10.1111/j.1467-971X.2009.01592.x

Smith, S. W., Scholnick, N., Crutcher, A., Simeone, M. & Smith, W. R. (1991). Foreigner Talk revisited. In J. Blommaert & J. Verschueren (Eds.), *The Pragmatics of International and Intercultural Communication* (pp. 173–186). John Benjamins Publishing Company. https://doi.org/10.1075/pbns.6.3

Stec, K. (2012). Meaningful shifts: A review of viewpoint markers in gesture and sign language. *Gesture, 12*(3). 327–360. https://doi.org/10.1075/gest.12.3.03ste

Suppes, A., Tzeng, C., & Galguera, L. (2015). Using and Seeing Co-speech Gesture in a Spatial Task. *Journal of Nonverbal Behavior*, *39*(3), 241–257. https://doi.org/10.1007/s10919-015-0207-7

Tellier, M., & Stam, G. (2010). Découvrir le pouvoir de ses mains : La gestuelle des futurs enseignants de langue. *Spécificités et Diversité Des Interactions Didactiques : Disciplines, Finalités, Contextes*, 1–14, Lyon, France.

Trujillo, J. P., Simanova, I., Bekkering, H., & Özyürek, A. (2018). Communicative intent modulates production and comprehension of actions and gestures: A Kinect study. *Cognition*, *180*, 38–51. https://doi.org/10.1016/j.cognition.2018.04.003

Uther, M., Knoll, M. A., & Burnham, D. (2007). Do you speak E-NG-L-I-SH? A comparison of foreigner- and infant-directed speech. *Speech Communication*, *49*(1), 2 – 7. https://doi.org/10.1016/j.specom.2006.10.003

van Engen, K. J., Baese-Berk, M., Baker, R. E., Choi, A., Kim, M., & Bradlow, A. R. (2010). The wildcat corpus of native-and foreign-accented English: Communicative efficiency across conversational dyads with varying language alignment profiles. *Language and Speech*, *53*(4), 510–540. https://doi.org/10.1177/0023830910372495

Walsh, S. (2002). Construction or obstruction : Teacher talk and learner involvement in the EFL classroom. *Language Teaching Research*, *6*(1), 3–23. https://doi.org/10.1191/1362168802lr095oa

Warren-Leubecker, A., & Bohannon III, J. N. (1982). The Effects of Expectation and Feedback on Speech to Foreigners. *Journal of Psycholinguistic Research*, *11*, 207 – 215. https://doi.org/10.1007/BF01067564

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., Sloetjes, H. (2006). ELAN: a Professional Framework for Multimodality Research. *Proceedings of the Fifth International Conference on Language Resources and Evaluation*.

Wooldridge, B. (2001). "Foreigner talk": An important element in cross-cultural management education and training. *International Review of Administrative Sciences*, *67*(4), 621–634. https://doi.org/10.1177/0020852301674002

Yanfen, L., & Yuqin, Z. (2010). A study of teacher talk in interactions in English classes. *Chinese Journal of Applied Linguistics*, *33*(2), 76– 86.

Yule, G. & Tarone, E. (1990). Eliciting the performance of strategic competence. In R.Scarcella, E. Andersen & S. Krashen (Eds.), Developing Communicative Competence in a Second Language (pp. 179–194). Newbury House.

Appendix

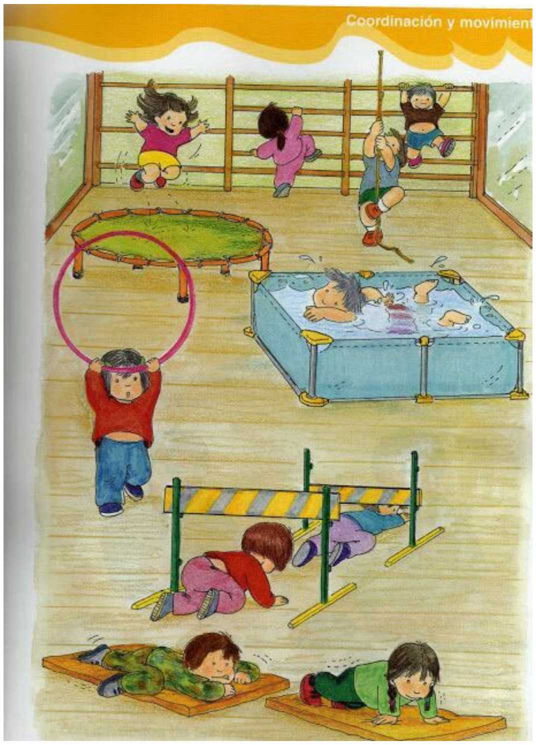

Figure 7. Illustration of the stimulus 'gym', version for the director to describe.

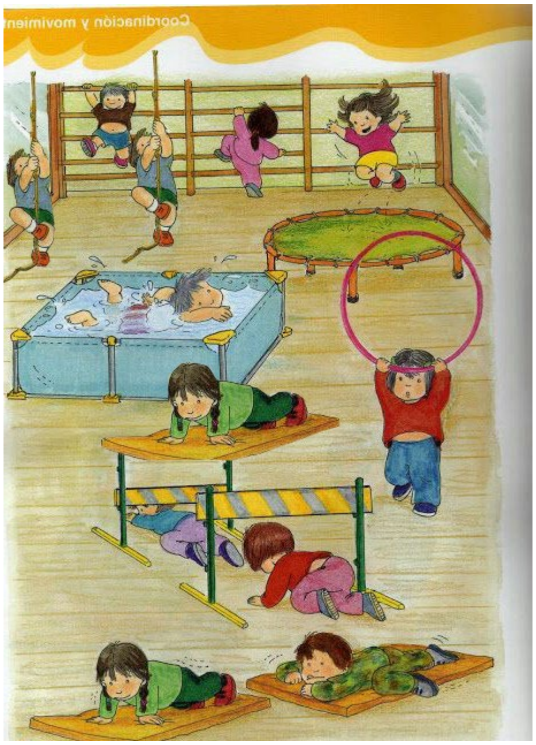

Figure 8. Illustration of the stimulus ‘gym’, alternative version for the matcher.

Figure 9. Illustration of the stimulus ‘forest’, version for the director to describe.

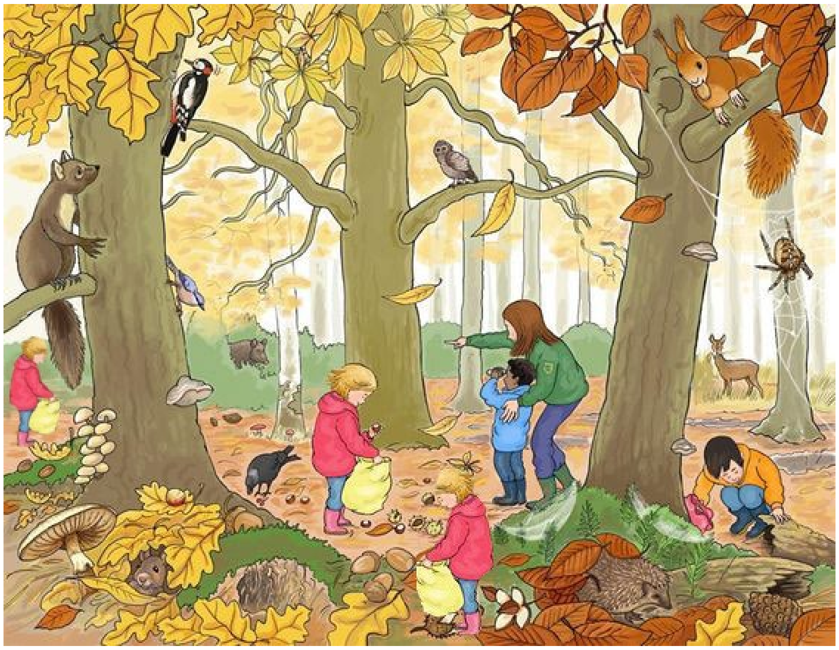

Figure 10. Illustration of the stimulus ‘forest’, alternative version for the matcher.

Figure 11. Illustration of the stimulus ‘lake’, version for the director to describe.

Figure 12. Illustration of the stimulus ‘forest’, alternative version for the matcher.

Figure 13. Illustration of the stimulus ‘playground’, version for the director to describe.

Figure 14. Illustration of the stimulus ‘playground’, alternative version for the matcher.