“XR Mark Test” Reveals Sensorimotor Body Representation in Toddlers

Michiko Miyazaki[1]*, Tomohisa Asai[2], Norihiro Ban[3] and Ryoko Mugitani[4]
[1] Department of Social Information Studies, Otsuma Women’s University, Tokyo, Japan
[2] Advanced Telecommunications Research Institute International (ATR), Kyoto, Japan
[3] HEIMEI Corporation, Kanagawa, Japan
[4] The Faculty of Integrated Arts and Social Sciences, Japan Women’s University, Tokyo, Japan

*Correspondence: myzk@otsuma.ac.jp

Abstract

The mark test is a popular test for self-recognition. Although the extent to which self-recognition can be assessed remains controversial, the test elicits visually guided, self-oriented, and spontaneous reaching movements. In this study, we demonstrated that this self-oriented reaching is suitable for estimating sensorimotor body representation in toddlers. We developed a non-verbal task (Bodytoypo) to assess the localization of body parts by gamifying the mark test and conducted it with thirty 2- and 3-year old children. Specifically, we detected the children's skeletal data in real-time, displayed virtual marks on various parts of their body, and estimated their reaction time and accuracy of body part localization. Subsequently, a statistical-based automated analysis using 2-D image processing and conventional frame-by-frame coding were performed. The results revealed developmental changes in the children’s reaching strategies. A few errors were observed around the face. A reduction in the error rate for joint and movable areas was observed in children between the ages of 2 and 3 years. An analysis of movement trajectories using a combination of image processing and machine learning algorithms showed that 2-year-olds acquired visually guided reaching (feedback control) from ballistic exploratory reaching and 3-year-olds acquired rapid and predictive reaching (feedforward control) from visually guided cautious reaching. It was also found that the accuracy of localization could be predicted by examining the coordination of body parts. Evaluation of the developmental changes in self-oriented reaching reveals new possibilities for the mark test and development of body representation.

Keywords

toddlers, mark test, sensorimotor body representation, No, gamification, augmented reality, Openpose, internal model, feedback control, feedforward control

Introduction

The mark test is a classical test for measuring “self-recognition” and it has been featured in numerous studies over the past 50 years on a variety of species [1–5]. When facing a mirror, participants will try to touch or remove a mark secretly placed on their face. Researchers have argued that passing the test could be a measure of self-recognition. Although this interpretation remains controversial, the mark test prompts participants to make visually guided and self-oriented reaching movements. The importance of these hand movements occurring naturally has been ignored until recently.

Observing these movements and clarifying the cognitive processes behind them may lead to new discoveries about children's body representation. In previous studies, an insightful error was discovered when mark tests were applied using mirrors or video images and some toddler participants tried to locate the mark behind their head while it was placed on their forehead. We named this curious initial search error the "rear-search error" [6,7]. In a mirror mark test, 38% of 2-year-olds and 11% of 3-year-olds demonstrated the rear-search error initially [6]. With a live condition, 55% of 3-year-olds and 38% in a 2-second delay condition demonstrated the error (these findings prompted reanalysis [8]). In 2-year-olds, the video mark test had pass rates of less than 20% compared to the mirror mark test [8,9]. Is this error similar to the distortions of body perceptions observed in adults? [10,11] Or is the error reflective of a specific representation of body perception in this age group (Other phenomena that have the possibility of specific bodily representation are scale errors [12] and tadpole human drawings [13,14])?

Examining the accuracy of reaching for various body parts may help us to clarify developmental changes in sensorimotor body representation. Detailed trajectory analysis for the whole body should specifically reveal the development of sensorimotor internal models (e.g., weight shift between feedforward-back motor control) that represent the relationship between sensory inflow (mark detection) and motor outflow (reaching).

There are some technical difficulties to be solved. First, the experimental procedure, especially for trial repetition, must be sophisticated. We updated the gamification procedure [15] with AI-based online bone estimation using Microsoft Kinect (an infrared sensor device) to Openpose (an image processing library) [16]. In this study, we used Openpose to extend the targeted body parts from the face to the whole body. Although Openpose detects bones from two-dimensional RGB images, it has been reported that its detection is reliable even for infants with much smaller body sizes [17]. As a result, we implemented augmented reality (AR) to display virtual marks in real-time on various body parts. When the marks were located accurately by participants, a pleasant visual and auditory reward was presented, which was expected to keep their motivation to repeat trials. In this sense, Bodytoypo is an interactive “XR” mark test where AR marks “can be touched,” crossing the real-virtual boundary. Second, a quantified evaluation of the reaching movements is essential. Recently, several studies using image processing of 2D images for motion trajectory analysis have been reported [17]. In the study of infants, the advantage of detecting motion trajectories from image

processing without markers is significant. In this study, we conducted an offline analysis using AI statistics. Because these factors are tightly connected with each other, we developed a collaborative work among engineering (online task processing), machine learning statistics (offline analysis of full-body motion), and developmental psychology (administering experiments).

With the ability to repeat dozens of trials using AR, we are now able to quantitatively evaluate the accuracy and speed of localization for each body part. In other words, body parts that are easy and difficult to localize can be plotted topologically. Proprioceptive body images are perceived as distorted, even in adults [10,11]. We investigate whether the distortions observed in adults are observed from childhood, or whether there is a specific mode of perception in childhood. The relationship with the acquisition of language related to the body will also be examined.

Results

Experiment

Method

Participants

Thirty-six children aged 2.5–3.5 years old participated in this study. The final sample consisted of 30 children (mean = 34.7 months, range: 28–44 months, SD = 4.49), 18 2.5-year-olds (11 females), and 12 3.5-year-olds (9 females). Six participants were excluded from the analysis (attrition rate = 16.7%) owing to fussiness or shyness (n=5) or machine trouble (n =1). This study was carried out in accordance with the recommendations of the Otsuma Women's University's Life Sciences Research Ethics Committee and written informed consent was obtained from all participants' parents. The study protocol was approved by the Otsuma Women's University Life Sciences Research Ethics Committee (2019-012).

Apparatus

We developed a body part localization task using augmented reality (AR) with the image processing library Openpose [16]. We named this system "Bodytoypo." An image of each participant's whole body was recorded using a USB camera (Logicool C920) and displayed via a projector (Epson EB 485WT) onto a screen (KIMOTO RUM60N1) (see Figure 1: it was presented in a mirror-like or ipsilateral relationship). The 2D coordinates of their body parts were automatically detected using Openpose on a GPU machine (Mouse computer NEXTGEAR i690PA2). The detected skeletal data were processed in real-time so that the AR character (favored cartoon characters recognized by children) was displayed as a mark on the coordinates of the target body parts. The

net delay (transmitting and image processing) was approximately 10 frames (0.33 s), and the processing speed of the AR overlay was 15–22 fps (best-effort format). Participants stood on a sign 115 cm away from the screen. The angle of the camera was adjusted so that the entire mirrored-like body was reflected on the screen. Each participant's behavior was recorded using a capture device (Avermedia AVT-C878) on a stand-alone PC (Galleria QSF1070HGS). The recorded video was further used for offline analysis both for manual coding and movement analysis.

Tasks

Body part localization task (Bodytoypo):

The participants were prompted to touch the AR marks like in a traditional mark test with a mirror. The AR marks were displayed individually on 30 body parts (upper forehead, forehead, eye (R/L), ear (R/L), nose, cheek (R/L), mouth, chin, collarbone (R/L), shoulder (R/L), upper arm (R/L), elbow (R/L), navel, forearm (R/L), hand (R/L), thigh (R/L), and big toe (R/L)). Note that the 30 body parts shown in Bodytoypo were based on the Openpose definition, which was slightly different from those in the independent offline analysis (see also Figure 1C). The order of the marks was randomly arranged in a sequence to avoid adjacent body parts. These sequences were presented in forward and reverse directions. The presentation order was counterbalanced among participants.

The experimenter was observing whether the participants could touch (or point at) the actual body part corresponding to the mark position correctly. If they could touch the correct body part, a cheerful audio-visual reward was presented by the experimenter’s manual key control. If they had trouble responding to the task (e.g., lost their motivation), the experimenter re-presented the mark or skipped the trial. In total, we prepared 30 trials for each participant.

Questionnaire of body part vocabulary

After the task, we asked parents about the vocabulary acquisition status of their children using a questionnaire. The parents were asked about their children’s responses to speech and comprehension of 60 body parts and body-related vocabularies in Japanese (e.g., tail, feathers, and horns).

Procedure

The experimenter first explained the study to the parents to obtain informed consent. The experimenter and participating children built rapport through free playing. Then, she demonstrated a 10-trial game of Bodytoypo the child. During the demonstration, she deliberately mistouched the mark, which would not give any reward. When participants were suitably positioned (standing straight and still in front of the screen) an AR mark would appear with a beep sound. The experimenter encouraged the child to participate in the task by prompting them in a rhythmic manner. If the participants correctly touched the body part with the AR mark, the mark disappeared with an audio-

visual reward (fun images and sound). If participants lost their motivation, the nose trial (an easy motivation catcher) was initiated. Aside from those additional nose trials, a maximum of 30 trials (30 body parts) was prepared for each participant and further analyzed. The experiment was completed within 5 minutes.

Analysis

We performed both a manual coding analysis (conventional frame-by-frame visual coding) and statistics-based automated analysis (the offline detection, change point detection of hand motion, and body-parts association in terms of predicting touch errors). The former by multiple coders was conducted to analyze the number of trials engaged, error rate (overall and by each body part), and the relationship between the error rates and word acquisition of body parts. The latter was conducted to analyze the participants' sensorimotor coordination among body parts during the Bodytoypo task, as a type of visually guided reaching motion.

The main focus of these analyses was the period between the mark appearing and children's first touch on each trial. By examining the accuracy of the first touch (simply, the location of touch and reaction time), we expected to reveal the child's body representation and its distortions.

Results

Manual analysis with frame-by-frame visual coding

Number of executed trials

Overall, the average number of executed trials was 27.4 for the 2-year-olds and 29.8 for the 3-year-olds. Most children successfully completed 30 trials, regardless of age ($t(28) = –1.57$, $p = .13$, n.s.). This result is consistent with our previous study [15]. This suggests that we were able to maintain the motivation of the participants during the whole-body version. It seems that the positive effect of gamification (i.e., producing a pleasant animation and sound) was demonstrated.

Error rate

The overall error rate for the first touch was 40.1% for the 2-year-olds and 35.0% for the 3-year-olds. There was no significant difference between age groups ($t(28) = 1.12$, $p = .27$, n.s.). Figure 2 summarizes the mean error rates for each age and body part (Figure 2A) and in mark locations (midline, left side, and right side; see Figure 2B). We examined the differences in mean error rates by combining body parts by mark locations [18] and found that the error rate for midline body parts was significantly lower regardless of age. The main effect of age was not significant ($F(1,28) = .83$, $p > .05$, effect size $f = .17$). However, the main effect of mark location was significant ($F(1,28) = 35.65$, $p < .01$, effect size $f = 1.13$). Holm's multiple comparison showed that the error

rates in the central location were significantly lower than those in other body parts (*ps* = .05).

Error rates by bodily function

Next, we exploratively analyzed the error rates of each body part combined by bodily function rather than mark location-based. The 30 body parts were re-categorized by bodily function and are summarized in Figure 3. A two-way ANOVA (age and bodily function) was used to explore whether the error rates were different. The four bodily functions examined were whether the body part was a joint, a moveable part, a face, or a named body part. The main effects of all bodily functions were confirmed (main effects of joint: $F(1,28) = 12.26$, $p < .01$; moveable: $F(1,28) = 25.51$, $p < .01$; face: $F(1,28) = 104.94$, $p < .01$, and named body part: $F(1,28) = 58.48$, $p < .01$). In addition, the interaction of age and bodily function was significant for whether it was a joint and whether it was a moveable part (age * joint: $F(1,28) = 4.90$, $p < .05$; age * moveable: $F(1,28) = 4.03$, $p < .10$). Therefore, these results suggest that the accuracy of the first touch increases between the ages of 2 and 3 years for parts that are joints and parts that can be moved.

Error rates and word acquisition of body parts

The questionnaire results revealed that for the 60 words asked, 28.7 (SD: 10.6) (2-year-olds) and 40.8 (SD: 14.6) (3-year-olds) of the words on average were able to be said as well as comprehended (see also Supplementary Figure S2C). There was a significant difference in the mean number of words acquired between the age groups ($t(28) = –2.63$, $p < .01$).

We examined whether the error rate of reaching decreases if the child has acquired more body part names. A partial correlation analysis was conducted with age in months as a control variable. However, this tendency was not observed ($r(27) = .148$, $p = .45$, n.s. Supplementary Figure S2A). We also conducted a correlation analysis between the vocabularies that were easy to produce and the error rate, but no significant difference was found in this analysis ($r(21) = -.346$, $p = .106$, n.s. Supplementary Figure S2B).

Result of statistics-based automated analysis

The overall pipeline.

The necessary process was first the change of point detection since each trial (from the mark appearing to the first touch) consisted of two periods: response latency and reaching movement (Figure 4). For this purpose, the change point was estimated by using the machine-learning R library “changepoint” (https://github.com/rkillick/changepoint/) over the time series of variance among distances between the used hand and other body parts (Figure S4). During a reach,

these distances change dramatically compared with those during latency. This automated estimation (separation of latency and reaching) was further validated in a later process (Figure 5ABC and 6AE). Once we distinguished them, the whole-body associations were calculated as a temporal-correlation matrix during those two periods (Figure 5), which indicates sensorimotor coordinates for touching the AR mark. This summarized matrix exhibits a relative similarity of trajectory among body parts so that the original configuration (physical body) was partly recovered from the matrix through multidimensional scaling (MDS) projection during latency (Figure 6A), but not during reaching (Figure 6E). Furthermore, this correlation matrix was further vectorized and entered into the regression analysis (as inputs) to predict children's touch errors (as outputs) (Figure 6BD). During each period, body part associations were visualized at a glance in terms of the x-y joint dendrogram (tanglegram, Figure 6CF). Finally, all variables as continuous variables were put into a linear mixed-model (all-in-one statistical model) to predict touch error (Figure 7A). To summarize, we developed that 2- to 3-year-olds' sensorimotor representation can be depicted where the feedback-forward weight for touching seems to be adaptively changing.

Body-parts association during latency and reaching

Figure 5A suggests the averaged temporal correlation among 30 body parts (29 parts except for the reference). The upper triangles are for x-, while the lower triangles are for y-coordinates where some "clusters" are observed on the diagonal (e.g., "face" cluster). The body parts (Figure 5A) or clusters (parts-averaged, Figure 5BC) were positively correlated in general, except for L-R upper body associations in the reaching period (blue squares). This negative correlation in the x-coordinate suggests bi-manual coordination in reaching behaviors that could validate the change point estimation (see Figure S4). Both aged groups behaved similarly in terms of these correlation matrices (Figure 5BC) as well as of the reaching trajectory in 2D (Figure S5). However, the "face" (or head) behaved oppositely between 2- and 3-year-olds if we see the subtracted matrix between correct and incorrect trials (Figure 5D) or the cross-correlation between the face and the hand (Figure S7A). We observe what happened for those children as follows:

Correlation matrix for configuration, prediction, and coordination

Figure 6A shows the recovered body configuration from the physical positions (averaged x-y coordinates of all trials, the upper panel) to the MDS projection (based on the averaged temporal correlation matrix among body parts and the lower-left panel) during latency. The comparison between them (through the Procrustes transformation in the lower right panel) suggests that the body parts association during latency is relatively stable so that we can see the original body configuration, although it is horizontally biased since the hand easily moved horizontally even during latency. This is not the case for reaching even where the average position was the same (Figure 6E, upper panel). Even though the correlation matrix was also largely the same as during

latency (see Figure 5A), the original body configuration was not recovered (Figure 6E in the lower-left and right panels) because various reaching movements were included which could validate the change point estimation again.

Although the correlation matrix between 2- and 3-year-olds was congruent (Figure 6BD, reprinted from 5BC), there was a statistical difference in predicting children's touch errors (overlaid asterisks). The correlation matrix was vectorized and entered into a linear regression model to predict the touch error distance. For 2-year-olds, the latency matrix as the assembled dataset (regardless of the participants or target positions at this stage) can predict touch error ($F(30,436)=1.666$, $p=0.016$), while the reaching matrix can for 3-year-olds ($F(30,308)=1.636$, $p=0.022$). On the other hand, the latency matrix for 3-year-olds or the reaching matrix for 2-year-olds cannot predict their touch errors ($F(30,288)=1.475$, $p=0.057$; $F(30,435)=1.19$, $p=0.229$, respectively). For each model, the significant (and marginal for reference) variables are indicated by symbols (Figure 6BD, $p<0.01$**, $P<0.05$*, $p<0.10$·) with signs (positive or negative effects), which included the L-R upper body associations (i.e., hand movements) in x-coordinates regardless of age.

Figure 6CF further depicts the body parts association in tanglegram (R library "dendeextend" (https://talgalili.github.io/dendextend/)) in terms of "correct" or "incorrect" labels by the manual codings. For the latency (Figure 6C), the incorrect matrix for 2-year-olds is characterized where Rupper-Rlower parts are associated (blue ellipse), while the other tanglegrams suggest that the Rupper cluster was independently moving (yellow rectangle). For reaching (Figure 6F), the incorrect matrix for 3-year-olds is characterized where the Rupper-Rlower parts are associated again (blue ellipse), while the other tanglegrams suggest that the Rupper and Face clusters were associated in x-coordinates (yellow rectangle) (see also Figure S7A for the cross-correlation between the face and hand). If face movement is critical for precise touching (see also Figure 5D again), this could be related to their feedback/forward control in the AR mark test [19]. This might have emerged as the weight between latency/reaching duration as a function of feedforward/back control strategy between both age groups as follows.

The weight between feedback/forward strategy changes among ages

So far, the Bodytoypo task mainly produced participants' touch errors and latency/reaching duration (Figure 7B) as well as body-parts association. These depend on the target location (see Figures 5 and 6). This specifically varies the initial mark distance between the hand used and the target location when the mark appeared (Figure S6 for a descriptive correlation matrix among variables). Although categorical comparisons, such as 2- or 3-year-olds or correct/incorrect trials, have been useful to depict what was happening in the current task qualitatively during latency or reaching, the all-in-one statistics with a linear mixed model attempted to conclude these relationships in a continuous manner with some other covariates as mentioned above. For that purpose, the best-fit model was explored using the R library ("lme4" (https://github.com/lme4/lme4/) and "lmerTest" (https://github.com/runehaubo/lmerTestR), where the initial inputs were latency duration, reaching duration, trial repetition, initial mark distance, and participants' age in

months as fixed effects as well as participants' ID and target labels as random effects, while the touch error value was the output. This every-effect-model produced significant interaction effects (e.g., the interactions among reaching, age, repetition, and initial distance [t(422.5)=2.721, p=0.0068]). The step function further identified the model with the minimum Akaike information criterion (AIC). This selected-effect-model still produced a significant interaction among latency/reaching, age, repetition, and initial distance (t(557.0)=2.30, p=0.0218 for latency; t(554.0)=2.38, p=0.0177 for reaching). Therefore, Figure 7A suggests the relationship among latency, reaching, and age in months in terms of predicting the touch errors (see Fig. S7B for other interactions when total RT was used instead of latency/reaching duration).

For younger months, both latency and reaching should be longer for correct touch (minimized touch error), although there was a robust anti-correlation between latency and reaching duration in general (Supplementary Figure S6). For older months, however, these should be shorter for correct touch. This might imply that younger months rely on slower feedback control when correct. Accordingly, if they administer a quick movement without visual feedback (i.e., ballistic exploration), they might make mistakes (see Figure 7C, as we see below, latency ratio did not vary for incorrect trials for 2-year-olds). When several months developed they no longer exhibit ballistic movements. The older months may rely on feedforward control with a shorter latency/reaching when correct. Accordingly, if they exhibit a longer reaching as feedback control, they may mistouch. This may sound conflicting with a conventional understanding of feedback/forward weight with development [20]. Given that visually guided reaching through a mirror (e.g., mirror drawing [21]) can confuse even an adult, more weight on feedback control for the AR mark test is not the optimal strategy. Indeed, the "trial repetition" factor suggests that, for older months, the repetition increased their touch errors presumably due to habituation-driven exploration or playing with feedback control (Figure S7B, the upper panels).

The weight between latency/reaching duration is depicted in Figure 7B, in relation to the touch error and correct/incorrect labels. During a reach, the duration distribution peaked for 2-year-olds in incorrect trials and peaked for 3-year-olds in correct trials. When the relative duration of latency (the ratio for total RT) was summarized between 2- and 3-year-olds (Figure 7C, a contour plot for visualization), the target-dependent variance of the latency ratio might be observed for correct trials with 2-year-olds, and incorrect trials with 3-year-olds, which could be measuring their feedback weight. On the other hand, for the correct trials in 3-year-olds, the initial mark distance was positively correlated with the latency ratio (Figure S6) and interacted with the total RT for touch errors (Figure S7B, the lower panels). The hand was synchronized with face movements in y-coordinates (Figure S7A), which could be measuring their feedforward weight. The best strategy for the Bodytoypo task should be the optimal integration of feedback/forward control depending on the target location (e.g., initial mark distance). The current results suggest a developmental weight shift from ballistic explorative movements (incorrect trials) to a feedback-based slow control (correct trials) for 2-year-olds, and from a feedback-based slow control (incorrect trials) to a feedforward-based fast control (correct trials) for 3-year-olds. The complex but optimal weight between the internal forward output and external feedback input could

be revealed by the current task. We now go back to visual observations over the original videos at last, according to the statistical results.

Qualitative re-evaluation of the videos based on automated analysis with manual analysis

Automated analysis revealed developmental changes in the reaching strategies of toddlers. To demonstrate the validity of this analysis, we applied the idea to both automated analysis and qualitative manual analysis. First, we examined whether the three reaching strategies found in the automated analysis were supported by manual analysis. Two coders re-watched the videos of each trial and classified them as ballistic exploration, feedback control (FB), or feedforward control (FF). Ballistic exploration is when the participant reaches for a mark quickly without aiming. Feedback control is a slower reaching strategy where participants adjust their hand movements while watching a video monitor or their actual body parts. Feedforward control is a strategy of quickly reaching the target with a prediction of localization. All trials were classified and 32.5% of the total trials were calculated with a concordance rate (kappa =0.33). As a result, the concordance rate was low and a difference in judgment of approximately 25% was observed, especially for the classification of ballistic exploration, FF, and FF and FB. This suggests the difficulty of classifying reaching strategies at a glance. The fact that reaching strategies switched during the same trial or that responses were ambiguous may have reduced the concordance rate.

During the qualitative re-evaluation, an interesting indicator was found; the difference in hand shape used for reaching. For example, touching a mark with a pointed finger was likely to be more confident in the accuracy of localization and less likely to be combined with ballistic exploration. Touching extensively with the hand indicated a lack of confidence in localization or an attempt to increase the likelihood of a correct response by covering a large area. We focused on the hand shape during a reach and classified whether the participants reached with their hand (hand) or pointed with one finger (index). As a result, 735 trials (87.7%) were classified as reaching with the hand, and 103 trials (12.3%) were classified as reaching with the index. The concordance rate calculated for 32.5% of the trials in the entire sample was also high (kappa = 0.83). To evaluate the relationship between hand shape and reaching strategies, we analyzed only trials in which the judgments were consistent among coders (237/275 trials). There were 237 trials of hand use in which the classification between coders agreed, and more than 80% of the trials were FF reaching, regardless of age or correct/incorrect answers (see Supplemental Figure S3A). There were 26 trials of index used with consistent classifications among coders. The 3-year-olds were all FF reaching, and the 2-year-olds were a mixture of FB and FF reaching. There was no index used in ballistic exploration at either age (see also Supplemental figure S3B). In addition, the percentage of index finger use was particularly high when the target was placed on the nose, which validates the above discussion.

Altogether, the three typical reaching strategies were suggested in the automated analysis. However, the three strategies are not switchable, but rather

changes in the online weighting of the internal model, so they are occasionally mixed. Therefore, it was clarified that complete "classification" would be difficult. Meanwhile, the shape of the hand, especially the way the index was used, was found to be related to reaching strategies and developmental changes.

Discussion

The purpose of this study was to demonstrate the potential of the classical mark test as a window for revealing children's sensorimotor body representation through visually guided, self-oriented, and spontaneous reaching. Our XR mark test with a machine-learning 2D-bone estimation in real-time (Bodytoypo task) clarified the developmental changes of participants' reaching strategies, from ballistic explorative movements to an optimal integration between internal output and external input. This is in line with our previous observation about rear-search errors in the classical mark test, which disappeared in 2- to 3-year-olds.

We examined whether the mark test can be used to quantitatively evaluate their sensorimotor representation. We obtained sufficient task repetitions regardless of age. We were able to analyze the full body-parts associations and the assembly distribution of reaction time in terms of touch accuracy. This was due to our gamification policy with a real-time and interactive XR implementation where the automated skeletal detection and the experimenters' skillful manual procedure were combined, suggesting a rich collaboration between the updated technology and a classical but still ingenious task for self-body recognition.

The mark test or Bodytoypo is a task that can be intuitively executed by participants without verbal instruction. Therefore, it is important to note that Bodytoypo could be completed not only by toddlers or adults but also by other species with adequate updates.

Our second purpose was to show the developmental changes in toddlers' body representations since their mistouch patterns during the mark test can be insightful (e.g., rear-search error). However, it was challenging to analyze toddlers' motions. For example, in another study, a relatively large number of participants were excluded because they removed 3D markers by themselves (attrition rate: 31.8%) [22]. Recently, the bone was detected through 2D RGB images in real-time as our task was implemented. In addition to this online processing, we also conducted an offline re-estimation of participants' bones for the recorded videos as accumulated 2D images because the online process was based on a best-effort format (i.e., changeable). This was motivated by a previous study that revealed toddlers' motor behaviors in similar offline motion analysis with their natural dynamics without wearing any devices [17]. As a result, we were able to show evidence of delicate changes in their reaching strategies where a preprocess of the children's noisy data was inevitable at first. The estimated change point between latency and reaching finally suggested their optimal integration between feedback/forward weights to correctly touch the mirror-reversed body parts.

We believe that the changes from ballistic exploratory movements to feedback and/or feedforward control between the ages of two and three years provide an

important insight into the acquisition of body representation, especially regarding the role of visual information in their sensorimotor control. In contrast to ballistic exploratory movements, reaching without modifying the motion trajectory and feedback control requires online visual adjustment (through a mirror in the current case) while referring to the learned self-body representation. This suggests that body part localization (i.e., touching or pointing) through ballistic exploratory movement weighs more on proprioceptive information, while that by FB/FF control should be an integrated process between proprioceptive and visual information [23]. In particular, 3-year-olds exhibited an optimal integration for correct touch where the FF might have more weight than FB since mirror visuals (based on FB control) can confuse even adults. The use of the index finger, which was observed only during FF, may also reflect appropriate motor prediction. This indicates the development of our sensorimotor integration with reliability among proprioceptive, visual, and motor predictions. The weights should always be changed depending on the circumstances or tasks. Our results suggest quick adaptability for 3-year-olds since latency and reaching were rapid while utilizing motor prediction (FF weights for correct touch).

Taken together, in addition to integrating vision and proprioception, children gradually acquire the ability to control their own body as an object or representation in a predictive manner. The classical mark test would still be useful in tandem with recent technical advantages, as the current study indicates, which enlightens toddlers' sensorimotor internal models.

Acknowledgements

We thank Ayano Ryoke and Moe Kobayashi for their help with the data collection and analysis. We would also like to thank all caregivers and children who participated in this study. This work was supported by JSPS KAKENHI Grant Numbers 19H04019 to Michiko Miyazaki, Tomohisa Asai, and Ryoko Mugitani.

Author Contributions

All authors contributed to the study concept, design, and interpretation.
N.B programmed the Bodytoypo task. M. M. collected the data. T.A. developed and performed the analyses. All authors wrote the manuscript and approved the final version of the manuscript for submission.

Declaration of Interests

The authors declare no competing interests.

References

1. Gallop, G.G., Jr (1970). Chimpanzees: self-recognition. Science *167*, 86–87.

2. Amsterdam, B. (1972). Mirror self-image reactions before age two. Dev. Psychobiol. *5*, 297–305.

3. Lewis, M., and Brooks-Gunn, J. (1979). Social Cognition and the Acquisition of Self (Springer).

4. Suddendorf, T., and Butler, D.L. (2013). The nature of visual self-recognition. Trends Cogn. Sci. *17*, 121–127.

5. Anderson, J.R., and Gallup, G.G., Jr (2015). Mirror self-recognition: a review and critique of attempts to promote and engineer self-recognition in primates. Primates *56*, 317–326.

6. Miyazaki, M., and Hiraki, K. (January 7 2017). Does rear-search error in the mark test indicate a uniqueness of body-representation in young children?

7. Imaizumi, S., Asai, T., and Miyazaki, M. (2021). Cross-referenced body and action for the unified self: empirical, developmental, and clinical perspectives. In Body Schema and Body Image: New Directions (Oxford Scholarship Online), p. 194.

8. Miyazaki, M., and Hiraki, K. (2006). Delayed intermodal contingency affects young children's recognition of their current self. Child Dev. *77*, 736–750.

9. Suddendorf, T., Simcock, G., and Nielsen, M. (2007). Visual self-recognition in mirrors and live videos: evidence for a developmental asynchrony. Cogn. Dev. *22*, 185–196.

10. Mora, L., Cowie, D., Banissy, M.J., and Cocchini, G. (2018). My true face: unmasking one's own face representation. Acta Psychol. *191*, 63–68.

11. Fuentes, C.T., Longo, M.R., and Haggard, P. (2013). Body image distortions in healthy adults. Acta Psychol. *144*, 344–351.

12. DeLoache, J.S., Uttal, D.H., and Rosengren, K.S. (2004). Scale errors offer evidence for a perception-action dissociation early in life. Science *304*, 1027–1029.

13. Freeman, N.H. (1975). Do children draw men with arms coming out of the head? Nature *254*, 416–417.

14. Cox, M.V. (2013). Children's Drawings of the Human Figure (Psychology Press).

15. Miyazaki, M., Asai, T., and Mugitani, R. (2019). Touching! an augmented reality system for unveiling face topography in very young children. Front. Hum. Neurosci. *13,* 189.

16. Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E., and Sheikh, Y. (2021). OpenPose: realtime multi-person 2D pose estimation using part affinity fields. IEEE Trans.

Pattern Anal. Mach. Intell. *43*, 172–186.

17. Ossmy, O., and Adolph, K.E. (2020). Real-time assembly of coordination patterns in human infants. Curr. Biol. *30*, 4553–4562.e4.

18. Auclair, L., and Jambaqué, I. (2015). Lexical-semantic body knowledge in 5- to 11-year-old children: how spatial body representation influences body semantics. Child Neuropsychol. *21*, 451–464.

19. Shen, H., Baker, T.J., Candy, T.R., Yu, C., and Smith, L.B. (2010). Using the head to stabilize action: reaching by young children. In 2010 IEEE 9th International Conference on Development and Learning (ieeexplore.ieee.org), pp. 108–113.

20. Bushnell, E.W. (1985). The decline of visually guided reaching during infancy. Infant Behav. Dev. *8*, 139–155.

21. Starch, D. (1910). A demonstration of the trial and error method of learning. Psychol. Bull. *7*, 20.

22. Kahrs, B.A., Jung, W.P., and Lockman, J.J. (2014). When does tool use become distinctively human? hammering in young children. Child Dev. *85*, 1050–1061.

23. Sailer, U., Randall Flanagan, J., and Johansson, R.S. (2005). Eye–hand coordination during learning of a novel visuomotor task. J. Neurosci. *25*, 8833–8842.

Figure Titles and Legends

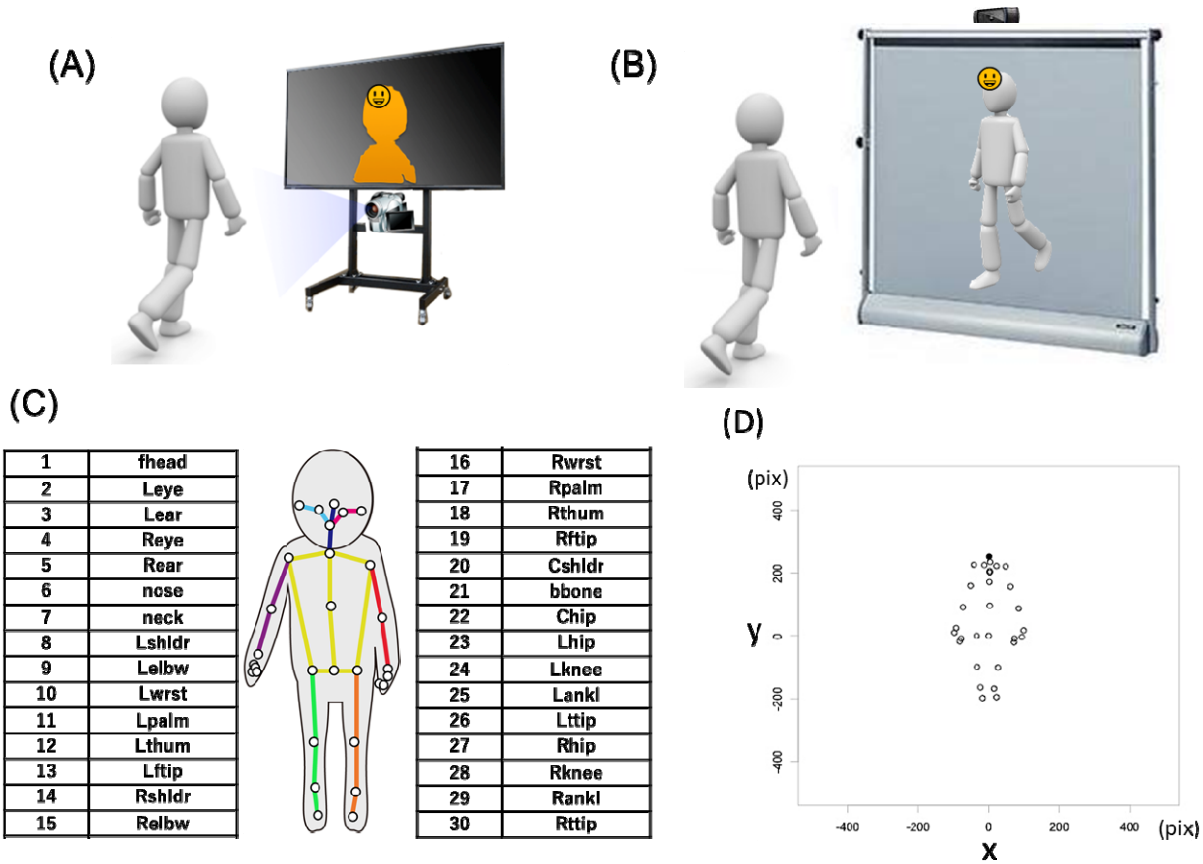

Figure 1. Real-time estimation of body parts for the XR mark test (“Bodytoypo”).
Note: (A) The concept of the XR mark test as a translational task. (B) Online presentation for participants touching the AR mark in the current study. (C) The input original image (frame of the video) with the estimated bone (30 parts) using a machine-learning algorithm. (D) The depicted output for x-y coordinates of each body part (30 fps) for the following offline motion analysis. The location of the center of the hip (Chip, 22) was always fixed at the origin (0,0) as the reference point.

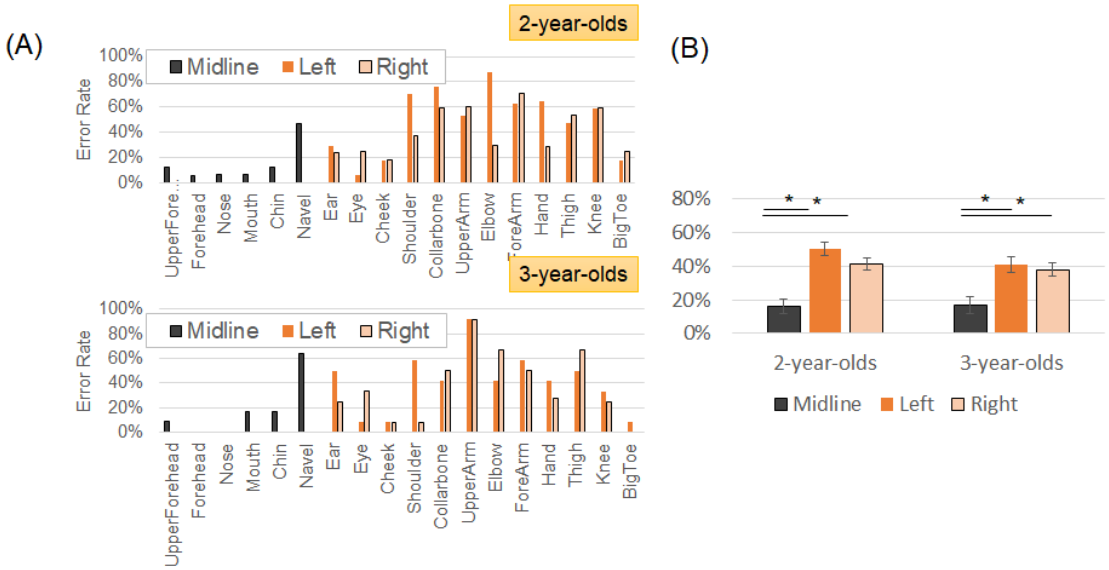

Figure 2. Error rates for each body part (based on Bodytoypo game).
Note: (A) Error rates for each body part by age. Face-related parts showed low error rates. The body part around the face had a low error rate regardless of age. (B) Error rates for each age by position category (midline, left, and right side). Error rates for body parts located on the midline were significantly lower than those for other body parts.

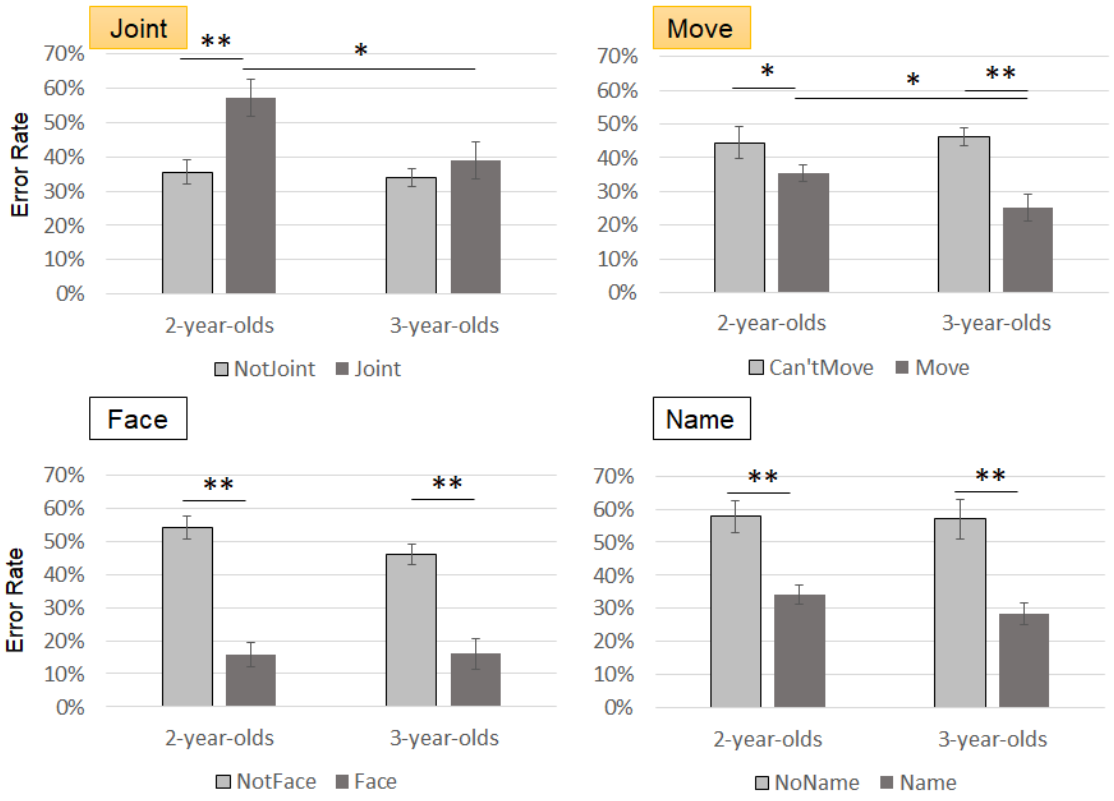

Figure 3. Error rates by property of body parts.
Note: The 30 body parts presented by Bodytoypo were categorized into two groups from the following four perspectives, and the differences in error rates were examined. We used the following categories: joint or not, movable or not, face or not, and named or not. In all four categories, the main effects of categories were observed, but only for the categories of joint and movable, an interaction with age was observed.

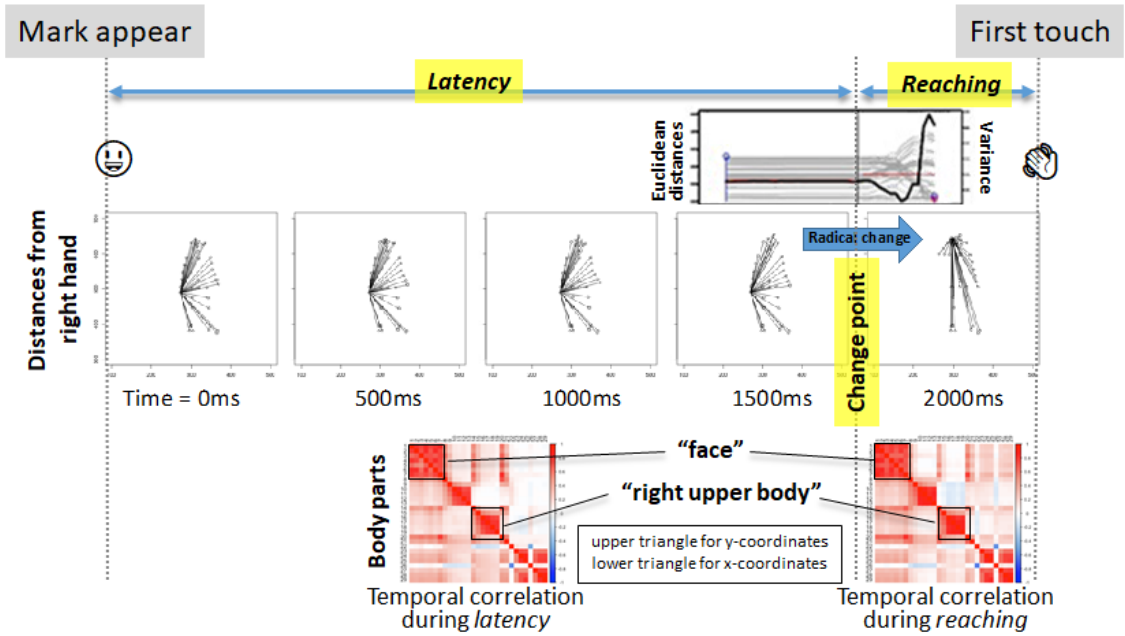

Figure 4. Schematic illustration for each trial.

Note: Each trial (from the mark appearing to the first touch) has two periods: response latency and reaching duration. This was defined by the detected change point over the time series of variance among distances between the used hand (in this example, right hand) and other body parts (the upper panel or supplementary Figure S4). During reaching, these distances dramatically change compared with those during latency (middle successive panels). The association among body parts (temporal correlation of x- or y-components) was calculated for each period (lower two panels), where some "clusters" are observed on the diagonal (e.g., "face" cluster). This correlation matrix was further vectorized and used in the statistical analysis as inputs for predicting participants' touch errors as outputs.

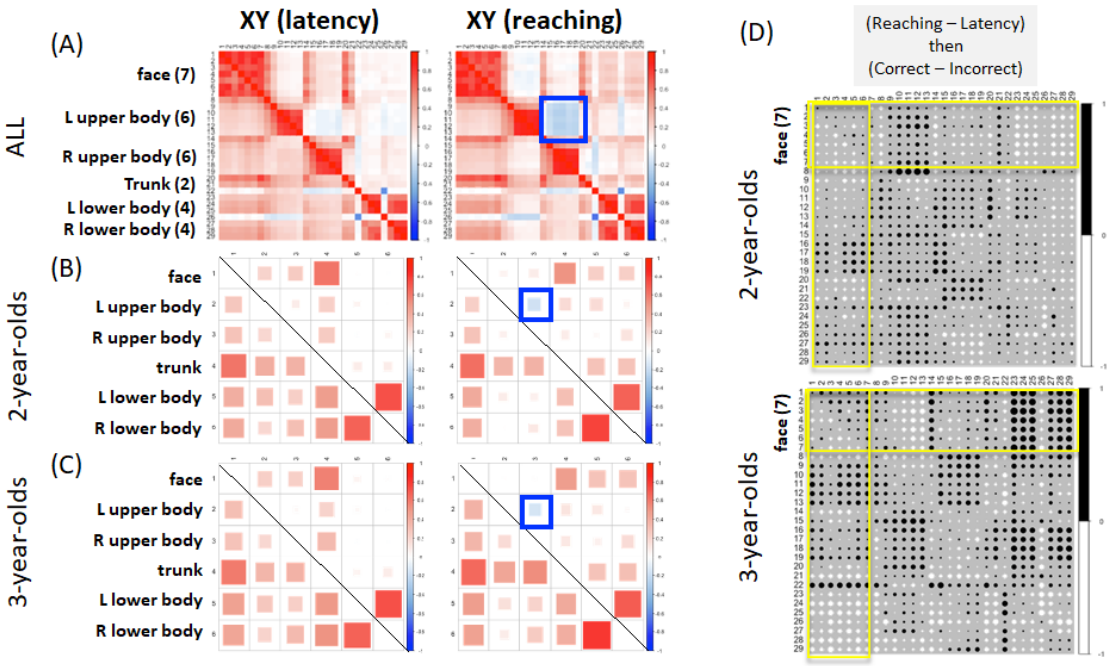

Figure 5. Temporal correlation among body parts for 2- or 3-year-olds.
Note: Temporal correlation matrices for body parts during latency or reaching. The upper triangles are for x-, whereas the lower triangles are for the y-coordinates. Body parts (A) or clusters (B, C) were positively correlated in general, except for L-R upper-body relations during the reaching period (blue squares). This negative correlation in the x-coordinate suggests bimanual coordination in reaching behaviors that could validate the change point estimation. Both aged groups behaved similarly in terms of the association among body clusters (B, C) as well as of the reaching trajectory (see Figure S5). (D) However, the face behaved oppositely between 2- and 3-year-olds if we see the subtracted matrix between correct and incorrect trials (see the main text for the detailed discussion).

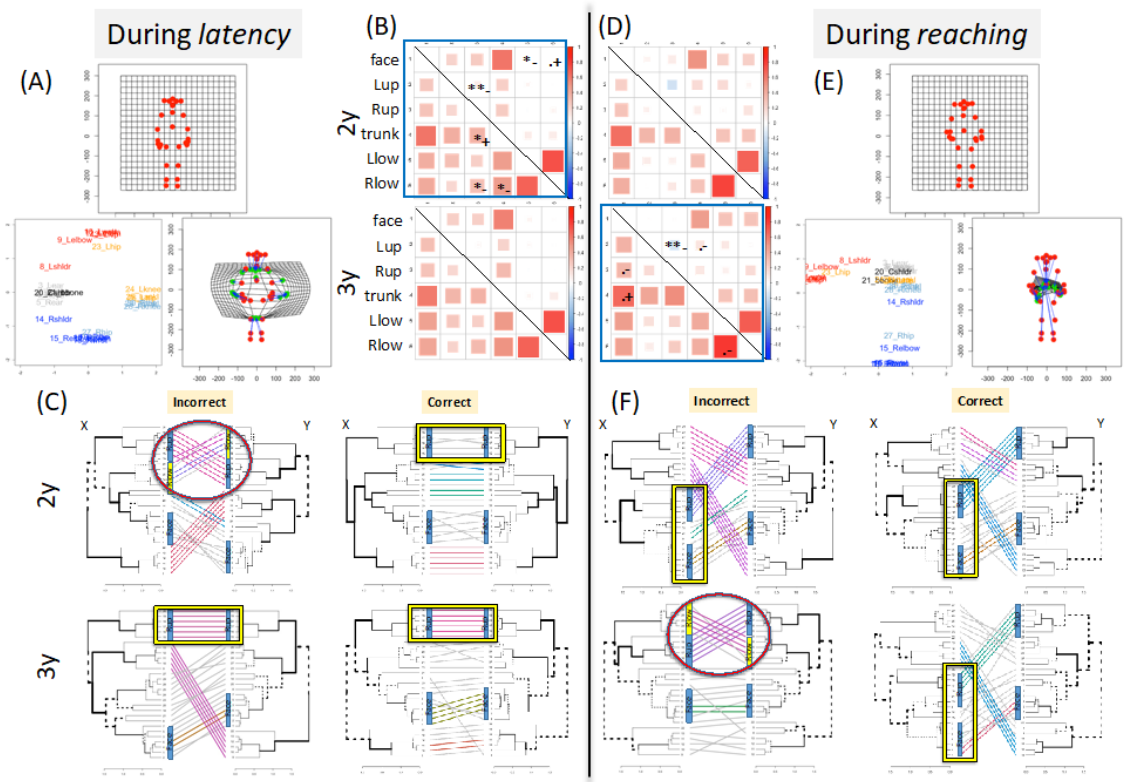

Figure 6. Body parts associations for configuration (AE), prediction (BD), and coordination (CF). Note: The temporal correlation matrix for latency (A) globally recovered the physical body-parts configuration through MDS projection and Procrustes transformation, while that for reaching (E) was not due to the various movements included. The latency matrix for 2-year-olds can statistically predict the later touch errors (B), while the reaching matrix for 3-year-olds can (D). During the latency period, the incorrect tanglegram for 2-year-olds was characterized as the right upper and lower parts association (D). During the reaching period, the same association was observed for the incorrect tanglegram for 3-year-olds (F). See the main text for further discussion.

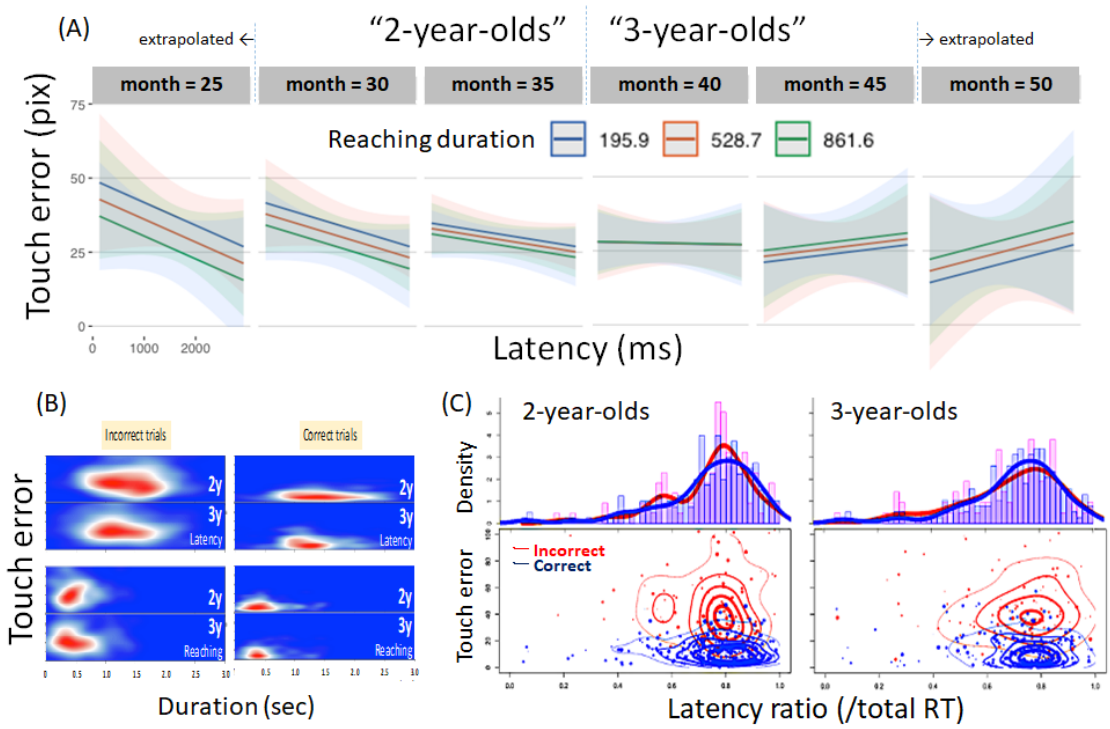

Figure 7. Estimated parameters for predicting the touch error regarding the age in months. Note: (A) A linear mixed model among potential inputs was explored for the minimum AIC. Specifically, the relationship among latency, reaching duration, and age (months) for predicting touch errors was illustrated, where the slopes were gradually reversed as age increased. (B) The distributions as heatmaps for the duration (latency/reaching) and touch error in terms of age and in/correct trials. (C) Contour plots of (B) for comparison (lower), and probability distributions for the latency ratio. In sum, the touching strategy seems to be opposite between 2- and 3-year-olds in terms of feedback/forward weights.