Tracking Action in the XR Mark Test Reveals Developing Body Representation Among Toddlers

Michiko Miyazaki[1]*, Tomohisa Asai[2], Norihiro Ban[3] and Ryoko Mugitani[4]

[1] Faculty 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
(4982 words; in the main text, excluding the Materials and methods, References, and Figure legends.)

Abstract

The mark test is a popular index for self-recognition or -consciousness, but has not been regarded as an index for assessing the development of self-body representations. We extend the mark test to a cross-reality (XR) task (referred to as *Bodytoypo*) that assesses whole-body part localization using augmented reality, to evaluate sensorimotor body representations among toddlers. Specifically, our system detects participants' skeletal data in real-time, and displays virtual marks on various parts of their bodies, which they could "touch." Thirty 2- and 3-year-olds participated in the task; their parents reported their bodily word acquisition. The accuracy of first touch reaching and reaction time were further analyzed for each body part. Specifically, in addition to conventional frame-by-frame coding, we performed the AI-statistics-based automated analysis using 2-D image processing. The results showed decreases in errors at joints and movable body parts between 2–3 years of age. However, there was no correlation with body part vocabulary richness. Localization error can be predicted by dynamical body part coordination. Regarding acquired internal models of reaching movement, 3-year-olds obtain fast and predictive (feedforward) reaching rather than visually guided (feedback) careful reaching. Analyzing reaching behaviors in the XR mark test revealed the development of sensorimotor body representations.

Keywords

mark test, reaching movement, body representation, augmented reality, Openpose, statistical analysis, internal model, feedback control, feedforward control, development

Introduction

The mark test, which was developed by G. Gallup in 1970, is a classical test for evaluating "self-recognition," and has been studied extensively among various species for the past 50 years (Amsterdam, 1972; Anderson & Gallup, 2015; Gallup, 1970; Lewis & Brooks-Gunn, 1979; Suddendorf & Butler, 2013). In the mark test, the participants discover, through their

mirror images, a mark secretly placed on their faces, which they spontaneously touch or attempt to remove. Such self-oriented behavior has been observed among human infants and a limited number of other species. Whether passing this test indicates the emergence of self-recognition has long been controversial (Anderson & Gallup, 2015; de Waal, 2019; Rochat & Zahavi, 2011; Suddendorf & Butler, 2013).

To date, the mark test has been too preoccupied with the emergence of self-oriented behavior (de Waal, 2019; Rochat & Zahavi, 2011). The mark test has often been a one-shot. Moreover, aspects of the body and the act, such as the various body parts targeted and the various ways of reaching for the mark, have rarely been considered. We believe, however, that we can clarify a part of the acquisition process of self-body representation by analyzing the reaching for marks on the entire body. The image of one's body in a mirror or video becomes the objectified self for those who have just acquired mirror self-recognition. The acquisition of the recognition and manipulation of the body reflected in mirrors and videos, for example, helps induce the emergence of a sense of body ownership and a sense of agency, as examined in the rubber hand illusion (RHI) and robot hand illusion. This study proposes that new aspects of the mark test can be found by reinterpreting it as a "paradigm" that naturally evokes self-oriented, visually-guided, and spontaneous reaching movements.

We attempt to reinterpret the mark test because we recognize that toddlers' reaching errors contain a wealth of information that can be used to capture developmental changes in body representations. For example, 2-year-olds exhibit interesting errors in the test from the perspective of body representation acquisition. Approximately 40% of 2-year-olds looked at a mirror image, found a mark secretly placed on their forehead, and tried to remove it from the "back of their heads" (the rear-search error (Imaizumi et al., 2021; Miyazaki & Hiraki, 2017)). This error may result from the immaturity and entanglement of cognitive abilities related to the acquisition of body representations, including (1) the acquisition of the meaning and relative associations between body parts (body representation), (2) multimodal integration with visual information, and (3) motor control of the arm used to reach (acquisition of internal models). By analyzing the cognitive background of reaching errors from these three characteristics, it may be possible to highlight the development of body representations from multiple perspectives. Below is a summary of body representation development that can be approached from the three characteristics.

First, the aspect of the mark test to evoke self-oriented behavior involves the acquisition of meanings and relative associations between body parts (semantic body representation (Schwoebel & Coslett, 2005)). For example, the upper forehead, where the rear-search errors occur, has no independent body name and may have a less clear relative positional relationship than the nose (the rear-search errors never happen with the nose). Several studies show that the speed of name acquisition and the difficulty of body part pointing vary by body part (Brownell et al., 2007, 2010; Camões-Costa et al., 2011; Waugh & Brownell, 2015; Witt et al., 1990). For example, Camões-Costa et al. conducted a task in which 17 2- to 3-year-olds were asked to label and point to 50 body parts. The results showed that the participants better comprehended body parts related to the face or represented a relatively large body area. To examine the relationship between body part name acquisition and reaching behavior, we analyze the relationship between a questionnaire on vocabulary acquisition status and reaching for marks. We hope that this will clarify the differences in localization difficulty and coordination by body part. Describing the differences in localization difficulty by body part may provide a clue to the elucidation of the so-called body map or sensory homunculus.

Second, regarding multimodal integration with visual information, we assume that we can examine the specificity of children's visual processing in terms of how they map their mirrored body to their real body, that is, how they integrate visual information (mirrored body) and interoceptive information (real body based on internal sense; e.g., proprioception, touch).(e.g., proprioception, touch). The rear-search errors may have been caused by a misalignment of the real body and the mark owing to the priority given to the visible mirrored body. The tendency for children to prioritize the processing of the visible body in multisensory processing was also reported in a study of RHI among 6- to 7-year-old children, suggesting that integrated processing of vision, touch, and proprioception may require a lot of time (Gottwald et al., 2021). However, RHI among younger children is considered difficult to comprehensively examine owing to the nature of the task, which requires verbal skills for subjective reports and fine motor control that indicates proprioceptive drift (Cowie et al., 2013, 2016). A multisensory processing task applicable to younger children is the task of reaching a visual target; Shen et al. (2010) analyzed spontaneous reaching to a toy that appeared at various distances (Shen et al., 2010) among 1- to 5-year-olds. The results indicate that smooth reaching requires head coordination. They visually perceive the target by stabilizing the head before reaching, resulting in smooth reaching. Visually guided reaching, which can be observed using the mark test, integrates the body as a mirror image as a manipulation target from the inside and outside. Moreover, clarifying the process of its mastery is important for understanding the learning mechanism of multisensory processing related to the acquisition of body representations.

Finally, the property of spontaneous reaching movements, tied to the property of being visually guided, can reveal developmental changes and changes in weightings in the internal model. The rear-search error may have been caused by an attempt to map the mirror image (vision) to the real body (proprioception) through feedforward control to further localize the mark when it should have been done carefully. Studies of the visuomotor control of the task of reaching for external objects suggest that acquisition proceeds from feedback control, followed by gradual visual-motor integration, and feedforward control acquisition with prediction (Bernstein, 1967; Scott Kelso, 2014). However, in the 1980s, even newborns were able to reach for nearby targets if their trunks were supported, and their posture was stable (Hofsten & von Hofsten, 1984), despite not seeing their hands (Claxton et al., 2003; Corbetta et al., 2014, 2018). Inexperienced neonates also reach in a feedforward control-like manner depending on the situation. We believe that by examining the developmental process of feedback and feedforward control weighting in reaching, in which each body part is a target, we can clarify the development of the internal model and the dynamics of behavioral control.

Therefore, the three properties of the mark test can be considered a paradigm that illuminates new aspects of the mark test as a body part localization task. The mark test becomes a new task that allows for the examination of how visual information and prediction are incorporated to achieve correct localization, and how each body part is coordinated to touch the target on body parts. The three points listed here are intricate and difficult to examine individually. However, all are essential windows for capturing developmental changes in body representations. We already developed a body part localization task several years ago, that applied the paradigm of the mark test to spontaneously and repeatedly elicit reaching for an invisible body part, the face (Miyazaki et al., 2019). Therefore, in this study, we extend the previously developed task to a whole-body version (see Materials and Methods, Figure 1 and Figure S1). Integrating skeletal detection with Openpose (Cao et al., 2021) and augmented reality (AR) technology, virtual

marks were displayed on 30 parts of the participants' bodies. Their localization errors were further analyzed comprehensively. For the analysis of the reaching movement, the reaching motion was divided into latency and duration according to the change point. Bodytoypo is an interactive "XR" mark test that crosses the boundary between real and virtual, where participants can "touch" the AR marks. We further discuss the developmental transition of body representations, multisensory integration, and internal models.

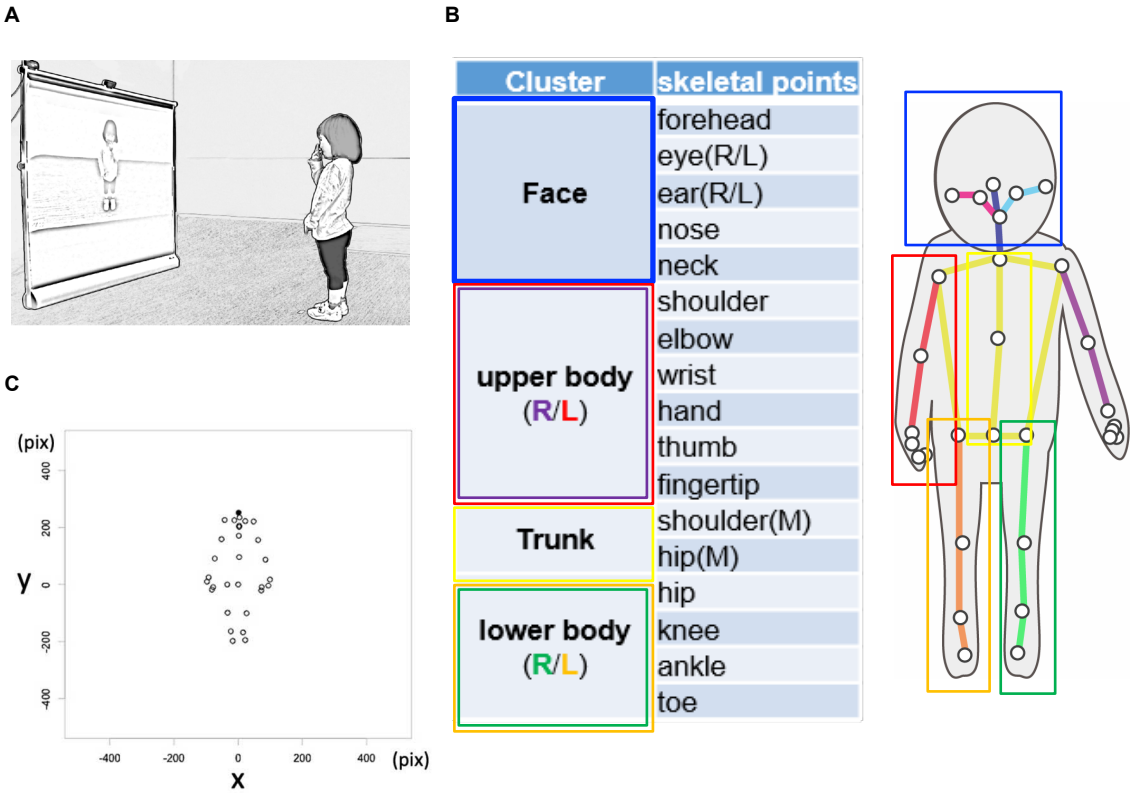

Figure 1. Real-time estimation of body parts for the XR mark test ("Bodytoypo").
Note: (A) Online presentation for participants touching the AR mark in this study. (B) The original input image (frame of the video) with the estimated bone (30 parts) and six body clusters using automated AI analysis. R/L/M in the table indicate right, left, and midline, respectively. (C) 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.

Results

Bodytoypo is a task that can maintain participants' motivation and has no age difference in the overall error rate

The average number of executed trials were 27.50 (SD = 5.371) and 29.75 (SD = .452) for the 2- and 3-year-olds respectively. Most participants completed 30 trials, regardless of their age (t (28) = -1.439, p = .161, *n.s.*, effect size d = -.522). This result is consistent with our previous study (Miyazaki et al., 2019) and suggests that we could maintain participants' motivation during this whole-body version. Furthermore, the positive effect of gamification (i.e., producing an enjoyable animation and sound) was demonstrated (see Materials and Methods).

The overall error rate for the first touch was 39.9% (SD = .132) and 35.0% (SD = .095) for the 2- and 3-year-olds respectively. The age groups differed insignificantly (t (28) = 1.098, p = .282, *n.s.*, effect size d = .398). 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 integrating 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 and the interaction between age and mark location were insignificant (F(1,28) = .830, p = .370, MSe = .044, *n.s., partial* η^2 = .029, F(2,56) = .954, p = .389, MSe = .019, *n.s., partial* η^2 = .033, respectively). However, the main effect of mark location was significant (F (1,28) = 35.647, p < .01, MSe = .019, *partial* η^2 = .560). Holm's multiple comparisons showed that the error rates in the midline location were significantly lower than those in other body parts, and that the left-fight difference was insignificant (midline-left: p < .01, effect size d = -1.635; midline-right: p < .01, effect size d = -1.338; left-right: *n.s.,* effect size d = .385).

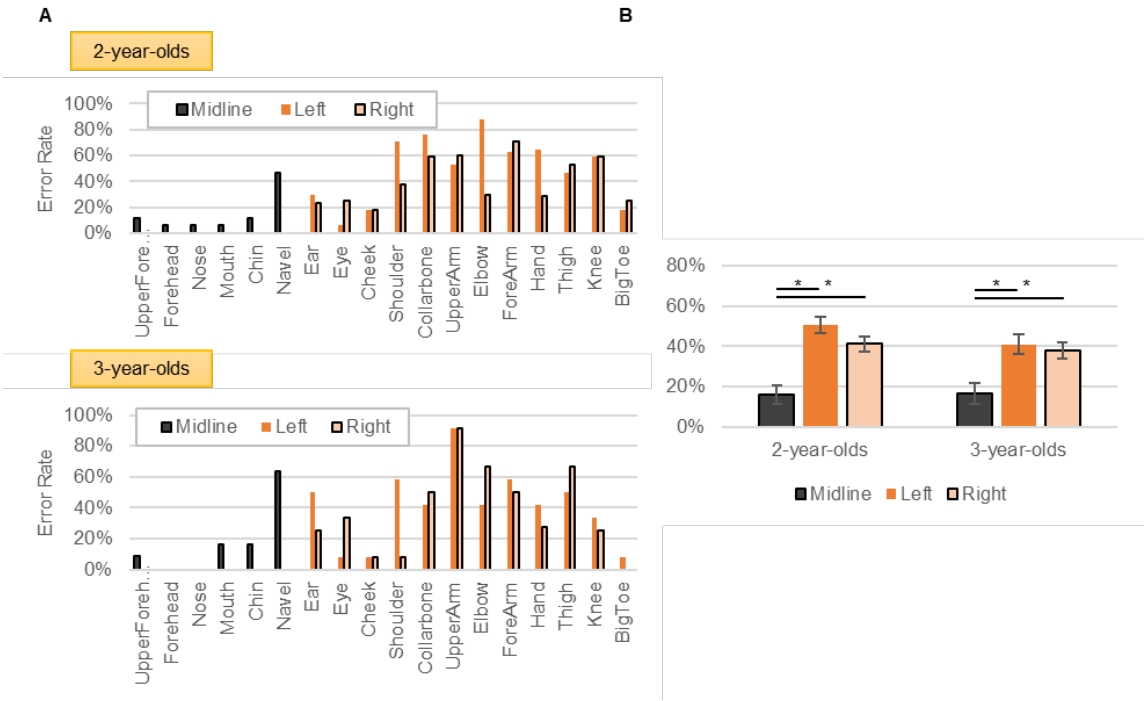

Figure 2. Error rates for all 30 body parts in Bodytoypo.

Note: (A) Error rates for all 30 body parts by age. 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). Error rates for body parts located on the midline were significantly lower than those for other body parts.

Localization errors were fewer for faces, named, moveable parts, and joints and were uncorrelated with word acquisition of body parts

We further analyze the error rates of each body part combined with bodily function. 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 differed. We examined four bodily functions, that is, whether the body part was a joint, moveable part, face, or named body part. The main effects of all bodily functions were confirmed (main effects of joint: $F(1,28) = 12.259$, $p = .002$, , $MSe = .021$, *partial* $\eta^2 = .305$; moveable: $F(1,28) = 25.507$, $p < .01$, $MSe = .013$, *partial* $\eta^2 = .149$; face: $F(1,28) = 104.939$, $p < .01$, $MSe = .016$, *partial* $\eta^2 = .477$, and named body part: $F(1,28) = 58.481$, $p < .01$, $MSe = .017$, *partial* $\eta^2 = .676$. Additionally, the interaction of age and bodily function was significant regardless of whether it was a joint or a moveable part (age * joint: $F(1,28) = 4.898$, $p = .035$, $MSe = .021$, *partial* $\eta^2 = .149$; age * moveable: $F(1,28) = 4.026$, $p = .055$, $MSe = .013$, *partial* $\eta^2 = .126$. These results suggest that the accuracy of the first touch increases between the ages of 2 and 3 years for moveable joints and parts.

The results of the body word acquisition questionnaire for parents revealed that for the 60 words asked, 28.7 (SD = 10.599) (2- year-olds) and 40.8 (SD = 14.640) (3-year-olds) of the words overall could be produced and comprehended (see Figure S2C). There was a significant difference in the mean number of acquired words between the age groups (t (28) = –2.632, $p = .014$, effect size $d = -.955$). A partial correlation analysis was conducted on the number of acquired body vocabulary and Bodytoypo's error rate, with age in months as a control variable (Figure S2A). A correlation analysis was also conducted between the acquisition rate for body parts used in Bodytoypo and the error rate (Figure S2B). However, the results were insignificant.

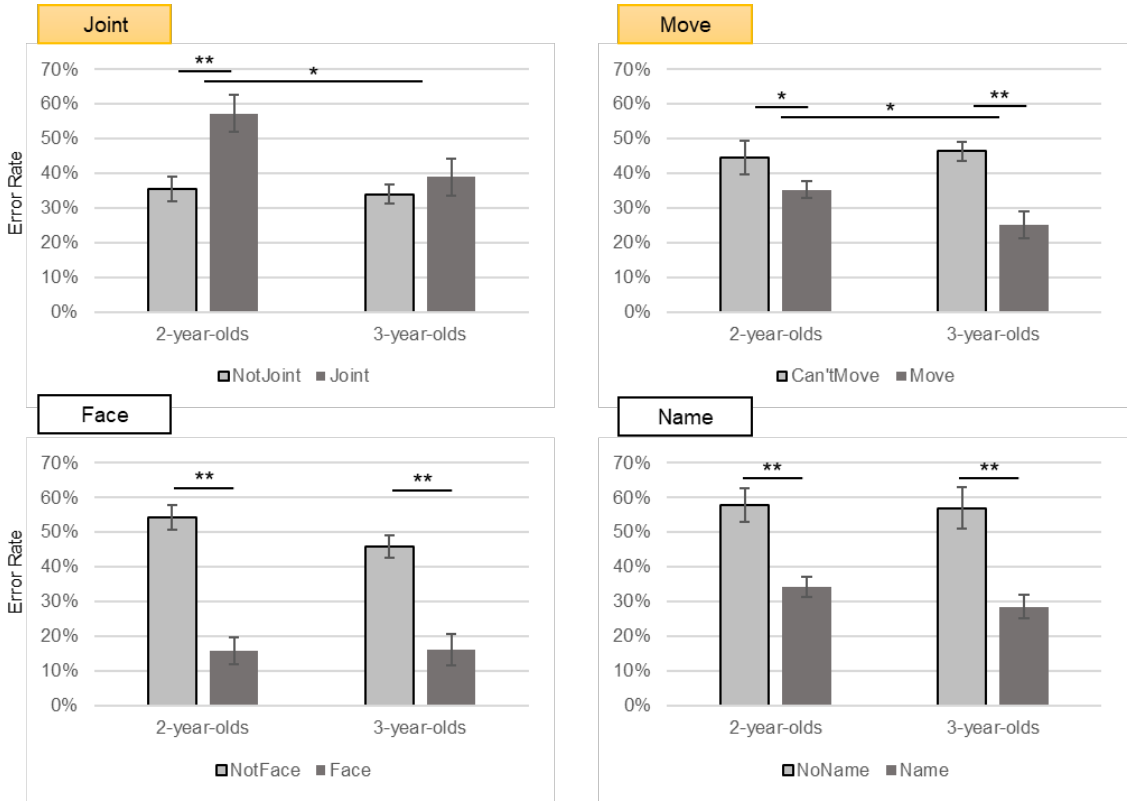

Figure 3. Error rates by bodily function.

Note: The 30 body parts presented by Bodytoypo were categorized into two groups from the following four perspectives. The differences in error rates were further 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. However, an interaction with age was observed only for the joint and movable categories.

Visual guided reaching develops as part of multisensory integration

Body-parts association during latency and reaching:

Figure 4 summarizes the procedure of each trial, where Figure 5 suggests the averaged temporal correlation among 30 body parts (29 parts except for the reference). The upper triangles represent x-, while the lower triangles represent 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 generally positively correlated, except for L-R upper body associations during the reaching period (blue squares). This negative correlation in the x-coordinate suggests bi-manual coordination in reaching behaviors, which 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 in the case of subtracted matrix between correct and incorrect trials (yellow parts in Figure 5D for reverse monochrome patterns) or the cross-correlation between the face and the hand (Figure S7A). We observe what happened to those participants as follows:

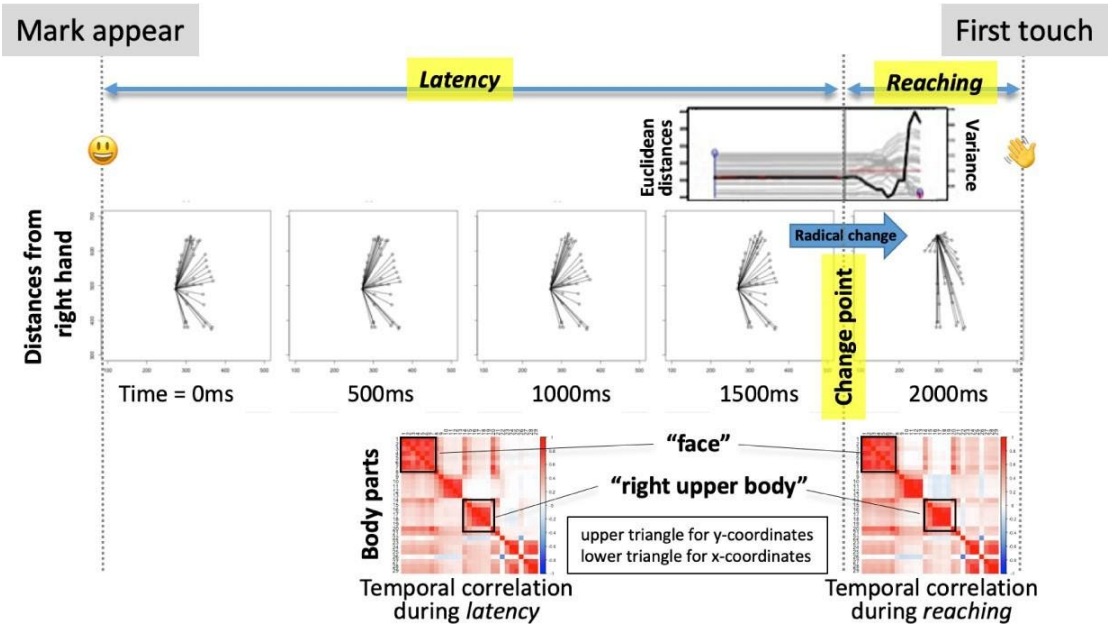

Figure 4. Schematic illustration for each trial.

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 markedly change compared with those during latency (middle successive panels). The association between 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.

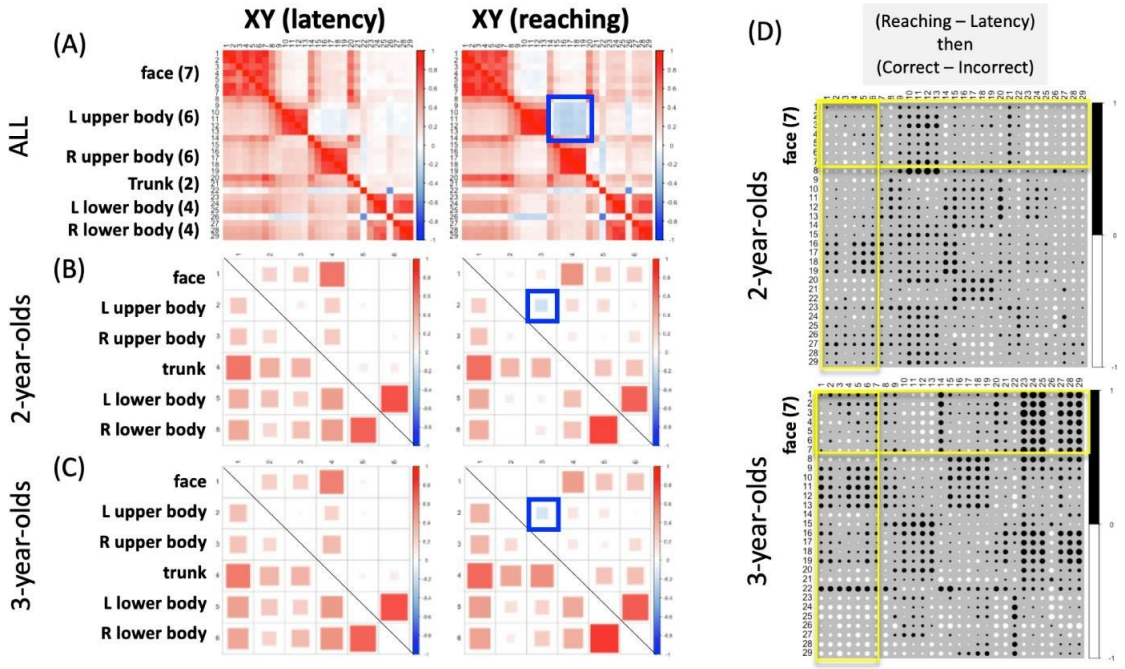

Figure 5. Temporal correlation among body parts for 2- or 3-year-olds.

Note: Temporal correlation matrices for body parts during latency or reaching. The upper triangles represent x-, whereas the lower triangles represent the y-coordinates. Body parts (A) or clusters (B, C) were generally positively correlated, 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 between 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 in the case of subtracted matrix between correct and incorrect trials (see the main text for a comprehensive discussion).

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 association between body parts during latency is relatively stable so that we can see the original body configuration. However, it is horizontally biased as 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). Although the correlation matrix was also markedly similar to that 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 re-validate the change point estimation.

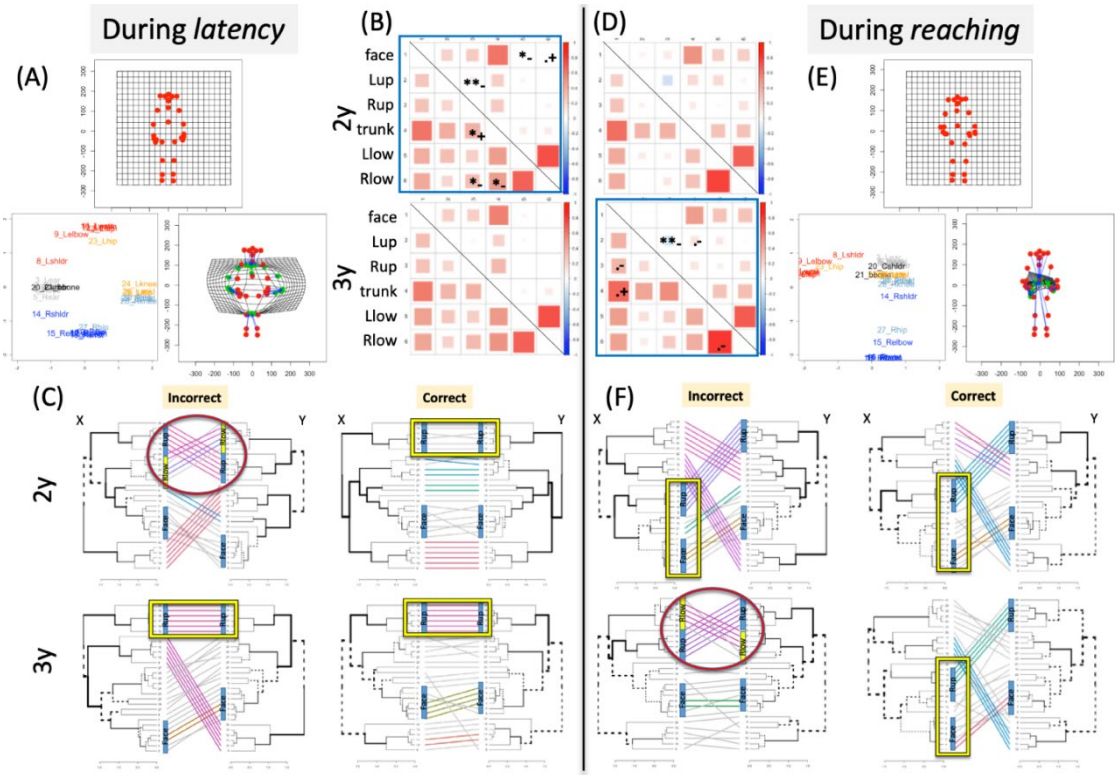

Figure 6. Associations between body parts for configuration (AE), prediction (BD), and coordination (CF).

Note: the temporal correlation matrix for latency (A) globally recovered the physical body-parts configuration through MDS projection and the Procrustes transformation while that for reaching (E) did not, owing 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 predict the later touch errors (D). During the latency period, the incorrect tanglegram for 2-year-olds was represented as the right upper and lower parts association (C). During the reaching period, the same association was observed for incorrect tanglegram for 3-year-olds (F). See the main text for a comprehensive discussion.

Although the correlation matrix between 2- and 3-year-olds was congruent (Figure 6BD, reprinted from 5BC), there was a statistical difference in predicting their 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 ($F(30, 436) = 1.666$, $p = .016$), while for 3-year-olds, the *reaching* matrix can predict touch error ($F(30, 308) = 1.636$, $p = .022$). However, the latency matrix for 3-year-olds or the reaching matrix for 2-year-olds cannot predict their touch errors ($F(30, 288) = 1.475$, $p = .057$; $F(30, 435) = 1.19$, $p = .229$, respectively). For each model, the significant (and marginal for reference) variables are indicated by symbols (Figure 6BD, $p < .01$**, $p < .05$*, $p < .10$.) with signs (positive or negative effects), which include the L-R upper body associations (i.e., hand movements) in the x coordinates regardless of age. Figure 6CF further depicts the association between body parts in tanglegram (R library “dendeextend” (https://talgalili.github.io/dendextend/)) in terms of “correct” or “incorrect” labels using the manual codings. Regarding latency (Figure 6C), the incorrect matrix for 2-year-olds is characterized where upper (R)-lower (R) parts are associated (red ellipse), while the other tanglegrams suggest that the upper (R) cluster was independently moving (yellow rectangle). Regarding reaching (Figure 6F), the incorrect matrix for 3-year-olds is characterized where the upper (R)-lower (R) parts are associated again (red ellipse), while the other tanglegrams suggest that the upper (R) and Face clusters were associated in the x coordinates (yellow rectangle) (see Figure S7A for the cross-correlation between the face and hand). If face movement is critical for precise touching (see Figure 5D again), this could be related to their feedback/forward control in the AR mark test (Shen et al., 2010). This might have emerged as the weight between the latency/reaching duration as a function of feedforward/back control strategy between both age groups as follows.

Feedforward/feedback control weights changes among ages

Weight between feedback/forward strategy changes among ages:

Thus far, the Bodytoypo task mainly generated participants’ touch errors and the latency/reaching duration (Figure 7B) as well as the association between body-parts. 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 effective in qualitatively depicting what was happening in the current task during the latency or reaching periods, the all-in-one statistics with a linear mixed model attempted to conclude these relationships continuously, using some other covariates as mentioned above. Therefore, 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; the touch error value was the output. This every-effect-model generated significant interaction effects (e.g., the interactions among reaching, age, repetition, and initial distance [$t(422.5) = 2.721$, $p = .0068$]). The step function further identified the model with the minimum Akaike information criterion (AIC). This selected-effect-model still generated a significant interaction among latency/reaching, age,

repetition, and initial distance (t(557. 0) = 2.30, p = .0218 for latency; t (554. 0) = 2.38, p = .0177 for reaching). Therefore, Figure 7A presents the relationship between 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).

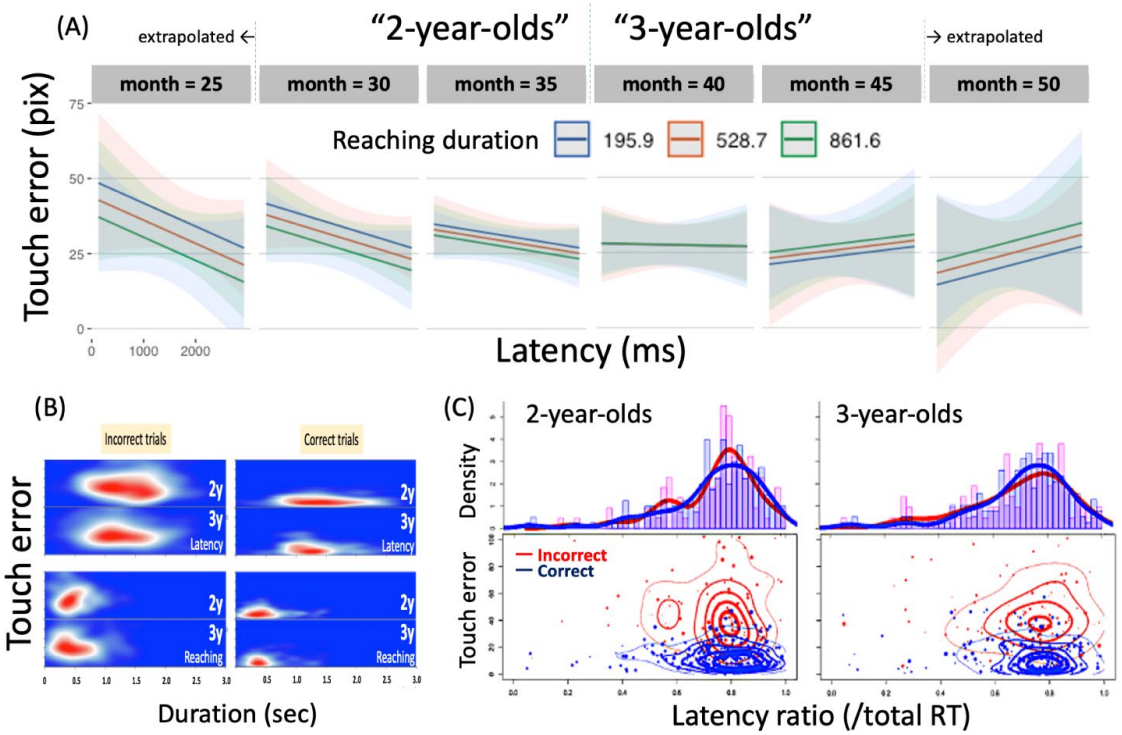

Figure 7. Estimated parameters for predicting the touch error regarding the age in months.

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

For younger months, both latency and reaching should be longer for correct touch (minimized touch error). However, generally, there was a robust anti-correlation between latency and reaching duration (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 shown below, the latency ratio did not vary for incorrect trials for 2-year-olds). Several months later, they no longer exhibited ballistic movements. The older participants 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 (Bushnell, 1985).

Given that visually guided reaching through a mirror (e.g., mirror drawing (Starch, 1910)) 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 owing to habituation-driven exploration or playing with feedback control (Figure S7B, the upper panels).

Figure 7B shows the weight between latency/reaching duration 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 evaluating their feedback weight. However, for the correct trials among 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 evaluated by 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 current task could reveal the complex but optimal weight between the internal forward output, as well as the external feedback input. We further re-consider visual observations over the original videos, according to the statistical results.

Reaching strategies are mixable and switchable. Therefore, it is appropriate to evaluate weights rather than classifications

Automatic analysis revealed developmental changes in toddlers’ reaching strategies: ballistic search, feedback control (FB), and feedforward control (FF). Ballistic search refers to the participant reaching for the mark quickly, without aiming (Supplementary Video S1). Feedback control is a slower reaching strategy in which hand movements are adjusted while watching the video monitor or actual body part (Supplementary Video S2). Lastly, feedforward control is a strategy for reaching the target quickly while anticipating localization (Supplementary Video S3).

To examine the validity of this categorization, we test whether the three reaching strategies found in the automatic analysis were classifiable to human observers: two coders rewatched the video and classified all trials into one of the three strategies. An agreement was calculated using 32.5% of the total trials. However, the agreement was low (κ = .33), especially for the ballistic exploration and FF, and FF and FB classifications, which showed a difference of approximately 25% in judgment. This result suggests that the classification of reaching strategies is difficult for the human eye. Instead, we thought it reasonable to explain the switching of reaching strategies during the same trial (Supplementary Video S4) and the remaining ambiguity of reaching methods as an online weighting of the internal model rather than as an exclusive choice of three strategies. However, the qualitative re-evaluation revealed that differences in the shape of the hand used for reaching, especially the use of the index finger (see Discussion), are valid indicators of developmental changes in reaching strategies (Figures S3).

Discussion

This study aims to demonstrate another potential of the classical mark test as a window to reveal toddlers' sensorimotor body representation. Focusing on the three properties of the mark test (self-oriented, visually guided, and spontaneous reaching), and through the analysis of the first touch of the Bodytoypo, we describe body representation development from multiple perspectives, multimodal sensory integration, and the acquisition of internal models. The main findings are as follows. In terms of semantic body representation, touch errors and vocabulary acquisition scores for each body part were not correlated. However, from the perspective of multimodal integration, touching body parts that are easily used in daily life, such as named parts, facial parts, joint parts, and movable parts, was relatively easy regardless of age. Furthermore, from the perspective of internal model acquisition, three types of reaching strategies could be assumed: ballistic search, feedback control, and feedforward control. Moreover, the strategies and weightings varied with age. These reaching strategies indicate a process of optimal integration of internal outputs and external inputs. This is consistent with our previous observation regarding the rear search error in the classical mark test, which disappears among 2- to 3-year-olds (Imaizumi et al., 2021).

The evaluation of body representations from the three perspectives proposed in this study may have created a new avenue for approaching the general question of self-recognition. Particularly, the research on the perspective of enhancing one's understanding of self-mirrored image based on own actions is scant. The self-image/mirror image is not merely an image of oneself but rather a reflection of one's own body and an object to be controlled. This may allow us to approach the elucidation of the early developmental mechanism of a sense of body ownership and subjectivity without using the RHI task, which is difficult to apply to younger children. Some possible future extensions include, Bodytoypo, which inserts a time delay or uses a self-image with arms unnaturally extended. Bodytoypo is also a task in which participants can perform intuitively without verbal instructions. Therefore, there is hope that it could be applied not only to young children and adults but also to other species with appropriate updates.

We also describe the significance of two new technical approaches that have been attempted in extending this task. The first is the integration of new technologies with behavioral experiments. For real-time skeletal detection of the whole body, we applied an image processing technology referred to as Openpose, which detects the skeleton from 2D images and performs posture estimation using deep learning, allowing even small infants to detect the skeleton in the same way as adults (Ossmy & Adolph, 2020). The advantage of being able to detect motion trajectories from image processing without the use of markers is significant for research with young children. For example, in another study, a relatively large number of participants were excluded (attrition rate: 31.8%) because they removed the 3D markers themselves (Kahrs et al., 2014); the attrition rate in this study was further reduced by half. AR markers were then displayed on each body part in near real-time. These results are obtained from an advanced program that uses a GPU machine to process skeletal detection, and mark display at ultra-high speed in an online task. However, the correct and incorrect judgments of the participant's touch were made by manual control. This is because automatic decisions based on 2-D image processing do not necessarily realize the touch subjectivity (for example, even if the participant did not touch the mark and their arm passed over the mark, the 2-D image processing would show that they had touched it). By appropriately introducing manual judgments by the experimenter, we succeeded in learning a rule that rewarded touching a mark on the screen.

The second is to automate the analysis by incorporating machine learning. This study uses AI statistics for offline analysis. In previous developmental studies, most of the analysis was based on human's classification of actions in video images of children taken by the human eye. However, a series of actions may involve multiple strategies; the first touches in this study were no exception. Significantly, this study was able to show changes in the weighting of the reaching strategies rather than in the form of classification.

Therefore, this project is not simply a digitization of the mark test but a rich collaboration of engineering (online task processing), machine learning statistics (offline analysis of whole-body movements), and developmental psychology (theoretical consideration of crucial phenomena and conducting experiments).

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 FB/FF control should be an integrated process between proprioceptive and visual information (Sailer et al., 2005). Particularly, 3-year-olds exhibited an optimal integration for correct touch where the FF might have more weight than FB as 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 that 3-year-olds can adapt quickly as latency and reaching were rapid while utilizing motor prediction (FF weights for correct touch).

This process of taking in visual information has also been observed in studies of RHI and self-reaching. Children seem to prioritize visual information more than adults and therefore, respond more to the visible hand than the felt hand (Cowie et al., 2018). Additionally, studies that have placed vibrators on various body parts and analyzed the relevant responses show that reaching while looking at the target is not observed before six months (Somogyi et al., 2018). It is possible that a certain amount of time and motor experience is required for us to be able to balance and smoothly incorporate visual information into our proprioceptive processing of self.

Altogether, in addition to integrating vision and proprioception, young children gradually acquire the ability to control their own bodies 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 present study shows, it can approach the acquisition and adjustment of young children's sensorimotor internal model.

Materials and Methods

Participants

Thirty-six children aged 2.5–3.5 years old participated. The sample size was determined based on previous studies on body representations (Camões-Costa et al., 2011; Somogyi et al., 2018). All participants were Japanese. This study was conducted under the recommendations of the Otsuma Women's University's Life Sciences Research Ethics Committee, and written informed consent was obtained from all the participants' parents.

The study protocol was also approved by the Otsuma Women's University Life Sciences Research Ethics Committee (2019-012). The final sample comprised 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).

The final dataset used for AI-statistics analysis comprised 607/803 trials of 30 children (2.5-years-old, n= 310; 3.5-years-old, n= 297). Of these, 196 trials were excluded from the AI analysis (24.4% attrition) because the RTs were too long (>4 seconds) and/or failed to detect the skeleton through offline image processing.

Apparatus

We developed a body part localization task using augmented reality (AR) with the image processing library Openpose. 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 by children) was displayed as a mark on the coordinates of the target body parts. The net delay (transmitting and image processing) was approximately ten 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 camera angle was adjusted so that the entire 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 manual coding and movement analysis.

Tasks

Body part localization task (Bodytoypo):

The experimenter manually displayed the AR mark on the participant's life-size self-image, and observed whether the participant could correctly touch (or point to) the actual body part corresponding to the mark's location. The AR marks were displayed individually on 30 body parts (upper forehead, forehead, eye (Right/Left, hereafter 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 lightly differed from those in the independent offline analysis (see Figure 1B, Figure S1B,C). 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. If the participants could touch the correct body part with the AR mark, the mark disappeared with a fun audio-visual reward presented by the experimenter's manual keyboard control.

Questionnaire of body part vocabulary:

After completing the task, parents were asked about their children's vocabulary acquisition status through a questionnaire. They were asked about their speech and comprehension of 60 body parts and body-related vocabulary (e.g., tail, wings, horns) in Japanese.

Procedure

The experimenter first explained the study to the parents to obtain informed consent. The

experimenter and participating children built rapport through free playing. The experimenter further demonstrated 10-trials of Bodytoypo to the participants. During the demonstration, she intentionally touched the mark incorrectly and indicated that if they made a mistake, the reward would not appear. When participants were suitably positioned by standing straight and still in front of the screen, an AR mark would appear with a beep sound. The experimenter encouraged them to participate in the task by rhythmically presenting the AR marks. If the participants did not readily respond to the presented mark, the experimenter re-presented it or skipped the trial. If participants lost their motivation, the nose trial for recapturing motivation was inserted. Aside from those additional nose trials, a maximum of 30 trials (30 body parts) was executed for each participant. The experiment lasted 5 minutes.

Quantification and statistical analysis

The analysis in this study focused on the period between the appearance of the mark and the children's first touch in each trial. A detailed analysis of the accuracy and error of the first touch and the motor trajectory for each body part revealed developmental changes in the acquisition of body representations from the three perspectives presented earlier.

We employed both traditional frame-by-frame manual coding analysis and statistics-based automated analysis. The former analysis examined the number of trials performed, overall and body part-specific error rates, and the relationship between Bodytoypo performance and body part-related lexical acquisition. This analysis would allow us to examine whether the Bodytoypo task can sustain participants' motivation. We can then show (1) the development of body representation. The latter analysis focused on (2) developmental changes in multimodal processing with visual information and (3) developmental changes in internal models and their weighting through a detailed analysis of movement trajectories using two-dimensional image processing.

The overall pipeline in the latter analysis is shown below. The necessary process was first, the change of point detection as each trial (from the mark appearing to the first touch) comprised two periods: response latency and reaching movement (Figure 4). Therefore, the change point was estimated 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 drastically compared with those during latency. This automated estimation (separation of latency and reaching) was further validated during a process conducted later (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), indicating 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) rather than during reaching (Figure 6E). Furthermore, this correlation matrix was 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). Lastly, all the variables were put into a linear mixed-model (all-in-one statistical model) as continuous variables, to predict touch error (Figure 7A). In summary, 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. Furthermore, the effectiveness of the automated analysis is evaluated by applying the findings from the statistics-based automated analysis to the manual analysis, from a qualitative perspective.

Acknowledgments

We thank Ayano Ryoke and Moe Kobayashi for their help with the data collection and analysis. We also 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.

Competing interests

The authors declare no competing interests.

Reference

Amsterdam, B. 1972. “Mirror self-image reactions before age two.” *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology* 5(4): 297–305. https://onlinelibrary.wiley.com/doi/abs/10.1002/dev.420050403

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; Journal of Primatology* 56(4): 317–326. https://doi.org/10.1007/s10329-015-0488-9

Bernstein, N. 1967. The Co-ordination and Regulation of Movements. Pergamo. *Press, London*.

Brownell, C. A., Nichols, S. R., Svetlova, M., Zerwas, S., and Ramani, G. 2010. “The head bone’s connected to the neck bone: when do toddlers represent their own body topography?” *Child Development* 81(3): 797–810. https://doi.org/10.1111/j.1467-8624.2010.01434.x

Brownell, C. A., Zerwas, S., and Ramani, G. B. 2007. ““So big”: the development of body self-awareness in toddlers.” *Child Development* 78(5): 1426–1440. https://doi.org/10.1111/j.1467-8624.2007.01075.x

Bushnell, E. W. (1985). The decline of visually guided reaching during infancy. *Infant Behavior & Development*, *8*(2), 139–155. https://doi.org/10.1016/S0163-6383(85)80002-3

Camões-Costa, V., Erjavec, M., & Horne, P. J. (2011). Comprehension and production of body part labels in 2- to 3-year-old children. *The British Journal of Developmental Psychology*, *29*(Pt 3), 552–571. https://doi.org/10.1348/026151010X523040

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 Transactions on Pattern Analysis and Machine Intelligence* 43(1): 172–186. https://doi.org/10.1109/TPAMI.2019.2929257

Claxton, L. J., Keen, R., and McCarty, M. E. 2003. “Evidence of motor planning in infant reaching behavior.” *Psychological Science* 14(4): 354–356. https://doi.org/10.1111/1467-9280.24421

Corbetta, D., Thurman, S. L., Wiener, R. F., Guan, Y., and Williams, J. L. 2014. “Mapping the feel of the arm with the sight of the object: on the embodied origins of infant reaching.” *Frontiers in Psychology* 5, 576. https://doi.org/10.3389/fpsyg.2014.00576

Corbetta, D., Wiener, R. F., Thurman, S. L., and McMahon, E. 2018. “The Embodied Origins of Infant Reaching: Implications for the Emergence of Eye-Hand Coordination.” *Kinesiology Review* 7(1): 10–17. https://doi.org/10.1123/kr.2017-0052

Cowie, D., Makin, T. R., and Bremner, A. J. 2013. Children’s responses to the rubber-hand illusion reveal dissociable pathways in body representation. *Psychological Science* 24(5): 762–769. https://doi.org/10.1177/0956797612462902

Cowie, D., McKenna, A., Bremner, A. J., and Aspell, J. E. 2018. “The development of bodily self-consciousness: changing responses to the Full Body Illusion in childhood.” In *Developmental Science* (Vol. 21, Issue 3, p. e12557). https://doi.org/10.1111/desc.12557

Cowie, D., Sterling, S., and Bremner, A. J. 2016. “The development of multisensory body representation and awareness continues to 10 years of age: Evidence from the rubber hand illusion.” *Journal of Experimental Child Psychology*, 142: 230–238. https://doi.org/10.1016/j.jecp.2015.10.003

de Waal, F. B. M. 2019. “Fish, mirrors, and a gradualist perspective on self-awareness [Review of *Fish, mirrors, and a gradualist perspective on self-awareness*].” *PLoS Biology* 17(2): e3000112. https://doi.org/10.1371/journal.pbio.3000112

Gallup, G. G., Jr. 1970. “Chimpanzees: self-recognition.” *Science* 167(3914): 86–87. https://doi.org/10.1126/science.167.3914.86

Gottwald, J. M., Bird, L.-A., Keenaghan, S., Diamond, C., Zampieri, E., Tosodduk, H., Bremner, A. J., and Cowie, D. (2021). “The developing bodily self: How posture constrains body representation in childhood.” *Child Development 92*(1): 351–366. https://doi.org/10.1111/cdev.13425

Hofsten, C. von, and von Hofsten, C. (1984). “Developmental changes in the organization of

prereaching movements." *In Developmental Psychology* 20, 3: 378–388. https://doi.org/10.1037/0012-1649.20.3.378

Imaizumi, S., Asai, T., and Miyazaki, M. 2021. "Cross-referenced body and action for the unified self: empirical, developmental, and clinical perspectives." *Body Schema and Body Image: New Directions*, 194. https://books.google.co.jp/books?hl=ja&lr=lang_ja|lang_en&id=ObEzEAAAQBAJ&oi=fnd&pg=PA194&dq=Imaizumi+Asai+Miyazaki&ots=c3GL0r1h_l&sig=CyixjqQUBFrpy22dSu7nCh5nFpw

Kahrs, B. A., Jung, W. P., and Lockman, J. J. 2014. "When does tool use become distinctively human? Hammering in young children." *Child Development* 85(3): 1050–1061. https://doi.org/10.1111/cdev.12179

Lewis, M., and Brooks-Gunn, J. 1979. *Social Cognition and the Acquisition of Self*. https://doi.org/10.1007/978-1-4684-3566-5

Miyazaki, M., Asai, T., and Mugitani, R. (2019). "Touching! An Augmented Reality System for Unveiling Face Topography in Very Young Children." In *Frontiers in Human Neuroscience* 13. https://doi.org/10.3389/fnhum.2019.00189

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?* The sixth annual Budapest CEU Conference on Cognitive Development, Budapest, Hungary. https://www.bcccd.org/down/bcccd_final_program_2017.pdf

Ossmy, O., and Adolph, K. E. 2020. "Real-Time Assembly of Coordination Patterns in Human Infants." *Current Biology: CB* 30(23): 4553–4562.e4. https://doi.org/10.1016/j.cub.2020.08.073

Rochat, P., and Zahavi, D. 2011. "The uncanny mirror: a re-framing of mirror self-experience." *Consciousness and Cognition* 20(2): 204–213. https://doi.org/10.1016/j.concog.2010.06.007

Sailer, U., Randall Flanagan, J., and Johansson, R. S. 2005. "Eye–Hand Coordination during Learning of a Novel Visuomotor Task." *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 25(39): 8833–8842. https://doi.org/10.1523/JNEUROSCI.2658-05.2005

Schwoebel, J., and Coslett, H. B. 2005. Evidence for multiple, distinct representations of the human body. *Journal of Cognitive Neuroscience* 17(4): 543–553. https://doi.org/10.1162/0898929053467587

Scott Kelso, J. A. 2014. *Human Motor Behavior: An Introduction* (J. A. S. Kelso (ed.)). Psychology Press. https://doi.org/10.4324/9781315802794

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. *2010 IEEE 9th International Conference on Development and Learning*, 108–113. https://doi.org/10.1109/DEVLRN.2010.5578856

Somogyi, E., Jacquey, L., Heed, T., Hoffmann, M., Lockman, J. J., Granjon, L., Fagard, J., and O'Regan, J. K. 2018. "Which limb is it? Responses to vibrotactile stimulation in early infancy." *The British Journal of Developmental Psychology* 36(3): 384–401. https://doi.org/10.1111/bjdp.12224

Starch, D. 1910. "A demonstration of the trial and error method of learning." *Psychological Bulletin* 7(1): 20. https://psycnet.apa.org/journals/bul/7/1/20/

Suddendorf, T., and Butler, D. L. 2013. "The nature of visual self-recognition." *Trends in Cognitive Sciences* 17(3): 121–127. https://doi.org/10.1016/j.tics.2013.01.004

Waugh, W., and Brownell, C. 2015. "Development of Body Part Vocabulary in Toddlers in Relation to Self-Understanding." *Early Child Development and Care* 185(7): 1166–1179. https://doi.org/10.1080/03004430.2014.983915

Witt, A., Cermak, S., and Coster, W. 1990. "Body part identification in 1- to 2-year-old children." *The American Journal of Occupational Therapy: Official Publication of the American Occupational Therapy Association* 44(2): 147–153. https://doi.org/10.5014/ajot.44.2.147