Body representations may be considered in terms of both their first- and third-person characteristics. By first-person, we mean those aspects of body awareness and knowledge that are typically exclusively available to the “owner” of the body — the self. By third-person, we mean those aspects of body awareness and knowledge that are at least potentially available to any observer. Although there is overlap between first- and third-person characteristics — I can observe my own hand movements as I type just as anyone who happens to be watching me type can also observe these movements — there are also some qualitatively distinct characteristics. Most importantly, the owner of the body gets various kinds of information in a qualitatively different form and in some sense “privately.” For example, the proprioceptive sense yields a type of first-person information that allows for awareness of one’s movement and the relative position of various parts of one’s body. Thus, even without visual information, it is possible to determine fairly accurately both dynamic and spatial aspects of one’s body parts through proprioception. And so, in the dark, I know a lot more about my body than you do.

At least in humans, however, bodies are represented simultaneously in terms of both first- and third-person characteristics. Bodies have both externally observable properties and internal conditions. As adults, our experience and representation of our own bodies, for the most part, consist of integrated multisensory input. That is, when we perform an arm movement, we do not perceive the visual and proprioceptive information as separate yet correlated, but instead as unified and integrated. Similarly, the observed movement of our faces when we look in a mirror is perceived in integration with felt movement of our face. This integration depends upon the perfect temporal coincidence of the relevant information. In illustration, we know that adults will misperceive a fake rubber hand as their own under conditions where they are presented with a tight temporal correlation between first person information (either tactile or proprioceptive) from their own real hand and “third person” visual information from the rubber hand (e.g. Botvinick and Cohen, 1998; Dummer et al., 2009).

It is worth noting here that such integrated multimodal body representations are applied equivalently to both self and other (Barresi and Moore, 1996;
Moore, 2006) – there is a common code. The notion of a common code for the perception of others’ and one’s own body further demonstrates that body representations must, to some degree, involve an integration of both first- and third-person characteristics of the body (Barresi and Moore, 1996; Moore, 2006). However, it may also be the case that the ability to detect the presence or lack of a correlation between visual and proprioceptive information during observed movement aids in the development of self-awareness, particularly in the ability to discriminate between self-produced effects and other-produced effects (Lewis and Brooks-Gunn, 1979; Bahrick and Watson, 1985). We return to this issue later in the chapter.

Our primary goal in this chapter is to consider the developmental origins of the multimodal first- and third-person integration that supports human body representations. In particular, we focus on studies of visual-proprioceptive intermodal integration. In the first half of the chapter, we review those studies that have examined this topic in infants, ending with some recent work that has examined integration through the manipulation of the synchrony of visual-proprioceptive intermodal perception. This review will reveal that our current understanding of the development of visual-proprioceptive integration remains at a very early stage. Thus, in the second half of the chapter, we consider a variety of issues that may guide future work on this topic.

**Visual-proprioceptive intermodal perception in infants**

Infants are able to detect the relations between their own action and contingent events in the environment from very early in life. Indeed, such ability would seem to be essential to any form of instrumental learning. For example, Watson and Ramey (1972) placed 8-week-old infants in a crib and gave them experience with a visual event – a moving mobile – contingent upon the production of a particular movement – in this case, a head movement. Infants easily learned the correspondence or contingency between their own head movements and activating the mobile and increased their rate of movement. A large literature now exists on the parameters of infants’ instrumental learning and attests to the capacity of young infants to detect the relation between the first-person experience of action and the experience of external events (e.g. Rovee-Collier, 1987).

However, visual-proprioceptive body perception and representation is different. As noted earlier, when we perceive our own body movements we do not have an experience of correlated information. Rather, the multimodal sources of information are perceived as integrated. Integration in this context means a unified percep of the movement that has both visual and proprioceptive characteristics. Following the intersensory redundancy hypothesis of Bahrick and Lickliter (2000), we assume that this integrated perception is generated
The development of body representations

through selective attention to amodal properties, such as temporal characteristics, present in the intermodal stimulation, although we will not elaborate on this process further.

The method for exploring infants' visual-proprioceptive integration draws on the visual preference method for intermodal perception. For example, in classic work by Spelke and colleagues (e.g. Spelke, 1976, 1979; Spelke and Owsley, 1979), infants watched two movies while listening to a soundtrack played from a centrally located loudspeaker. The soundtrack matched one of the movies as, for example, when regular beats of a drum coincided with an object bouncing on a surface. Under these conditions, visual preference for one or other of the movies indicates that the correspondence between the matched movie and soundtrack has been detected. In general, in Spelke's studies, whereas there was no preference for either movie when played silently, the addition of the matching soundtrack led to a visual preference for the matched movie in infants under 6 months. When applied to visual-proprioceptive intermodal perception, the approach retains the presentation of two streams of visual information presented side-by-side on video monitors. Now, however, one of the monitors presents infants with live video of their own movements, whereas the other presents control video of similar movements that are not live. The proprioceptive information provided by the infants' experience of their own movements serves as the information stream that is matched to one of the movies (the live one). Thus, visual preference for one of the videos indicates detection of the correspondence between the visual and proprioceptive information.

A number of studies have employed this general approach to examine infants' visual-proprioceptive intermodal perception. Papousek and Papousek (1974), for example, placed 5-month-old infants in a highchair facing two television screens, one of which displayed a live video of the infant's face and one of which displayed a pre-recorded and therefore non-contingent video of the infant's face. Infants attended to the pre-recorded video for a significantly longer time, thereby demonstrating their ability to discriminate between the videos. Without direct visual information about their own face, infants had to rely solely on the detection of the relationship between the proprioceptive feedback provided by their facial movements and the visual feedback provided on the TV screens.

Employing a similar experimental paradigm, Bahrick and Watson (1985) sought to further elucidate infants' abilities at discriminating between self-produced and other-produced movement. In their first experiment, 5-month-old infants simultaneously viewed both a contingent (live) and a non-contingent video (pre-recorded video of a peer) of their body from the waist down for up to 4 minutes. Infants wore brightly colored, striped stockings to encourage them to attend to the screens and hide any distinctive cues infants may use to discriminate between the videos. Bahrick and Watson (1985) used legs as the visual
stimuli instead of faces as in Papousek and Papousek (1974) in order to avoid cues such as differential eye contact between the videos that might be used by infants to discriminate between videos. Differential eye contact could occur as a result of using pre-recorded videos of the infant’s face. Unless special precautions are taken, live video prevents the infant being able to make direct eye contact with his or her own image, that is, when the infant looks at the live screen the image in the live screen ‘looks’ back. This is compared to the pre-recorded video in which there will not be eye contact, adding another possible cue the infant could use to discriminate between the two videos. Also, as infants were viewing pre-recorded videos of other infants in Bahrick and Watson’s (1985) study, feature recognition, that is identifying a distinctive cue that indicates it is one’s own body (e.g. a mole) may have been used as an additional cue by which they discriminated between videos. Therefore having infants all wear the same stockings eliminated this possibility.

Infants looked significantly longer at the non-contingent video suggesting that 5-month-olds are capable of detecting contingencies between their own movement and the movement they view on the video. However, in this study a barrier such as a highchair tray did not impede infants’ visual access to their legs. It was therefore impossible to determine whether or not infants identified the visual-proprioceptive contingency or alternatively matched visual-visual information in order to discriminate between self- and other-produced movement.

To better understand the extent to which infants rely solely on visual-proprioceptive contingencies, Bahrick and Watson (1985) repeated the experiment; however, this time they occluded the infants’ view of their own legs, which meant that they had to rely solely on proprioceptive feedback from their legs and the visual information presented on the screens. Infants again preferred to look to the non-contingent view of their legs, adding further support to the hypothesis that 5-month old infants are capable of detecting intermodal (visual and proprioceptive) contingencies. A third study was conducted to rule out other possible explanations for discrimination, such as feature detection. Although the infants were all fitted with the same leggings, it is possible that they detected features such as leg length or width that may have aided in discrimination. Bahrick and Watson (1985) eliminated these cues by using a pre-recorded video of the participant for the non-contingent video as opposed to a pre-recorded video of a peer. Again, without the use of feature detection infants still showed a preference for the non-contingent video, further supporting the hypothesis.

Together these results raise other interesting questions. Why do the infants prefer the non-contingent video and is there a time in development when one might observe a preference for viewing the contingent video? In order to investigate this, Bahrick and Watson (1985) repeated the paradigm described above using 3-month-old infants; however, no overall preference for either video was seen. It did appear, though, that 3 months of age is a period of transition as the
participants were bimodally distributed, with one group preferring the contingent view and another group preferring the non-contingent view. Bahrick and Watson (1985) suggested that this pattern of results indicates that 3 months of age may serve as a transition period through which infants move from being interested in the contingent pattern of information specifying the self to being interested in the non-contingent pattern of information specifying the other.

Bahrick and Watson’s (1985) findings demonstrated that infants as young as 5 months of age are capable of discriminating between a contingent and non-contingent image of their own movement; however, the study did not isolate what kind of information infants used to make this discrimination as it did not control for the temporal or spatial aspect of the videos. Using 3- to 5-month-olds, Rochat and Morgan (1995) manipulated both viewing perspective and directionality in order to investigate the impact of spatial information in the detection of self-produced stimuli. Infants were reclined at a 60-degree angle, which impeded them from viewing their own legs, and presented with two different live views of their own legs on a TV screen. One view was referred to as the “ego” view (i.e. congruent), which presented the infants’ legs from the infant’s viewpoint. The second view was referred to as the “observer’s” view in which the infants’ legs were presented as if from an observer’s perspective (i.e. non-congruent). The two views differed spatially but were temporally identical. That is, leg movement was temporally synchronous between the two videos but due to the different views being showcased (i.e. ego versus observer) the legs moved either toward (observer) or away from (ego) the infants. Further, directionality was left-right reversed. Both of these manipulations vary spatial information. Results supported Rochat and Morgan’s (1995) prediction that infants would prefer the spatially non-congruent video, as nine out of ten infants in both age groups spent more time looking at observer’s view of their legs.

In a second experiment, Rochat and Morgan (1995) examined infants’ sensitivity to discrepancies in directionality alone. To do this infants viewed two videos as in the first experiment. Both images presented an ego view (thereby controlling for viewing perspective). However, one image displayed a left-right reversal (i.e. non-congruent), thereby manipulating the directionality of movement. Results replicated the findings of the first experiment in that infants showed a preference for the non-congruent view, suggesting that a change in directionality of movement is sufficient for an infant to identify a discrepancy in visual and proprioceptive feedback.

Having shown that infants are sensitive to the directionality of movement, Rochat and Morgan (1995) next isolated the viewing perspective (i.e. ego and observer) in order to determine whether or not it was a sufficient enough cue to discriminate between congruent and non-congruent information. In a third experiment, only viewing perspective was manipulated by again presenting infants with both an ego and observer’s view of their legs, while maintaining
directionality and temporal information. Contrary to expectations, infants did not show a preference for either the spatially congruent (i.e. ego view) or non-congruent (i.e. observer's view) video. This suggests that the preference observed in the first two experiments was driven by differences in directionality and not differences in perspective.

The studies described so far have either compared a live video with a recorded one (e.g. Bahrick and Watson, 1985) or two live videos in which spatial information alone was manipulated (e.g. Rochat and Morgan, 1995). The latter approach has demonstrated that 3- to 5-month-olds are capable of discriminating between two live videos of their own body on the basis of directionality. However, discrepancy in viewing perspective does not appear to be sufficient for discrimination if the temporal information remains contingent. This may be due to their inability to form a representation of their body from an observer's perspective. The former approach compares attention to synchronous and contingent intermodal information with attention to asynchronous and non-contingent information. As such, it is well designed to examine the discrimination of contingent intermodal information from non-contingent intermodal information. The results using this approach appear to be similar regardless of whether faces (Papousek and Papousek, 1974), legs and feet (Bahrick and Watson, 1985; Rochat and Morgan, 1995), or arms and hands (Schmuckler, 1996) are used as stimuli. The latter approach compares attention to two cases of intermodally synchronous but spatially divergent information. This approach is useful for examining the extent to which common spatial information is detectable from visual-proprioceptive intermodal sources.

However, we suggested at the outset of the chapter that human body representations involve an integration of first-person and third-person information, an integration that depends on the synchrony (or simultaneity) of those forms of information. Thus, a third approach is to compare attention to two cases where contingency is maintained but synchrony is disrupted. The idea here is to present infants with a visual preference situation in which one of the video monitors displays live visual feedback of the infants' movements while the other displays visual feedback of the infants' movements that has been delayed by a short amount of time. The goal is to determine the degree of temporal lag that can be detected. From our point of view, this approach is of particular importance for the examination of the integration of first- and third-person information because contingency is controlled – both videos display visual information that is contingent on the proprioceptively perceived information – but temporal coincidence is varied. Therefore, this approach isolates the temporal properties of visual-proprioceptive – or first- and third-person information – integration. Studies of the temporal properties of visual-proprioceptive integration may be seen as an examination of the window of simultaneity within which infants have a unified percept of the body.
Rochat and Striano (2000) were the first to report such a study. Using the preferential-looking paradigm, they presented 1- to 5-month-old infants with two views of their legs; one view was live while the other was delayed by 0.5, 1, 2, or 3 seconds. Infants viewed each delay for 1 minute in counterbalanced order. Infants did not show preferential looking in any of the delay conditions and this led Rochat and Striano (2000) to conclude that for young infants visual and proprioceptive information could be integrated over all delays between 0 and 3 seconds. However, given that this study produced only null results, it is in fact impossible to draw any firm conclusions. It is possible, especially given the relatively short exposure (i.e., 1 minute) at each level of delay, that the lack of preference observed reflected essentially random performance. Further, a failure to discriminate between the two videos based on a failure to demonstrate a preference should not exclusively be interpreted as an inability to detect a discrepancy as it is equally plausible that infants were capable of discriminating between the delayed and live images but simply did not have a preference.

To further investigate this phenomenon, Hiraki (2006) had 5- and 7-month-old infants view two displays of their own legs. One screen was live while the other was delayed by 2 seconds. The short delay ensures that the spatial information presented on each screen was as similar as possible. Seven-month-olds looked significantly longer at the delayed screen; however, there was no difference in looking time at the two screens for the 5-month-olds. Hiraki (2006) also reported that in a pilot study infants did not show a preference when the delay was 1 second, although it did not appear to be the case that delay was systematically manipulated. This study was the first to report a discrimination of a short delay in visual feedback of self-produced movements in infants and suggests that the temporal threshold for discrimination is between 1 and 2 seconds. The failure to find discrimination at 5 months might suggest that infants’ ability to detect an asynchrony between the visual and proprioceptive stimulation changes over this period of development, although such a conclusion must remain tentative, given the difficulty of making firm inferences from a failure to show visual preference in the younger infants.

We (Collins and Moore, 2008) have also conducted preliminary research following on from Hiraki’s approach. Like others, we used a visual preference approach. We presented forty-six infants (5–12-months-old) with two video images of themselves simultaneously – one live and one delayed by varying degrees (we used 1 second, 2 seconds, and 10 seconds as conditions in a between subjects design) for about 4 minutes. The delayed video was achieved through the use of a video delay unit (Prime Image Pipeline D1), which can insert delays in playback between 0 and 30 seconds in frame increments. Infants viewed video of their own faces in our study, as our pilot attempts using infants’ legs were not able to sustain infants’ attention for long enough. We then examined looking times to the delayed and live image and calculated the
proportion of time spent looking at the delayed screen. Our initial results showed that older infants in this age range demonstrated a preference for the delayed image when the delay was 2 seconds, but no infants demonstrated a preference when the delay was either 1 or 10 seconds. This initial study is encouraging in that it is consistent with the results of Hiraki (2006) as well as extending his findings to faces. The non-linear pattern of results is intriguing because it suggests that there may be two separate explanations for the lack of preference at 1 and 10 seconds respectively, although it needs to be replicated in a larger sample. The results from these three studies seem to suggest a developmental pattern in which infants younger than 5 months of age are integrating visual and proprioceptive input (e.g., not discriminating a temporal delay) whereas infants 7 months and older, as seen in both our own work and Hiraki’s (2006) study, become capable of identifying short temporal discrepancies between their visual and proprioceptive inputs.

We can see that very little research examining the temporal parameters of visual-proprioceptive integration has been conducted with infants. When one examines the research findings to date, it is clear that the current understanding of infants’ visual-proprioceptive temporal discrimination threshold is far from complete. As we have seen, all of the relevant studies have used the visual preference paradigm, which is a method with intrinsic limitations, and, apart from evidence of a discrimination of a 2-second delay in infants in the second half of the first year, they have generated mostly null results. The downside to the use of this visual preference paradigm is that if infants do not demonstrate a looking-time preference, the inference from the results is ambiguous. Indeed, there are three competing explanations for the null results. First, of course, is that the infants are unable to make the discrimination under study. A second is that the discrimination is possible for the infants but that they do not have a preference for either stimulus. Finally, a third is that there is a preference but that the method is not sensitive enough to detect it. We believe that there is good reason to believe that the null results so far achieved may reasonably be attributed to a lack of preference as well as a lack of sensitivity, rather than a lack of discrimination. In the next section, we consider a variety of issues that may inform our inferences from the available data and suggest further empirical approaches to determine more accurately the nature of the development of visual-proprioceptive integration.

**What is the temporal threshold of visual-proprioceptive integration?**

We draw on two sources of evidence to inform our understanding of this question. First, a small number of experiments with adults have provided relevant data on visual-proprioceptive integration. Over a series of trials,
Leube et al. (2003) asked adults to slowly open and close one hand while watching the produced movement on a video monitor. The video feedback was randomly delayed between 0 and 400 milliseconds (ms) with each trial lasting 3 seconds. Participants were asked to identify whether or not the resulting video they were viewing was delayed or live. Adults responded at about chance levels when presented with a delay of 80 ms. That is, when adults were presented with an 80 ms delay, approximately 50 percent of the time they identified the video as delayed, and approximately 50 percent of the time they identified the video as live. Delays shorter than 80 ms were predominantly reported as being live and, conversely, delays longer than 80 ms were predominantly correctly identified as delayed. We have recently replicated this approach, with adults requested to make arm movements that they could only view using video feedback. Our results also indicated a threshold of about 80 ms (between 2 and 3 video frames, or between 66 and 99 ms). These findings suggest that, at least in adults, the lower limit of discrimination is much shorter than what studies with infants have so far revealed, but leaves open whether the difference between current estimates of discrimination in infants and adults corresponds to genuine developmental differences or only methodological differences.

Second, other studies of infants have examined the temporal parameters of bimodal integration across visual and auditory streams of information (see e.g. Lewkowicz, 1996, 2000). The use of these two streams of information has the advantage that both are under experimental control, and thereby allows a precise assessment of the temporal window over which bimodal information may be integrated. Using a habituation/test method, Lewkowicz (1996) first presented infants from 2–8 months with simple events involving synchronous visual and auditory information. After habituation, test events with delays varying from 100 to 550 ms were presented. Lewkowicz reported that when the auditory information led the visual information by as little as 350 ms, infants showed discrimination of live from delayed stimuli. When the visual information led the auditory information, discrimination of live from delayed occurred at as little as 450 ms. This level of temporal discrimination is substantially shorter than that observed in the research on infant visual-proptioceptive integration (Hiraki, 2006).

It is possible, of course, that the temporal parameters of intersensory integration for visual and auditory information are different than for visual and proprioceptive information. Interestingly, however, the threshold estimates for visual-proptioceptive integration and visual-auditory integration in adults are remarkably similar. For example, Lewkowicz (2000) habituated adults to a bouncing object in which the auditory and visual information was synchronous. Following habituation, participants were presented with a variety of asynchronous trials in which the auditory stimuli (i.e. percussive bounce) preceded the visual stimuli (i.e. the object hitting a surface). Adults in this study were
able to detect temporal asynchrony when presented with a 65 ms delay between the sound and the bounce. The close correspondence between the temporal parameters for visual-proprioceptive and visual-auditory integration suggests no significant difference in the temporal information processing properties between the two types of bimodal integration. If the same is true for infants, then one might predict that temporal discrepancies discriminable for visual-proprioceptive events would be similar to those for visual-auditory events, i.e. about 350 ms.

From this brief review of other relevant research a reasonable conclusion is that the research on visual-proprioceptive integration in infants has not yet used appropriately sensitive methods. It appears that the current methods being used to explore this phenomenon need to be revised and improved upon in order to more clearly delineate infants’ capabilities with regards to asynchrony detection. We would recommend a number of modifications. First, it is important to note that the discrimination threshold for infants’ integration of visual and auditory information was established using a habituation/test methodology (Lewkowicz, 1996). Therefore, use of the habituation method for visual-proprioceptive integration is warranted. Ideally, infants should first be habituated or familiarized with live visual feedback of their movements and subsequently presented with visual feedback delayed by varying amounts of time.

Second, the study of visual-proprioceptive integration is compromised in comparison to visual-auditory integration in that it is not possible to exert control over one of the streams of information – proprioception – in the same way as one can with visual or auditory information. Furthermore, it is not possible with infants, as it is with adults, even to instruct them to move. In these paradigms, infants are free to move, or not move, as they wish. There tends to be considerable variability in the amount of movement that infants produce in these studies, with some producing little or no movement. But, for an infant to have even a chance of discriminating between the test videos they must move; if they do not, both videos will be appear identical. Therefore, those infants that produce little or limited movement are unlikely to be able to discriminate the test videos. Clearly it would be an advantage to be able to encourage infants to produce sustained movement during the trial, as adults tend to do. One solution to this issue might be to increase the salience of the visual stimuli by a dressing up the body part infants are to view. For example, when the focus is on leg movements, one might consider attaching tassels or lights to the infants’ socks that react to movement, thereby increasing the infants motivation to move their legs and producing enough movement to allow them to discriminate between videos.

In short, it is very unlikely that we currently have an accurate picture of the temporal threshold for visual-proprioceptive integration in infants. We propose that a more valid approach to determining the threshold would be to use a
habituation/test procedure along with strategies to maximize the movement that infants produce during the session as they attend to the visual feedback.

**How might the temporal discrimination threshold vary?**

We have suggested that the temporal threshold for visual-propiroceptive integration may vary with development. Adults can discriminate visual-propiroceptive asynchronies of around 80 ms, whereas the limited data to date show that infants can discriminate no less than a 2-second asynchrony, and even an optimistic estimate based on ideal experimental conditions would suggest that about 350 ms may be the lower limit for infants. However, our knowledge of the possible developmental pattern is extremely primitive. All we know at present is that infants during the second half of the first year have a threshold for asynchrony detection that is considerably longer than that of adults. We know nothing about how that developmental difference changes with age. Thus, future research will need to be directed at a cross section of ages in an attempt to map the change in threshold with age. This work will be of most interest when carried out in relation to other measures of self and body awareness. For example, it will be interesting to examine whether there are significant changes in visual-propiroceptive integration associated with developments in objective self-awareness, such as self-recognition during the second year. We return to this issue in the final section of the chapter.

A second issue researchers are confronted with when studying variability in the temporal discrimination threshold is the lack of control not only over the amount of movement, as discussed in the previous section, but also over the type of movements produced by the infants. Although it is likely that there is a real developmental difference in discrimination threshold, it is important to consider that the apparent difference in threshold between adults and infants may be, in part, dependent upon a difference in the extent to which the participants take an active role in trying to determine whether an asynchrony exists. One major difference between adult and infant delay detection studies is that adults are explicitly told the purpose of the study — to detect an asynchrony — whereas infants are simply left to observe their own movements. As a result, adults’ movements will tend to be oriented toward the conscious goal to determine whether or not the visual input is delayed from the proprioceptive experience. If discrimination threshold also varies according to type of movement, then it is possible that adult participants quickly learn during the experiment to produce the kinds of movements that are most helpful. In contrast, infants are more likely to produce random movements, which will not necessarily be advantageous for detecting small temporal delays between visual and proprioceptive feedback.
Recent work with adults in our lab suggests that the type of movement produced is an important factor in asynchrony detection. We hypothesized that movements that involved a sudden onset and offset would facilitate the participants' ability to detect delays, whereas movements that were of a continuous nature with no sudden start or stop would diminish their ability to detect delay between visual and proprioceptive information. To test this idea, we manipulated the type of movement participants produced while viewing video of their hand either live or delayed by small lags. Participants were seated at a table facing a video monitor and asked to place their arm behind a vertical divider attached to the table thereby blocking visual access to their arm. The monitor provided the participants with a view of their arm that was either live or delayed by 1, 2, 3, 4, or 5 video frames. For each trial, participants were asked to move their hidden hand and arm from side to side in either a continuous motion (i.e. back and forth without stopping) or discretely (i.e. back and forth with a quick start and stop). Trials were 5 seconds in length. In order to control the amount of movement produced, participants were asked to move from one side of the table (i.e. right to left) every second thereby producing five side-to-side movements during every trial. Participants completed five trials for every delay; one for their right and one for their left hand, in a counterbalanced order in either the discrete or continuous condition. After each trial, participants were asked to verbally report whether the trial was delayed or live.

Analyses examined whether the type of hand and arm movement produced differentially affected participants' ability to detect temporal delays between visual and proprioceptive perceptual input. There was a significant difference in performance across condition, in that the threshold for participants producing discrete movements was approximately 81 ms, as compared to participants producing continuous movements whose threshold was about 113 ms. This study, therefore, demonstrates that the type of movement produced, that is discrete or continuous, significantly affects adults' ability to detect delays between visual and proprioceptive input.

The significance of this finding for studies with infants is that we can be reasonably certain that infants are not spontaneously producing the kinds of movements that would allow optimal asynchrony detection and therefore that the current measures of that threshold overestimate it. Further, apart from the type of movement, it is conceivable that infants may simply not be producing enough movement in order to provide them with a sufficient amount of information, therefore undermining their ability to discriminate. This may be particularly true for studies that use the infant's face as the stimulus in that legs and feet perform gross motor movements as opposed to one's face, which primarily performs fine motor movements. Given the difference in the type of movement, it is possible that discriminating small temporal delays between one's own movement and the visual feedback of that movement is easier with gross
motor movements as seen with feet and leg movement. Further, the movements produced by legs and feet are more discrete than the movement produced by the face. Perhaps the more continuous movement produced by the face will make temporal asynchrony detection much more difficult in that there was not a sudden onset and offset of movement as seen with legs and feet.

**Visual-proprioceptive integration and the development of the objective self**

The kind of integration of visual and proprioceptive information based on synchrony reviewed in this chapter so far provides an important basis for body representations. However, it is well known that infants do not acquire a more objective or explicit sense of the self’s body until the middle to end of the second year (Brownell et al., 2007; Moore, 2007; Moore et al., 2007). The paradigmatic manifestation of this development is mirror self-recognition, whereby infants show self-directed behavior when confronted with an unusual mirror-image of themselves (Amsterdam, 1972), although there are other manifestations of the objective self (see Brownell et al., 2007; Moore et al., 2007). For some years, it has been claimed that the self-awareness underlying mirror self-recognition depends upon the integration of the visual and proprioceptive information available when attending to a mirror image of the self (e.g. Lewis and Brooks-Gunn, 1979; Mitchell, 1993; Povinelli, 1995).

We know that for toddlers, self-recognition depends upon the **synchrony** of the dynamics of the visually perceived mirror image and the dynamics of the proprioceptively perceived movements in front of the mirror. When synchrony is disrupted in the self-recognition task in toddlers, they fail to treat the visual image in the same way. A number of experiments have explicitly tested children’s awareness of the self using modified versions of the self-recognition task that use delayed video feedback as opposed to a mirror or live video feed. For example, Povinelli et al. (1996) videotaped an experimenter secretly placing a sticker on children’s heads while they played a game with the experimenter. Children were 2, 3, or 4 years or age – substantially older than those that can pass mirror self-recognition tasks under normal conditions. About three minutes after the placement of the sticker, the child viewed the video. They found that almost no 2-year-olds, 25 percent of 3-year-olds, and 75 percent of 4-year-olds reached toward the sticker. These results have been taken to demonstrate that children younger than 4 years of age did not understand how the delayed video image related to their current self (Povinelli et al., 1996). However, by 4 years, children are able to connect their current self with previous states of self into a temporally continuous sense of self. Therefore, based on these results, Povinelli et al. (1996) distinguished between an “online” sense of self developed at the
end of infancy and a "proper" or temporally extended self developed later in the pre-school period (see also Povinelli and Simon, 1998).

More recently, Miyazaki and Hiraki (2006) presented children with a similar video self-recognition task in which a short delay of 1 or 2 seconds was imposed. They argued that such short delays would be within the limits for online processing of visual-proprioceptive information and thus that 3-year-olds would be capable of "passing" the delayed self-recognition task with such short delays in the visual feedback. Following the surreptitious placement of the sticker on the children's heads, the researchers then asked the children to attend to a TV. The image presented on the TV was either a live image of the child or an image delayed by 2 seconds. They found that over 80 percent of 4-year-olds reached up and touched the sticker on their heads in both the live and delayed conditions. In contrast, almost 90 percent of 3-year-olds touched the sticker in the live condition but only 38 percent in the delayed condition. Miyazaki and Hiraki (2006) repeated the experiment using a shorter delay of 1 second and found that 71 percent of 3-year-olds were capable of passing the task, suggesting that the temporal limit for an online visual-proprioceptive information processing system is between 1 and 2 seconds. This finding suggests that even a delay as short as 2 seconds might be too much for children's ability to detect the correlation between the visually presented image and their body representation.

It is worth noting here that the degree of delay – between 1 and 2 seconds – that has been shown to disrupt self-recognition in toddlers (Miyazaki and Hiraki, 2006) corresponds to the degree of delay shown to be discriminable by infants in the visual-proprioceptive integration tasks (Hiraki, 2006). Does this coincidence suggest that both self-recognition and visual-proprioceptive intermodal integration manifest the same temporal parameters? And, if so, does this mean that mirror self-recognition depends on the same information processing mechanisms as visual-proprioceptive intermodal integration? We suggest that this coincidence actually masks a significant difference between these types of task. As we have seen earlier in the chapter, the degree of asynchrony that can be detected in visual-proprioceptive integration tasks shows a decline over development such that by adulthood, people can detect an asynchrony of 80 ms. In contrast, the degree of asynchrony that children are able to ignore in self-recognition tasks shows an increase over development, so that by 4 years children show self-recognition over delays of minutes, not just seconds. The fact that the temporal thresholds for these two types of task change developmentally in opposite directions suggests that the tasks depend on different processes with different developmental histories. Nevertheless, we argue that these processes are not independent; indeed they may well interact in the development of body representations of the self.

Visual-proprioceptive intermodal integration is fundamentally a basic perceptual-attentional-motor process. Coincident and correlated visual and
 proprioceptive information during movement are processed from early in life with the net results being the generation of integrated visual-proprioceptive representations of the body – what, in the context of its evolution, Barth et al. (2004) refer to as the “self evolved for locomotor flexibility” (SELF) system. It is important to emphasize that this self representation operates “on-line,” i.e. during movement, but that it also incorporates (in the original meaning of that term) both first-person (e.g. proprioceptive) and third-person (e.g. visual) information. Established through the infancy period, the self-representation forms a necessary basis for the more explicit body representations that emerge at the end of infancy. In self-recognition tasks, the child is faced with a situation in which they have to relate the acquired self representation to the available image in the visual medium (mirror or video) in such a way that that visual image is taken to be “of” the self representation. This comparison is at first facilitated by the common information available in both – the dynamics of the mirror or video image match those of the online self representation – and thus the image in the mirror is linked back to the self.

This process that we have described for self-recognition is a representational one rather than a perceptual one, even if it is at first tied to online processing. Later, as children become able to hold in mind the self representation for longer, they become able to bridge longer temporal discrepancies between the presented visual image and the self representation.

Are there implications for the temporal parameters of visual-proprioceptive integration? The empirical work remains to be done, but one suggestion is that the development of the more explicit self-representation feeds back into visual-proprioceptive integration. Earlier we pointed out that there are likely real developmental differences between infants and adults in the degree of asynchrony that can be detected between proprioceptive information and visual feedback. Infants can detect asynchronies of 2 seconds (Hiraki, 2006), whereas adults can detect asynchronies of as little as 80 ms. It is possible that this developmental difference reduces with age in a continuous way as information processing systems mature. However, perhaps a more intriguing possibility is that the difference is overcome in a more discontinuous way as children acquire a more explicit self-representation. If so, then one might predict that assessments of visual-proprioceptive asynchrony detection would show rapid changes at just those points in development that are associated with changes in self representation. For example, we are currently investigating whether the onset of mirror self-recognition is associated with a significant decrement in the asynchrony detection threshold. Toddlers between between 15 and 21 months of age will be tested on mirror self-recognition and on their performance in a visual-proprioceptive integration task, in which the visual feedback delay will be manipulated. A direct comparison of visual-proprioceptive asynchrony detection for children who do and who do not show self-recognition will be carried
out to allow us to determine if the asynchrony detection threshold is indeed associated with self-recognition. If such an association is found, this will provide evidence that a discontinuous developmental pattern of asynchrony detection is linked to the onset of an objective sense of self.

Conclusion

Human body representations involve both first- and third-person characteristics. Thus the integration of synchronous visual and proprioceptive information about the body and its movements is fundamental to body representation in infants and adults. Over the past twenty years extensive progress has been made concerning our understanding of the development of visual-proprioceptive integration, but there is still much work to be done. We believe that infants’ temporal discrimination threshold for visual-proprioceptive integration has been greatly overestimated. Comparison of infant research and that with adults suggests that this work has been hampered by a lack of sensitivity of the methods as well as by the challenge of exerting control over the experimental preparations. Nevertheless, there does appear to be real development in the temporal parameters of visual-proprioceptive integration. Further work is required to elucidate the nature of the developmental pattern and, in particular, the relation between visual-proprioceptive integration and the development of more explicit concepts of self. A more complete understanding of infant visual-proprioceptive integration will allow us to not only understand how infants learn about themselves but also how they learn to discriminate between themselves, others, and the world around them.

References


