An Embodied Perspective of Early Language Exposure

Hanako Yoshida (yoshida@uh.edu)
Joseph M. Burling (jmburling@uh.edu)
Department of Psychology, University of Houston, 126 Heyne Building
Houston, TX 77204-5022 USA

Abstract

Advances in developmental research has made it clear that word learning has a long beginning. Recent work has demonstrated that infants learn words at 6 months of age—that is, before the traditional "first word" milestone in productive language—which is a full year before the usual "naming explosion" in productive vocabulary. Before infants talk, walk, or even point, how can the earliest stage of word learning take place at all? We used recent technology that allowed us to zoom in on the point of view of infants and also the traditional room-view observations to document how infants' visual input is dynamically synchronized with their own participation, as well as from social input in the context of parent-child word learning play. The parents' task was to play with the child with a set of toys as they taught the toys to them. To specifically document the child's dominant view and their participation, we coded the size of the toy object on which the child was focused and who was manipulating the toy at the moment. The results reveal systematic and dynamic links between infants' view and their level of participation.

Keywords: embodied perception; word learning context; child-centered view

Introduction

Recent advances have made it clear that word learning has an early beginning. A new study has demonstrated that infants comprehend at least *some* words at 6 months of age—that is, before the traditional "first word" milestone in productive language (Bergelson & Swingley, 2012)—and a full year before the usual "naming explosion" in productive vocabulary (Goldfield & Reznick, 1990). What is the nature of this very early stage of word learning? What are the experiences when such infants hear words, and are they fundamentally different from the experiences of older infants? During the first 2 years of life, infants learn and refine a whole set of new motor skills that dramatically change the ways in which the body moves and interacts with the environment, and their social interactions consequently also change. Six-month-old infants are relative novices at reaching for objects and do not sit steadily, and so for the most part, objects are brought to them or perceived from a distance; 12-month-olds, in contrast, walk and bring themselves to objects, and 18-month-olds are mobile, socially skilled, and capable of physically achieving their own desires. Do these changes in the ways infants physically interact in the world also change the way they socially interact, and determine the nature of their visual experiences that support word learning?

Needless to say, there are a number of factors that contribute to the process of later word learning, including social cues such as eye gaze and gesture, prosody, language structure, input frequency, pragmatics, and many others. Yet recent evidence demonstrating effective word learning at a much earlier stage than previously thought suggests the need for investigation of the language learning environment at the earliest stage. Moreover, learners, even very young babies, actively engage in the world by contingently responding to the social gestures of others. In doing so, infants—and perhaps in different ways at different ages—distort regularities and carve up the input in systematic ways. This means that one cannot really consider the input separately from the learners' own actions, because the learner selects and creates the input. Importantly, word learning takes place at any stage of early parent-child interactions, and most often co-occurs with infants' active exploration of objects where their head, hands, or body, and eyes coordinate and shape "input" and possibly optimize their view for initial learning of labels.

Recent work that motivated this present study investigated how early learners create their own visual input by observing the first-person-view during toy play (Smith, Yu, & Pereira, 2011; Yoshida & Smith, 2008; Yu, Smith, Shen, Pereira, & Smith, 2009). These studies typically used a small camera attached to young children's foreheads and documented how this child-centered view provides insight into factors relevant for early word learning. In the work presented here, we sought to study developmental changes in this child-centered view by longitudinally following children from 6 to 18 months in order to detail the quality of their visual input, how it changes over time, and the potential dynamic relation to their rapidly growing physical capacities in the context of parent-child play. Documentation of precise changes in visual experiences mediated through children's own physical growth is essential to studying how infants' changing motor skills serve as strong filters of their early learning experiences, because effective visual attention determines effective learning.

Emergence of Language Learning

One crucial question in language learning is how children overcome referential ambiguity—that is, how children selectively extract the referents to map the corresponding words. Effective learning of such a complex skill requires a highly selective process of sampling information from the environment; observing infants' eye gaze reveals a great deal about what they are processing and learning. Infants' language learning, however, takes place in a dynamically changing context—that of physical growth and motor development. Indeed, acting and knowing are inseparable aspects of human life, and early learning heavily depends on a person's physical capacities and environment. The potential link between language learning and bodily constraints may be uniquely different for different developmental stages. Yet, there are few studies looking directly at the how early bodily experiences influence language learning in relation to an individual's task involvement. As a first step, we document how the early visual input in a language learning tasks relates to the physical constraints of the child's sensory-motor engagement during the task.

Object Size and Language Learning

If infants shape the effective view of objects through their own physical growth, consequential actions, and mature social partners' participation, the degree of referential ambiguity or perceptual accessibility may be partially addressed at the level of sensory motor coordination. In a recent study of 18-month-old infants (even some as young as 17 months) in a naturalistic parent-child play context, both parent and infant physically participated in the play (holding objects) at the same level, but objects dominated the child's view (much bigger, thus occluding other objects) when the object was held by the child (Pereira, Smith, & Yu, under revision; Yu & Smith, 2012). Infants of this age (18 months) are capable of producing smooth head turnings and can manage a stable posture, suggesting that their own body coordination may help optimize their focus, and bodily movements having a relatively stronger role for the optimal view at this age. This quality of visual experiences was also related to learning object names (Yu & Smith, 2012), suggesting that how well 18-month-olds can zoom in on objects has important implications for their successful word learning. What happens if younger infants, who seem capable of learning words (Bergelson & Swingley, 2012), do not yet have the physical capacity to coordinate their bodies to support their optimal view? We specifically targeted younger infants to further our understanding of the role of physical development in the organization of early visual input. We looked at the size of the object in the infants' view at the moment of it being held by the infant or parent.

Visual Attention, Motor Skills, and Social Development

Infants and young children's ability to follow eye gaze and pointing directed toward objects is considered a major milestone in the development of joint attention. This type of attention emerges in infants as young as 3 months. It helps early learners identify word meaning and serves a potential communicative function. For infants to develop this ability, they have to coordinate their head and body so that they can capture the adult's head and eyes as well as the target. As they become more efficient at coordinating their view, they can produce head turnings to follow the adult's movement to maximize the input (Butterworth & Grover, 1990; D'Entremont, Hains, & Muir, 1997). If a stable view containing all the important elements for effective attention and learning is important, then one might think that 6-month-olds, whose head movements are less active (thus much more stable) and views more distal, would be in a better position to capture all the relevant components than more advanced infants who are capable of producing more movements. However, there is one study using a peripheral distracter to measure the focus of 7- to 10-month-olds' attention which found infants were more attentive—and showed potentially more effective attentional shifts—when engaged in *active play* than when receptively observing the task stimuli (Oakes, 1994). Another previous observation points to an interesting potential role of a social partner, suggesting that caregivers naturally respond to children's motor skill changes in order to support the transformation (e.g. Zumbahlen, 1997). These studies raise the question of how, exactly, an infant's own physical participation shapes the development of visual experiences and possibly influences the social partner's physical participation. If infants are not capable of bringing toys to their view, does this motivate the parent to help the infant gain a better view by bringing and holding the objects themselves? Or is the quality of the infants' view tightly linked to their own physical growth (with minimal parental support), thus emerging poorly yet dramatically improving as a function of their physical growth and advances in body coordination?

In the present study, we investigated two components that together influence an infant's visual focus: (a) who brought the object to the child's view (parent/social partner or the child), and (b) how parents' participation reflects their child's physical constraints. Studying how visual and bodily experiences support language learning reveals what contributes to effective visual attention, such as joint attention, and how such effective attention may emerge through both the child's own experiences and parental participation.

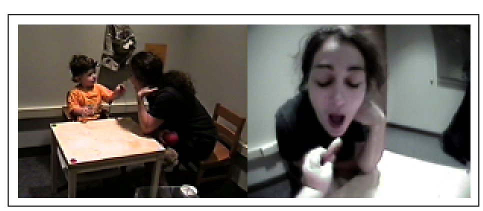

Figure 1: Snapshots showing the first-person (right) and third- person (left) view from Yoshida and Smith (2008).

Figure 2: Snapshot of third-person views of the same child across five sessions.

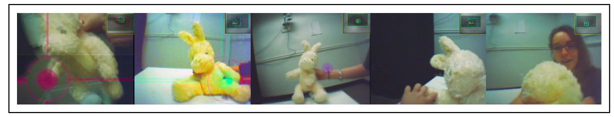

Figure 3: Images corresponding to a distinct developmental time point, which demonstrates the changes in object size over time and parent-child interactions.

Child-centered View

In a head camera study of parent-child play, Yoshida and Smith (2008) recorded 18-month-old infants' perspective in the context of a parent teaching a set of early learned words while the parent and child sat naturally at a table. The results revealed that the child's view was much more constrained (captured fewer items) compared to the room view (see Figure 1) and provided evidence of the coupling of head and eye movements. Yoshida and Smith (2008) independently measured eye gaze direction (frame by frame via a camera fixated on the infant's eyes) and head direction and found that eye and head directions were highly correlated, such that 87% of head camera frames coincided with independently coded directions of eye gaze.

In this study, we used the same procedure to capture younger infants' point of view. We monitored the development of their view by following a set of infants from 6 months to 18 months, testing them every 3 months to document how their focus changes over time, and how their own object exploration and parent's support (in bringing objects) relates to their view (demonstrated by Figures 2 and 3).

Because of the current focus on much younger children than have been studied in the past, we used a slightly different setup—the infant was supported in a child's chair and the parent sat diagonal to the infant—yet we maintained the relational position between the infant and parent used in the previous study. Moreover, a previous head camera study (Smith et al., 2011) evaluating head camera images from different seating arrangements (sitting naturally in a chair or on the floor) found no differences in any aspect of infant or parent behavior as a function of the task geometry, suggesting minimal impact from the current modification to the sitting arrangement. Another modification is that we used the head-mounted portable eye-tracking device (Figure 4) instead of a head camera to ease the difficulty of calibrations. With the head-mounted eye-tracking device, eye gaze can be directly measured and calibration issues can be better addressed without having the infants point to the object of their fixation (typically camera adjustments are made by asking infants to point to or touch what they are looking at).

To run the experiment, we placed an infant in the chair, and then one experimenter put a light weight head-mounted eye-tracking device on the infant's forehead while another experimenter distracted the infant by introducing him or her to a set of attractive toys (which made noise and had colorful moving lights). A standard camera recorded the play scene from a corner of the room, so that we collected synchronized third- and first-person-views, as in Yoshida and Smith (2008). In the present study we used naturalistic word learning in the context of playing with toys. The parent was instructed to teach the infant a set of words by selecting the toys the parent thought were most appropriate to play with from a collection of available toys.

Figure 4: Snapshot of an infant (at 6 months) with a head-mounted eye-tracking device.

Study

With a setup similar to that of Yoshida and Smith (2008), we observed seven infants, starting when they were 6 months old and ending when they were 18 months old, in the context of a parent-child word-learning play session (see Figs. 2 and 3). We specifically investigated the infants' moment-to-moment quality of object view by measuring the size of the focal object (as percentage of image pixels), that is, the image size of the largest object in the infant's view, and determining if the optimal visual moments for an object relate to who is holding it (the infant or parent), and whether this contributes to the infants' development of visual experiences.

Method

Participants Seven parent-infant dyads were brought to the laboratory five times (3-month intervals). There were three male and four female infants. Three additional infants began the study but did not contribute to the data because of refusal to wear the measuring equipment. The mean age of the infants was 6.4 months (first visit), 9.2 months (second visit), 12.4 months (third visit), 15.3 months (fourth visit), and 18 months (fifth visit), with each range less than 13 days.

Stimulus and Materials There were eight unique toy objects in a box, located on the floor beside the parent's chair for easy access to them by the parent. Each toy was a naturalistic toy whose name is listed on a developmental vocabulary inventory, McArthur Child Development Inventory (Fenson et al., 1993). Figure 5 shows eight toy objects that were used for parents to teach the typical early-learned words (open, bunny, car, bottle, cookie, eat, drink, put). The order and the duration of playing with them were controlled by the instructions given through the experimenter.

Figure 5: Toy objects used for the word-teaching play sessions.

Procedure Prior to entering the experimental room, the parents were presented with a set of instructions on their type of interaction. Parent and infant then entered the experimental room and were asked to sit in the designated chairs. The experimenter started one camera attached to the wall (5 meters away from the chairs) for recording the third-person view, and then put the head-mounted eye-tracking device on the child's forehead. The calibration procedure took place to adjust synchronization between camera placement and eye position. The experimenter left the room as she instructed the parents to pace their play and teaching according to the guidance from the audio prompt, which said the name of one of the objects every 40 seconds. The entire session took approximately 6 minutes.

Results

The results suggest that a child's view changes dramatically over the course of 12 months of development. At 6 months, the child's view was moderately selective. The focus of the object is then reduced in size at 12 months, then becomes most selective with a sharp increase in object size by 18 months. These changes may be the development of optimal focus, and the results from the room view suggest that the infant's frequency of reaching and holding of objects has a systematic influence on the infant's own view. Furthermore, the parent's participation level also reflected the development of this optimal view.

We first focus on two variables for the present study in order to address the nature of developmental growth in visual experiences. These variables of interest are the quality of the focused object (how the object looks to the child in terms of the size of the object from their perspective) and also the participation level of the parent and child interacting with the object, determined by the frequency in which either manipulates, holds, or touches the object within a given trial. Eye gaze was measured and tracking data was overlayed onto the video to capture the infant's central view. This was used to aid the coding process and to help determine to which object the child was currently attending.

The coding process involved evaluating individual frames of video to determine the size of the object in view and whether or not the parent and child were interacting with the object. For every 5 seconds of video at 30fps, multiple data entry fields were randomly generated, giving a total of 250 rows of data per video. Each video corresponds to a single subject and experimental session at a specified developmental time period. Research assistant coders were thoroughly trained to identify the object in view based on the tracking data, and to determine the area of the object in pixels as a proportion of the entire child-centered view. The mean and standard errors of the size proportions are based on sampled frame evaluations and then averaged across all the frames for each 320 second session. A separate set of coders were trained to enter information about the individual that was manipulating the object at each of the sampled frames (if any). Coders made judgments about who was interacting with the object and the type of object seen on screen. The frequency in which parent or child manipulated the object throughout the session was calculated as a proportion based on the total number of cells randomly generated for that session.

The first result, which is evident in Figure 6, shows that objects within an infant's view during the first session take up a moderately sized proportion of their view, yet at 12 months (their 3rd visit), the average size of objects focused within their view is reduced dramatically, and reaches its lowest point during this age. The proportion of all objects in terms of pixel area was greatest at 18 months of age (5th and final visit). There were significant developmental differences in object size between 6 months and 9 months, ($t = 2.56$; $p < .01$), and between 15 months and 18 months, ($t = 2.28$; $p < .05$). There were no significant differences found between ages 9 and 12 months, nor between 12 and 15 months.

The developmental shifts in focus is interesting and suggests that these optimal visual experiences do not come as easy, and do not appear to be a straight linear growth. At the 6, 9, and 12 month points, the size of objects in their visual field seem to fluctuate, and the dip at 12 months may be of developmental significance and contain important information about their physical growth and parental involvement. Initial parental interactions with the infant may help set the stage for determining

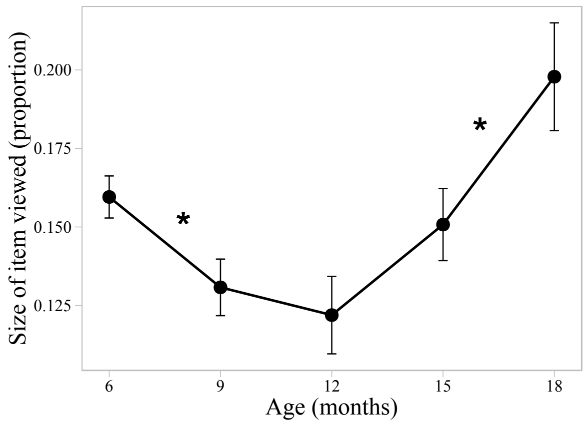

Figure 6: The average size of all objects as a proportion of the camera view at each longitudinal time point.

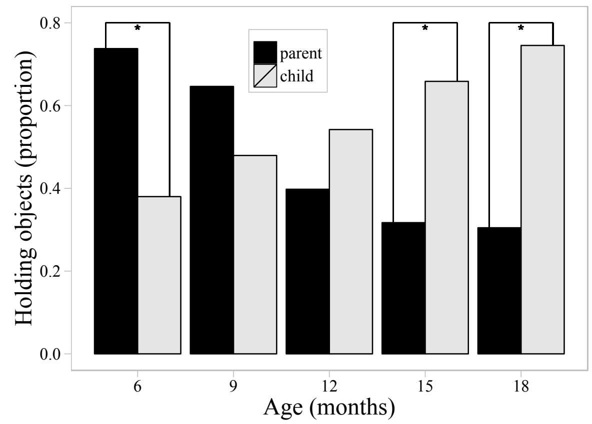

Figure 7: Frequency in which participants interact with objects over the course of 5 longitudinal sessions shown as a proportion of the number of coded frames.

optimal visual experiences while infants learn to properly handle and manipulate objects later on their own.

An infant's optimal view has been characterized by very few objects being focused upon and dominating the infant's visual field (Smith et al., 2011; Yoshida & Smith, 2008; Yu, Smith, et al., 2009). But, the magnitude of this type of focus appears to change dramatically in one year (see Figure 3), where at least by 6 months, infants show initial steps toward the development of language learning based on both physical growth and environmental interactions. To address this question, we analyzed the variable where coders made a judgment about who was interacting with the object. In absence of a longitudinal perspective, results might suggest that across all developmental time points (6 to 18 months) objects were held equally frequent between infants and parents. Yet, as can be seen in Figure 7, as early as 6 months, parents held the object of focus reliably more than infants ($\chi^2 = 6.25$; $p < .05$). This contrasts with later periods, in which children are interacting with objects much more frequently than adults. At 15 months, infants began to reliably hold the objects more often than adults ($\chi^2 = 4.9$; $p < .05$), and by 18 months, the magnitude of this difference in object interaction was most dramatic ($\chi^2 = 7.79$; $p < .01$). There were no significantly reliable differences found in object manipulation during the 9 month and 12 month time periods.

During the early stages of development, parental involvement seems to account for a majority of the visual experiences gained by infants, while at later stages, infants are responsible for their own experiences as they hold and manipulate objects with greater frequency. At the 12 month time point, there was no reliable difference in who was holding the object. Interestingly, this is the point where their center view quality drops to as small as 12.1%. This may be the point at which the parent begins to demonstrate less involvement in shaping the child's visual experiences—and thus a drop in object focus—which allows for the infant to independently start shaping what comes into their view and determine their own optimal focus, an ability that they apparently are not immediately successful in controlling during the early months.

General Discussion

The present findings suggest the contexts for early and later word learning are quite different, and that they shift from greater parental control to greater infant control, with parents ensuring early on that the named object is visually dominant, and the infant plays a more active role later in development. Clearly these findings are just a first step toward understanding the possible, developmentally changing pathways through which infants learn words. Recent advances make clear that infants start learning words much earlier than previously thought (as young as 6 months), before the traditional "first word" milestone in productive language (Bergelson & Swingley, 2012). The present results suggest that parents may isolate objects—focusing on objects one at a time, zooming in on a single object—to help this learning. Later, this isolation may be most effectively done when the child actively engages in object manipulation (holding and bringing objects) as suggested by the Yu and Smith (2012) results.

The inflexion point in the visual size of the named object at 12-months is intriguing, and suggests a possible transition from more parent control toward more infant controlled learning, and the parent following of the child's interests when naming. The 12-month mark has been noted by others as a period of change in social interactions. For example, whereas very young infants appear to automatically follow the eye-gaze of another (Farroni, Massaccesi, Pividori, & Johnson, 2004; Hood, Willen, & Driver, 1998) and follow head movement—not eye-gaze—when head and eye direction are in competition (Brooks & Meltzoff, 2005; Gergely, Nádasdy, Csibra,

& Bíró, 1995), 12-month olds appear to require more coherent cues as well as contingent interactions to follow these cues (Tomasello, Hare, Lehmann, & Call, 2007; Moore & Corkum, 1998; Johnson, Ok, & Luo, 2007). In brief, as infants' motor and cognitive skills make them more independent, social interactions and the structure of word learning may change in systematic ways.

The present finding that 12-month-old infants do not seem to experience this optimal view suggests that both parents and infants may be working out this transition. This finding does not mean that the infant and parent are not involved with each other or have no interest in playing. Rather, infants and parents appeared to participate relatively equally. This leads to new insight into how action coordinated through dominant participation (one agent or the other, but not both) maximizes support for the development of optimal view, and it raises a number of novel developmental predictions about the role of action coordination such as emergence of join attention and social contingencies in learning.

Linking word learning to the changing sensory-motor skills of infants may also be key to understanding the co-morbidity of language delay and motor deficits in developmental disorders such as autism (Iverson, 2010). Studying changes in visual experiences mediated through a child's own physical growth and experiences is a way to gain new perspectives on the nature of embodied language learning.

Acknowledgments

This research was supported by the National Institutes of Health grant (R01 HD058620), the Foundation for Child Development, and University of Houston's Grants to Enhance and Advance Research (GEAR) program. We especially wish to thank the families who participated in this study. We also thank the undergraduate research students in the Cognitive Development Lab for their support in coding for the present study.

References

Bergelson, E. & Swingley, D. (2012). At 6 to 9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences of the USA*, *109*, 3253–3258.

Brooks, R. & Meltzoff, A. N. (2005). The development of gaze following and its relation to language. *Developmental Science*, *8*(6), 535–543.

Butterworth, G. & Grover, L. (1990). Joint visual attention, manual pointing, and preverbal communication in human infancy. In M. Jeannerod (Ed.), *Attention and performance XIII: motor representation and control* (pp. 605–624). Lawrence Erlbaum Associates, Inc.

D'Entremont, B., Hains, S. M. J., & Muir, D. W. (1997). A demonstration of gaze following in 3- to 6-month-olds. *Infant Behavior & Development*, *20*(4), 569–572.

Farroni, T., Massaccesi, S., Pividori, D., & Johnson, M. H. (2004). Gaze following in newborns. *Infancy*, *5*(1), 39–60.

Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J. P., . . . Reilly, J. S. (1993). *The MacArthur Communicative Development Inventories: User's guide and technical manual.* Baltimore: Paul H. Brokes Publishing Co.

Gergely, G., Nádasdy, Z., Csibra, G., & Bíró, S. (1995). Taking the intentional stance at 12 months of age. *Cognition*, *56*(2), 165–193.

Goldfield, B. A. & Reznick, J. S. (1990). Early lexical acquisition: rate, content, and the vocabulary spurt. *Journal of Child Language*, *17*, 171–183.

Hood, B. M., Willen, J. D., & Driver, J. (1998). Adult's eyes trigger shifts of visual attention in human infants. *Psychological Science*, *9*(2), 131–134.

Iverson, J. M. (2010). Developing language in a developing body: the relationship between motor development and language development. *Journal of Child Language*, *37*, 229–261.

Johnson, S. C., Ok, S.-J., & Luo, Y. (2007). The attribution of attention: 9-month-olds' interpretation of gaze as goal-directed action. *Developmental Science*, *10*(5), 530–537.

Moore, C. & Corkum, V. (1998). Infant gaze following based on eye direction. *British Journal of Developmental Psychology*, *16*(4), 495–503.

Oakes, L. M. (1994). Development of infants' use of continuity cues in their perception of causality. *Developmental Psychology*, *30*(6), 869–879.

Pereira, A., Smith, L. B., & Yu, C. (under revision). A bottom-up view of toddler word learning. *Psychological Bulletin and Review.*

Smith, L. B., Yu, C., & Pereira, A. (2011). Not your mother's view: the dynamics of toddler visual experience. *Developmental Science*, *14*, 9–17.

Tomasello, M., Hare, B., Lehmann, H., & Call, J. (2007). Reliance on head versus eyes in the gaze following of great apes and human infants: the cooperative eye hypothesis. *Journal of Human Evolution*, *52*(3), 314–320.

Yoshida, H. & Smith, L. B. (2008). What's in view for toddlers? Using a head camera to study visual experience. *Infancy*, *13*, 229–248.

Yu, C. & Smith, L. B. (2012). Embodied attention and word learning by toddlers. *Cognition*, *125*(2), 244–262.

Yu, C., Smith, L. B., Shen, H., Pereira, A., & Smith, T. (2009). Active information selection: visual attention through the hands. *IEEE Transactions on Autonomous Mental Development*, *2*, 141–151.

Zumbahlen, M. R. (1997). *The role of infant locomotor onset in shaping mother-infant communication* (Doctoral dissertation, University of Illinois at Urbana-Champaign).