Using infant and toddler testing methods in language acquisition research

1. Introduction

Prior to the early 1970's, theories of language acquisition were largely based on analyses of children's speech productions. More recently the study of child language has been enriched by the development of new testing methodologies that allow researchers to investigate what infants know about language before they begin to speak. These procedures have uncovered remarkably sophisticated linguistic knowledge and learning abilities in very young infants.

In this chapter, we discuss three infant testing procedures that are well suited to investigate the receptive language abilities of young children: the visual fixation procedure (VFP), the headturn preference procedure (HPP), and the preferential looking procedure (PLP). Throughout our discussion, we point the reader towards example studies and additional readings. Note that this chapter is not meant to be a comprehensive review of all infant-testing methodologies. We do not discuss the conditioned headturn procedure (CHP), the high amplitude sucking procedure (HASP), or electroencephalography (EEG). For an excellent review of the CHP, see Werker, Polka, and Pegg (1997; see also Gout, Christophe, & Morgan, 2004). Likewise, for a review of the HASP, see Jusczyk (1997). For a useful discussion of advances in EEG research, see Friederici and Thierry (2008). Finally, for some examples of how hybrid methods can be created by combining elements of multiple different child testing procedures, see Kooijman, Johnson, and Cutler (2008) or Hollich (2006).

Before delving into the details of specific infant testing methodologies, we would like to express a brief precautionary note for readers who may be contemplating their first line of infant studies. Infant research is exciting, however it can also be enormously time consuming and frustrating. The testing methodologies outlined in this chapter require practice to master, and creating appropriately designed infant experiments can be trickier than it appears. For example, when lacking experience with a particular testing methodology, it can be difficult to determine how many trials to include in an experiment or whether to expect a novelty or familiarity preference (Hunter & Ames, 1988; Houston-Price & Nakai, 2004). Infant researchers must deal with these and other difficulties before they can collect useable data. Even small things, like getting the lighting just right in the testing booth, or knowing how to instruct caregivers, can have a substantial impact on the success of an infant research project.

A few potential pitfalls are very pervasive in infant studies, and being aware of them is important for designing good experiments as well as to critically evaluate published studies. First, controlling for both experimenter and caregiver bias is absolutely essential. Many infant testing procedures (e.g. HPP and VFP) rely on experimenter's ability to objectively observe children's behavior and relay this information to a computer via a button box. Because the experimenter must interpret infant's behavior (e.g. direction or length of fixation), these procedures are susceptible to experimenter bias. If experimenters have any way of knowing what stimuli are being presented at any given time, the experimenter may inadvertently influence the outcome of the experiment. To guard against experimenter bias, ensure that experimenters do not know what condition an infant is assigned to. It is also important that experimenters cannot hear what sound files are playing during the experiment. Thus, the optimum setup for running an infant-controlled procedure is to have the testing booth inside a double-walled sound attenuated booth and the experimenter outside of the booth in a separate control area. However, booths are expensive. Alternatively, a cheaper option is to use sound-absorbent paneling. If a booth exceeds the lab budget, the experimenter should listen to masking music over noise-reducing aviator-style headphones. The best masking music contains few pauses and is mixed with sound files used in the experiment. Regardless of whether or not a booth is used, caregivers should always listen to masking music because they will be in direct contact with the child participants. Note that cheap headphones will not guard the experiment from parental bias (Pinto, Fernald, McRoberts, & Cole, 1999). Use headphones that block caregivers from hearing the stimuli presented to child participants while at the same time minimizing the amount of sound that escapes from the headphones (sound escaping form the headphones could potentially be heard by the child during the experiment). Tight fitting active noise reducing headphones work very well.

It is also important to carefully consider what counts as a dropout (i.e. an infant whose data will not be included in the final analysis). If experimenters look at their data before deciding which children are dropouts, then there is a danger that they will preferentially exclude the data from children who do not perform as expected. For this reason, it is a good idea to have experimenters assign infants a fussiness rating immediately after the testing session (before the data has been examined). In addition, if the dropout is extremely high, it may be necessary to reconsider the design of the study. The greater the dropout rate, the lower the validity of the study. For example, if there is a very long experiment and two out of every three infants tested fail to complete the test, then the one third of the infants who actually make it through the study may not be representative of the normal population at large. The number of children tested in the procedures described below depends on the experimental design. In general, a typical study should be run with minimally 16 successful participants, and some experiments may require as many as 48 successful participants. By ‘successful’, we refer to those participants whose data can be included in the final data set, e.g. the number of infants left after excluding those participants who are exclude for reasons such as failing to complete the experiment due to fussiness. Every additional experimental manipulation typically requires an average of 16-20 additional successful participants.

Another important consideration is the language background of the participants. As language experience has a crucial impact on the participants’ language knowledge, it is imperative to obtain the language input of the participants in the study. For example, be sure to determine the specific language(s) infants are being exposed to and the percentage of exposure they receive from the different languages. Depending on the experiment design and question, it may also be important to set a criterion for inclusion in the study. For example, participants must receive at least 80% exposure to a language be considered a monolingual learner of that language. Note that some parents have great difficulty estimating how often they speak to their children in different languages. For example, when asked what percentage of time their children are exposed to English, they might answer 80%. But later in questioning, they may also report that their child hears French 60% of the time. Clearly a single child cannot hear English 80% of the time and French 60% of the time! Thus, it is important to take the time to make sure parents understand the question they are being asked so that they can provide the most accurate answer possible.

Finally, with between subjects designs or when comparing infants' performance across different experiments, avoid allowing spurious factors such as time of test to be confounded with condition or experiment. For example, testing infants in December is often challenging because ear infections and colds are common and the holiday season puts children off of their schedule. At the same time, testing infants in the middle of a heat wave is also problematic if the testing area is not air conditioned. The problem is that if all of the children in one condition are tested under suboptimal conditions, and all of the children in another condition are tested under optimal conditions, then there may well be more than one explanation for any differences observed between test conditions.

In sum, collecting good infant data is not always easy. There is a limit on how well one can prepare to run an infant study by reading about it. For this reason, discussing the research plan with an active infant researcher, and visiting an already established infant lab, should be seen as an important step in making sure that one's own research program gets off on the right foot. That said, we will now describe the three methodologies we have made the most use of in our own research.

2. Visual fixation procedure

The VFP, like most infant testing methodologies, can trace its roots back to the pioneering visual perception work by Robert Fantz (1967). A variant of this procedure can be used to assess infants' listening preferences (Cooper & Aslin, 1990). The variant we will discuss here was developed to study auditory discrimination (e.g. Zamuner, 2006).

2.1 Rationale

This procedure works because an infant's interest in a sound is linked to how long they will look at a concurrently presented visual stimulus. In a typical study participants sit on their caregivers lap and watch a checkerboard (or some other simple visual stimulus) paired on each trial with a sound (e.g. a repeating syllable). In the beginning, both the sound and checkerboard are very interesting. However, as time goes on, the sound and checkerboard become less interesting and infants gradually decrease their looking time to the screen. Infants are considered habituated once they decrease their looking time to the checkerboard by a preset criterion. They are then presented with the checkerboard paired with a new token of the same sound (e.g. new token of same syllable) or a new sound (e.g. new token of new syllable). If infants can discriminate the two sounds, then they will be more interested in the new sound than the old sound, i.e. they will dishabituate to the new sound as indicated by an increase in their looking time to the checkerboard (see Figure 1).

There is also a word-learning variant of the VFP that uses the same logic as the basic VFP paradigm. The word learning variant is often referred to as the 'switch' paradigm, and it is used to examine word-object associations rather than simple discrimination. In both variants, the auditory stimulus might be the same, but in studies with word-objects, participants are habituated to visual images such as objects and in discrimination, to visual stimuli such as a checkerboard. Participants are habituated to a novel word-object pairing (Stager & Werker, 1997) to a preset criterion (judged by a decrease of looking time). They are then presented with either the same word-object pairing that was presented during habituation (same-trial) or with a novel word-object pairing in which the auditory and visual stimuli is novel (switch-trial).[1] If infants have successfully learned the word-object association during the habituation phase, their looking times should continue to be low or decrease (continued habituation) on the same-trials. If infants are able to recognize that the association presented in the switch-trials is different, looking times should increase on those trials as infants dishabituate. Many studies combine the use of a pure discrimination version of the VFP with the word association 'switch' variant (e.g. Dietrich, Swingley & Werker, 2007; Stager & Werker, 1997).

2.2 Linguistic variables

The VFP paradigm has been used to study many areas of infant cognitive development (e.g. Cohen & Cashon, 2002). Within the field of language acquisition, the procedure has been used to investigate a wide variety of phenomenon including phonemic and phonetic discrimination (Burns, Yoshida, Hill, & Werker, 2007; Dietrich et al., 2007; Zamuner, 2006), language discrimination (Johnson, Westrek & Nazzi, 2005; Mehler, Jusczyk, Lambertz, Halsted, Bertoncini, & Amiel-Tison, 1988), word-object association (Stager & Werker, 1997; Werker et al., 1999) and bilingualism (Fennell, Byers-Heinlein & Werker, 2007).

2.3 Participants

The VFP is well suited for use with infants between 4 and 20 months of age, although it has been used with younger infants (Mehler et al., 1988). The design is limited with age groups younger than 4.5 months as infants' ability to maintain head support develops around this time. With older age groups, it may be difficult to maintain toddlers' attention over the course of the study. Depending on how quickly a participant habituates, a typical study lasts between 5 and 10 minutes. Use of this method to identify children at risk for speech or language delay is currently being investigated (Bernhardt, Kemp, & Werker, 2007). Different age groups require different habituation criteria, with younger participants requiring a lower habituation criterion (e.g. 50%) and studies with more complex stimuli typically requiring a higher habituation criterion (e.g. 65%).

2.4 Description of procedure

There are two phases to a VFP study: the habituation phase and the test phase. During habituation, infants are presented with a repeating audio stimulus paired with a visual

[1] At times it is important to counterbalance the order of stimulus presentation across habituation and test. For example, it can be theoretically important to determine whether there is an asymmetry in infants' ability to discriminate lexical contrasts, such as the switch between *bin* to *din* versus *din* to *bin* (Fikkert, 2007). Here, the order of presentation would become another experimental condition.

stimulus. In one version of the test phase, participants are presented with new and old trials, and looking times to the two types of test trials reveal whether infants discriminated between the old and new sound or word-object pairing. In other versions of this paradigm, half of the infants are assigned to a control condition (test trials are the same as habituation trials) and have are assigned to the test condition (test trials are different from the habituation trials). Note that both designs control for spontaneous drift back to the mean.

The length of the habituation trials can be fixed or infant-controlled. Fixed length trials continue regardless of whether infants are attending (e.g., a trial may be 15 seconds long and include 10 tokens of the word 'bin', which play regardless of whether the infants look at the screen). Infant-controlled trials terminate when an infant looks away for longer than a preset time (e.g. two seconds). Another important parameter to consider is whether participants must complete a fixed number of trials or whether they must meet a habituation criterion before proceeding to test. With a fixed number of trials, participants must complete the prespecified number of trials before moving to test. Note that this design can result in a familiarization paradigm rather than a habituation paradigm. Alternatively, a habituation criterion based on a sliding window can be used. For example, mean looking time during the first three trials of the experiment are compared to the mean looking time during the last three trials the infant has completed (e.g. trials 1 through 3 are compared to 3 through 6, then 4 through 7, etc). Once looking time over the last three trials decreases by a preset percentage relative to the initial three trials, the test phase begins. The number of test trials can also vary, depending on the study design.

VFP studies should also include pre-test and post-test trials. Typically these trials present a novel stimulus (such as a moving toy waterwheel), which may or may not be paired with an audio stimulus. These trials allow one to measure fatigue over the course of the experiment. Although looking times to the habituation stimuli may decrease over the course of the study, looking times to the pre-test and the post-test trials are not expected to differ significantly (Werker et al., 1999). If an infant does not dishabituate to the post-test stimuli, then their data should be excluded from the study.

2.4 Analysis and outcomes

Before analyzing the results data, taped experimental sessions need to be digitized, and participants' eye movements coded off-line (some free programs are available for off-line coding, see PLP). Typically a within subjects design is used, and a Repeated Measures ANOVA is sufficient to analyze infants looking times to same versus different trials.

3. Headturn preference procedure

The headturn preference procedure (HPP) was first developed by Fernald (1985) to examine infants' listening preference for infant- versus adult-directed speech. Since then, the HPP has undergone substantial modification (Jusczyk, 1997; Jusczyk & Aslin, 1995). Here we discuss two standard variants of the HPP. The first variant is a simple preference procedure suitable for examining what knowledge infants have when they enter the lab.

The second variant, often referred to as the Modified HPP, involves the addition of a training phase prior to testing.

3.1 Rationale

The HPP and the VFP both rely on the link between infants' looking behavior and their degree of interest in sounds to infer underlying linguistic knowledge or capabilities. However a major difference between the VFP and the HPP is that the latter requires infants to make a head turn response. In addition, the VFP is most often used to test discrimination whereas the HPP is rarely used to test discrimination, although there are exceptions to this generality (Nazzi, Jusczyk, & Johnson, 2000; Soderstrom & Morgan, 2007). Nonetheless, by far the most common use of the HPP is to test recognition of grammatical (as opposed to ungrammatical), familiar, or familiarized stimuli (e.g. Gout et al., 2004; Hirsh-Pasek et al., 1987; Johnson, Jusczyk, Cutler, & Norris, 2003; Mandel, Jusczyk, & Pisoni, 1995; Mattys, Jusczyk, Luce, & Morgan, 1999; Santelmann & Jusczyk, 1998).

3.2 Linguistic variables

The HPP has been used to investigate a wide variety of linguistic phenomenon such as phonotactic sensitivities (e.g. Mattys et al., 1999; Zamuner, 2006), artificial language learning (e.g. Saffran, Aslin, & Newport, 1996; Johnson & Jusczyk, 2001) ability to deal with phonetic variability (e.g. Singh, 2008), sensitivity to subphonemic contrasts (e.g. Johnson, 2008; McMurray & Aslin, 2005), and acquisition of grammatical dependencies (e.g. Santelmann & Jusczyk, 1998).

Simple preference variants of the HPP presents infants with two types of stimuli (sound files) and measures which stimuli infants listen to longer. Infants typically listen longer to grammatical (or frequent) than ungrammatical (or infrequent) constructions. For example, Santelmann & Jusczyk found that 18- but not 15-month-olds listen longer to sentences containing grammatical sentences like *The archeologist is digging for treasure* over ungrammatical sentences like *The archeologist can digging for treasure*. This version of the HPP has also be used to test infant's knowledge of the sound structure of their language (e.g. Mattys et al., 1999; Jusczyk, Cutler, & Redanz, 1993, Zamuner, 2006) and infant's understanding of how words are pronounced (e.g. Swingley, 2005; Vihman, Nakai, DePaolis, & Halle, 2004).

The second and perhaps more commonly used variant of the HPP, often referred to as the Modified HPP, involves two phases: familiarization and test. The addition of the familiarization phase allows one to test what information infants can extract from the speech stream, within a brief 2 to 3 minute familiarization phase. The Modified HPP was first used by Jusczyk & Aslin (1995) to study word segmentation. During familiarization, infants were presented with two alternating passages, each containing repetitions of a target word (e.g. *dog* or *cup*). During the test phase, infants were presented trials consisting of repetitions of the two familiarized words (e.g. *dog* and *cup*), whereas the other half consisted of repetitions of two words that were not in the familiarization passages (e.g. *bike* and *feet*). By 7.5 months, infants listened longer to familiar words than unfamiliar words, indicating that they had segmented these items from the continuous speech passages played in the training phase. Since this initial study, the modified HPP has been used to address a wide variety of additional topics, including artificial language learning (e.g. Gómez & Gerken, 1999; Johnson & Jusczyk, 2001; Saffran, Aslin, & Newport, 1996), language discrimination (Nazzi, Jusczyk, & Johnson, 2000), and sensitivity to prosodic structure (e.g. Johnson & Seidl, 2008). Moreover, longitudinal studies have suggested that infant's performance in HPP studies may be predictive of their subsequent language development (Newman, Bernstein Ratner, Jusczyk, Jusczyk, & Dow, 2006).

3.3 Participants

The HPP is optimally designed for use with infants between 6 and 9 months of age. However, as soon as infants have developed the muscles needed to control head turns (~4.5 months), they can be tested with the HPP (Mandel et al., 1995; Nazzi et al., 2000). At the same time, children as old as 18 to 24 months or more have also been tested with this method (e.g. Nazzi, Patterson, & Karmiloff-Smith, 2003; Santelmann & Jusczyk, 1998). Note, however, that drop out rates are higher with older children. As with the VSP, this method has also been used with children at risk for speech or language delay (Van Alphen, de Bree, Gerrits, de Jong, Wilsenach, & Wijnen, 2004; Wilsenach, 2006).

3.4 Description of procedure

The HPP is generally run in a dimly lit 6' x 6' booth. The child sits in the center of the booth on a caregiver's lap facing the front wall. A green light is mounted about 3 feet from the floor centered on the front wall, and a red light is mounted in the same manner on each of the two side walls. A camera is hidden just below the front green light, and an observer watches the infant's looking behavior on a monitor, preferably in a separate control room. At the beginning of each trial, the green light blinks. Once the infant orients towards the green light, the experimenter pushes a button on a button box, which causes the front green light to stop blinking and one of the two side red lights to begin blinking. Once the infant orients toward the side red light, the experimenter pushes another button and sound begins to play from a speaker hidden directly behind the blinking light. Sound continues to play until the infant looks away for more than two seconds. Then the next trial begins. Throughout the experiment, the experimenter simply tracks the infant's looking behavior to and away from the blinking lights, and relays this information to a computer via a button box. The computer is equipped with custom-designed software that controls the lights and presentation of sound files.

An HPP study generally lasts 2 to 5 minutes. Various factors, such as participant age and stimulus complexity, affect the optimal length of an experiment. HPP studies often present two types of items (e.g. half grammatical utterances and half ungrammatical utterances) over a course of 10 to 16 trials. Sometimes each infant is presented with as many as three different types of trials (e.g. Johnson, 2008). Including more than 3 types of test trials is not optimal since infant data is highly variable and unless an incredibly large number of infants are tested, there may not be enough power to pull out significant differences between conditions. Experiments presenting three or more types of test items might work best when run with a between subjects design (e.g. Johnson et al., 2003).

3.5 Analysis and outcomes

The dependent measure in HPP experiments is looking time to the side red lights. Figure 2 presents the results of a fictitious word segmentation study (error bars indicate Standard Error). Imagine that some infants were familiarized to words embedded in infant-directed speech (IDS), whereas others were familiarized to words embedded in adult-directed speech (ADS). At test, both groups of infants receive same test trials: repetitions of the familiarized and unfamiliarized words. The solid bars represent looking time to familiar words and the striped bars represent looking time to unfamiliar words. Note that the infants familiarized to IDS look longer to familiar words than unfamiliar words, whereas the infants familiarized to ADS do not. Visual inspection suggests that only those infants who had been familiarized with infant-directed speech succeeded in segmenting the familiarized words from speech. These data could be analyzed with a 2 (Looking Time: Familiar versus Unfamiliar) by 2 (Condition: IDS versus ADS) Mixed Design ANOVA. The ANOVA is called a Mixed Design because the Looking Time factor is a within-subjects measure (the same infant was presented with both familiar and unfamiliar test items) whereas the Condition factor is a between subjects factor (the same infants were not tested in Condition 1 and Condition 2).

With HPP it can sometimes be difficult to predict a novelty (i.e. longer looking to unfamiliar stimuli) or a familiarity preference (i.e. longer looking to familiar stimuli). For example, although natural language segmentation studies typically result in familiarity preferences (e.g. Johnson, 2008; Jusczyk & Aslin, 1995), artificial language segmentation tasks often result in novelty preferences (e.g. Johnson & Jusczyk, 2001; Saffran et al., 1996). Novelty effects are generally associated with tasks that are easier. For example, by decreasing the length of exposure to an artificial language, infants' looking times during the test phase can be caused to shift from a novelty preference to a familiarity preference (see Thiessen, Hill, & Saffran, 2005, for a detailed discussion). The important thing to remember is that for most HPP studies, it is the presence of a looking time difference rather than the direction of looking time difference that is crucial. However, the interpretation of *some* HPP studies is dependent on the direction of looking preference. In these studies, it is absolutely essential that a baseline study be carried out to determine the predicted direction of looking preference (see Johnson & Seidl, 2009, for further discussion).

Sometimes infants fail to exhibit either a novelty or familiarity preference. In this case, HPP results can be difficult to interpret because it does not necessarily indicate failure at a task. Perhaps the stimuli are not interesting or age appropriate, or the experiment may be too long (for discussion, see Aslin & Fiser, 2005; Gout et al., 2004). If the stimuli are simply uninteresting or unappealing, looking times may be short and drop out rates high. Block analyses can be useful in such a situation. Shifts between novelty and familiarity preferences can also impact HPP studies (e.g. Johnson & Seidl, 2009; Thiessen, Hill & Saffran, 2005). Some labs use data cleaning techniques to make sense out of data plagued by novelty/familiarity shift difficulties (Gómez, Bootz, & Nadel, 2006; McMurray & Aslin, 2005). These data cleaning techniques can come in handy when it is suspected that a portion of the participants tested are exhibiting a novelty preference while the others are exhibiting a familiarity preference. Some have argued that null results are so problematic with HPP data that the procedure itself is of limited utility (Aslin & Fiser, 2005). Although null results are a serious problem with this testing procedure, we nonetheless find the HPP to be a valuable tool for studying infant development. Alternative methodologies such as EEG and NIRS have just as serious interpretational problems as the HPP. Indeed, no infant testing methodology is perfect. A goal for the future may be to combine testing methods to compensate for weaknesses in individual testing procedures (Kooijman, Johnson, & Cutler, 2008).

4. Preferential looking procedure

Prior to the development of the preferential looking procedure (PLP), children who were highly mobile but not yet necessarily producing many words (e.g. ~14 to 26 months) were extremely difficult to work with in an experimental setting. The development of the PLP, which works well with this age group, led to major breakthroughs in the field of language development. There are many variants of the PLP, which go by different names. For example, in some labs this procedure is referred to as the intermodal preferential looking paradigm (Hollich, Hirsh-Pasek, & Golinkoff, 2000; Houston-Price, Mather, & Sakkalou, 2007) whereas in other lab a very similar procedure is referred to as the Looking-While-Listening Procedure (Lew-Williams & Fernald, 2007).

4.1 Rationale

In this section, we will simply use the general term PLP to refer to any procedure presenting pictures to children while recording their visual fixations.

4.2 Linguistic variables

The PLP is an incredibly versatile testing procedure that has been successfully used to study a wide variety of language phenomenon. It has been used to examine verb constructions (e.g. Hirsch-Pasek & Golinkoff, 1996), questions (Seidl, Hollich, & Jusczyk, 2003), grammatical gender (Lew-Williams & Fernald, 2007; Van Heugten & Shi, 2008), and sentence processing (e.g. Thorpe & Fernald, 2006). The PLP has also been used to study early word comprehension (e.g. Fernald, Swingley, & Pinto, 2001; Houston-Price, Mather, & Sakkalou, 2007; Swingley & Aslin, 2002) and word learning (e.g. Halberda, 2003; Hollich et al., 2000). More recently, researchers have started to look use the PLP to look at semantic development (e.g. Fernald, Zangl, Portillo, & Marchman, 2008; Johnson & Huettig, submitted). There are countless ways the PLP could be used study language development, and in the near future we fully expect the PLP to be put to work in creative new ways to address formerly intractable questions. Longitudinal studies have suggested that children's performance in PLP studies is related to subsequent language development (Fernald, Perfors, & Marchman, 2006).

4.3 Participants

Variants of the PLP have been used with children as young as 6 months (Tincoff & Jusczyk, 1999) or as old as 3.5 years (e.g. Lew-Williams & Fernald, 2007). There is no upper limit to the age that can be tested –simply have to adjust the experiment design to suit the age group being tested. Adult eyetracking studies use the same underlying logic as the PLP, and can be seen as an adult version of the same procedure. However, one major difference between the PLP and adult eyetracking is that the PLP is often run without an eyetracker. It is feasible to handcode PLP data frame-by-frame because children have limited attention spans, and PLP experiments usually have fewer trials than adult eyetracking studies.

4.4 Description of procedure

In this procedure, children sit on a caregiver's lap while viewing a video presentation on a large TV (or a number of screens). Caregivers wear headphones and/or opaque sunglasses so that they are unable to inadvertently bias their child's looking behavior. It is best to run the study in a dimly lit room with few visible distractions. Ideally the procedure is run in a booth or small room dedicated to the procedure, however if this is not possible, curtaining off the test area usually works well). Each trial generally consists of two pictures shown side by side (it is also possible to have three pictures using a triangular configuration with two on top and one on the bottom). The pictures should be equal in size and as equally interesting as possible. In between trials, we use an animated object accompanied by an attractive sound to entice children to fixate the center of the screen.

In Figure 3, a figure outlines the timecourse of a single trial in a fictitious word comprehension study. At the beginning of a trial, a star on a black background appears for two seconds. Then the test trials begins when pictures appear on the screen, and a child-directed voice says 'wow, see the ball?'. The release of the initial /b/ in the target word 'ball' occurs exactly 3 seconds after the test pictures first appear. The test pictures remain on the screen for a total of six seconds.

It is essential to have precise control over the timing of audio and video stimulus presentation in PLP studies. Otherwise, it will be impossible to analyze and interpret the data. There are many ways to achieve accuracy and control of stimulus presentation. One can present stimuli via a computer, and have the precise moment that trials begin as well as the moment target words are uttered time stamped on the data videos by customized software (see Fernald et al., 2008, for a detailed discussion). It is also possible to run PLP studies using a child-friendly eyetracker along with commercialized software (e.g. Tobii or Eyelink Remote). Another extremely simple and low tech strategy we have used in our lab is to create our stimulus videos on a computer in such a way that we know precisely how long each trial lasts, and how long after the onset of each trial the target word is uttered. We visually mark the onset and offset of trials by having a white background on the screen during test trials and a black background on the screen during intertrial intervals. We then export the test videos to digital tape. During the experiment we present our stimuli by playing the digital video for children while using another digital video camera to videotape their eye movements (Johnson, Huettig, & McQueen, 2008; Van Heugten & Shi, in press). If the lighting in the room is just right, then the onsets and offsets of trials show up as a dramatic lighting change reflected on the wall behind the child (a mirror on the back wall of the testing booth can achieve the same effect). The recordings of children's eye movements are then digitized and handcoded (at the time this chapter was written, a freeware eye movement program called SuperCoder was available from Dr. George Hollich at Purdue University). Using this procedure frees one from developing a time-stamping system since the videos are in digital format and therefore played back occurs at the same rate every time. In Europe, where the PAL video system is used, each video frame is 25 ms long. In North American, where the NTSC video system is used, each video frame is 29.97 ms long.

In addition to test trials, it is important to include filler trials to maintain children's interest. The number of trials in an experiment will depend on the nature of the study and the age group being tested. PLP videos typically last about 4 minutes. Note that some PLP studies have a short training phase before the beginning of the test phase. For example, children might be taught a new word or presented with a story. In this case, the number of trials may need to be decreased to prevent the experiment from getting too long.

PLP data coding is typically done offline from video recordings. This is why PLP data can be collected by experimenters who have not yet received enough training to run so-called infant-controlled procedures, such as the HPP or VFP. In our labs, the PLP procedure is often the first that students learn because all of their data can be easily checked by more experienced researchers. Regardless of how experienced coders are, periodic reliability checks are a must to ensure that all coders are judging eye movements consistently.

4.5 Analysis and outcomes

Deciding which dependent measure(s) to examine in any given experiment depends on the goal of the research. If a basic demonstration of overall comprehension (as opposed to a fine-grained measure of the precise timing of comprehension) is adequate, then analyzing overall looking times to the target versus non-target image(s) is appropriate (e.g. Van Heugten & Shi, in press; Houston-Price et al., 2007). When overall proportion of looking times is used, there are a few crucial decisions that need to be made. First, when will the window of analysis begin and end? Children need some time to process speech sounds and initiate eye movements (e.g. Swingley & Aslin, 2002) and most children do not just look to the target object and continue to stare until the trial ends. Instead, children tend to look to named targets and then after a second or two they may shift their gaze. For this reason, researchers typically choose a short window of analysis starting around 300 ms after the onset of the event of interest (e.g. the onset of a target word). Proportion of time spent fixating one type of target over another during this window of analysis can be compared to chance performance, or to performance in other conditions. It can also be compared to the proportion of looking time to target during a window of time preceding the onset of the interesting event (Swingley, 2003). This latter analysis is particularly useful if some of the images in a study are inherently more interesting to look at than others. Another decision to make is how to calculate the proportion of looking times. Should the denominator be the total duration of the window of analysis? Or should it be the total amount of time spent looking at either the target or distractor, i.e. excluding time spent looking away from the screen?

If the fine-grained timecourse of comprehension is crucial to a PLP study, then reaction times are often taken as a measure of processing speed (see Fernald et al., 2008, for a detailed description of reaction time measures). In this case, it is also useful to plot proportion of looks to target as a function of time. In Figure 4, we have plotted the data from our fictitious word comprehension study. Error bars indicated Standard Error. If enough trials are included in the experiment, data can be analyzed in much the same way as adult eyetracking studies (e.g. see analysis by Huettig & McQueen, 2007). Note however that the appropriate method for analyzing adult eyetracking data is still controversial (see Mirman, Dixon & James, 2008).

5. Advantages and disadvantages

There are a number of advantages and disadvantages with the three procedures discussed above. The following discussion will focus on issues specific to each individual procedure.

The VFP works with a wide range of ages and provides a very sensitive measure of speech sound discrimination. The experimental setup is also relatively simple; at the time this chapter was written software for running this procedure was freely avaliable.[2] Many researchers use the design, so there is a large research community on which to draw knowledge (which is good given that choosing the proper habituation criterion, trial length, etc. can be tricky). However, the procedure also involves an artificial exposure to speech (e.g. it is hard to imagine a situation in the real world where infants would hear the same syllable repeated 4 dozen times in isolation) with no measure of online speech processing. And finally, attrition rates can be high when testing older infants.

The HPP allows researchers to address language acquisition questions that would otherwise be extremely difficult or impossible to study. For example, the procedure enables researchers to easily study infant's perception of long stretches of speech. However, there are also disadvantages to the HPP such as the preference shifts and null results discussed above. Another drawback is that there is at present no commercially available software to run the HPP, so custom designed software and hardware are required to set up this procedure.

The PLP is an excellent infant testing methodology because the procedure itself is relatively easy to set up and run, and the results are generally easy to interpret. We also find that children of all ages really enjoy this procedure. However, one drawback is that very young infants may not shift their attention between two interesting events as efficiently as older children do. Also, coding PLP data by hand can be extremely time-consuming.

6. Do's and Don'ts

As with the previous section, there are a number of factors to consider for each procedure, and more general considerations and pitfalls to avoid that apply too all procedures.

[2] See homepage.psy.utexas.edu/homepage/Group/CohenLab/habit.html.

- Keep the experimental design simple. Allowing a study to become to long or complicated can be a recipe for disaster.

- Be very careful to thoroughly counterbalance. For example, for the HPP, counterbalance the side and order of presentation, and make sure experimenters are blind to what type of trial is being presented. In other paradigms, control for potential preferences for one picture over the other (e.g. have each picture serve as both a target and a distractor). Note that infants' attention and looking times are longest at the beginning of the study. Therefore, it is important to counterbalance across infants which trial appears at the onset of the study. This is especially important for the HPP.

- Try to use stimuli that sounds pleasant to infants (e.g. infant directed or happy speech). Infants respond best to these types of stimuli. Remember that sudden unexpected sounds can be frightening for a child.

- Be certain that there is good control over the timing of video and audio presentation.

- Do not allow parents to point and talk during the study. The parent's job is to basically be a big boring armchair throughout the experiment. Do not chat and play with children too long before beginning testing. If older children are having too much fun in the waiting room, they will be reluctant to enter a boring testing room.

- In studies that use habituation, carefully consider whether or not to include participants who do not reach the habituation criterion. Infants' preferences for familiar versus novel stimuli can change over the course of a study; therefore, including both habituators and non-habituators potentially leads to different groups in the study, i.e., those that are showing a familiarity versus a novelty preference (Cohen, 2001).

- Establish strict coding guidelines for what counts as a 'look' to ensure high coding reliability. Also consider whether to set a criterion for short looks, such as those that are less than 1 or 2 seconds long, which may be excluded from the analyses. At times, researchers will exclude short looks because on those trials participants will not have heard the crucial stimuli.

- If experimental sessions will be coded off-line or coded for reliability, be sure that recordings are of adequate quality for coding. There is no fully satisfactory way to do reliability coding on HPP data because the presentation of the experimental stimuli is reliant on the individual participants looking times during the study and the experimenter's coding of looking (this is also partly true with the VFP if the study is designed using infant-controlled trials.) Thus, for the HPP (and sometimes the VFP) it is absolutely essential from the onset to guard against bias and have only well-trained experimenters test infants.

We close this chapter with thoughts towards the future. As infant testing becomes more common, the range and methodologies available to study language acquisition continues to grow. Our discussion has only begun to scratch the surface of possible approaches to infant studies. Although all of the procedures we have discussed have their distinctive strengths, no method is perfect. Indeed, all testing procedures, whether they use behavioral or physiological measures, are simply indirect reflections of the underlying mental representations and processes of interest to researchers. Nonetheless, these indirect measures have provided us with a wealth of knowledge about early language development. In the future, we hope to see current infant testing methodologies adapted to study individual as opposed to group performance (see Houston, Horn, Qi, Ting, & Gao, 2007, for discussion). This in turn would make it more feasible for researchers to study special populations where only a very limited number of participants are available for testing. It would also open the door to exciting new avenues of research where individual language development can be tracked and more easily.

References

Albareda, B., Pons, R., Sebastián-Gallés, N. (2008, March). *The Acquisition of Phoneme Categories in Bilingual Infants: New Data from a New Paradigm*. Poster presented at the International Conference on Infant Studies, Vancouver, BC.

Aslin, R.N., & Fiser, J. (2005). Methodological challenges for understanding cognitive development in infants. *Trends in Cognitive Sciences*, *9*, 92-98.

Aslin, R.N., & Mehler, J. (2005). Near-infrared spectroscopy for functional studies of brain activity in human infants: promise, prospects, and challenges. *Journal of Biomedical Optics*, *10*, 11009.

Bernhardt, B.M., Kemp, N., & Werker, J.F. (2007). Early word-object associations and later language development. *First Language*, *27*, 315-328.

Bortfeld, H., Wruck, E., & Boas, D. (2007). Assessing infants' cortical response to speech using near-infrared spectroscopy. *NeuroImage*, 34, 407-415.

Burns, T.C., Yoshida, K.A., Hill, K., & Werker, J.F. (2007). Bilingual and monolingual infant phonetic development. *Applied Psycholinguistics*, *28*, 455-474.

Cohen, L.B. (2001, April). *Uses and Misuses of Habituation: A Theoretical and Methodological Analysis*. Symposium paper presented at the Society for Research in Child Development Meeting, Minneapolis, MN.

Cohen, L.B., & Cashon, C.H. (2002). Infant perception and cognition. In R. Lerner, A. Easterbrooks, & J. Mistry (Eds.), *Handbook of psychology: Vol. 6, Developmental psychology* (pp. 65-89). New York: Wiley and Sons.

Cooper, R.P, & Aslin, R.N. (1990). Preference for infant-directed speech in the first months after birth. *Child Development*, *61*, 1584-1595.

Dietrich, C., Swingley, D., & Werker, J.F. (2007). Native language governs interpretation of salient speech sound differences at 18 months. *Proceedings of the National Academy of Sciences of the U.S.*, *104*, 16027-16031.

Fantz, R. (1967). Pattern perception in newborn infants. *Science*, *140*, 296-297.

Fennell, C.T., Byers-Heinlein, K., & Werker, J.F. (2007). Using speech sounds to guide word learning: The case of bilingual infants. *Child Development*, *78*, 1510-1525.

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behaviour and Development, 8*, 181-195.

Fernald, A., Perfors, A., & Marchman, V.A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology*, *42*, 98-116.

Fernald, A., Swingley, D., & Pinto, J.P. (2001). When half a word is enough: Infants can recognize words using partial phonetic information. *Child Development, 72*, 1003-1015.

Fernald, A., Zangl, R., Portillo, A.L., & Marchman, V.A. (2008). Looking while listening: using eye movements to monitor spoken language comprehension by infants and young children. In I. Sekerina, E.M. Fernandez, & H. Clahsen (Eds.), *Developmental psycholinguistics: On-line methods in children's language processing* (pp. 97-135). Amsterdam: John Benjamins.

Fikkert, P. (2007). Acquiring phonology. In P. de Lacy (Ed.), *Handbook of phonological theory* (pp. 537-554). Cambridge, MA: Cambridge University Press.

Friederici, A., & Thierry, G. (Eds.). (2008). *Trends in Language Acquisition Research: Vol. 5, Early Language Development: Bridging brain and behaviour*. Amsterdam: John Benjamins.

Gómez, R.L., & Gerken, L. (1999). Artificial grammar learning by1-year-olds leads to specific and abstract knowledge. *Cognition*, *70*, 109-135.

Gómez, R.L., Bootzin, R., & Nadel, L. (2006). Naps promote abstraction in infants. *Psychological Science*, *17*, 670-674.

Gout, A., Christophe, A., & Morgan, J. (2004). Phonological phrase boundaries constrain lexical access II. Infant data. *Journal of Memory & Language, 51*, 548-567.

Halberda, J. (2003). The development of a word learning strategy. Cognition, 87, B23-B34.

Hirsh-Pasek, K., & Golinkoff, R. M. (1996). The preferential looking paradigm reveals emerging language comprehension. In D. McDaniel, C. McKee, & H. Cairns (Eds.), *Methods for assessing children's syntax* (pp. 105-124). Cambridge, MA: MIT Press.

Hollich, G. (2006). Combining techniques to reveal emergent effects in infant segmentation, word learning, and grammar. Language and Speech, 49, 3-19.

Hollich, G., Hirsh-Pasek, K., & Golinkoff, R. (2000). Breaking the language barrier: An emergentist coalition model of word learning. *Monographs of the Society for Research in Child Developmen*t, *65* (3).

Houston, D.M., Horn, D.L., Qi, R., Ting, J., & Gao, S. (2007). Assessing speech discrimination in individual infants. *Infancy*, *12*, 119-145.

Houston-Price, C., & Nakai, S. (2004). Distinguishing novelty and familiarity effects in infant preference procedures. *Infant and Child Development, 13,* 341-348.

Houston-Price, C., Mather, E., & Sakkalou, E. (2007). Discrepancy between parental reports of infants' receptive vocabulary and infants' behaviour in a preferential looking task. *Journal of Child Language*, *34*, 701-724.

Huettig, F., & McQueen, J.M. (2007). The tug of war between phonological, semantic, and shape information in language-mediated visual search. *Journal of Memory and Language, 54*, 460-482.

Hunter, M. A., & Ames, E.W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. In C. Rovee-Collier & L.P. Lipsitt (Eds.), *Advances in infancy research* (Vol 5, pp. 69-95). Norwood, NJ: Ablex.

Johnson, E.K. (2008). Infants use prosodically conditioned acoustic-phonetic cues to extract words from speech. *Journal of the Acoustical Society of America*, 123, EL144-EL148.

Johnson, E.K., Huettig, F., & McQueen, J.M. (2008). Conceptual attributes of heard words modulate 24-month-olds' attention to the visual scene. Poster presented at the 2008 Intl. Congress for the Study of Child Language, Edinburgh, Scotland.

Johnson, E.K., & Jusczyk, P.W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory & Language*, *44*, 1-20.

Johnson, E.K., & Seidl, A. (2008). Clause segmentation by 6-months-olds: A crosslinguistic perspective. *Infancy*, 13, 440-455.

Johnson, E.K., & Seidl, A. (2009). At 11 months, prosody still outweighs statistics. *Developmental Science*, 12, 131-141.

Johnson, E.K., Jusczyk, P.W., Cutler, A., & Norris, D. (2003). Lexical viability constraints on speech segmentation by infants without a lexicon. *Cognitive Psychology*, *46*, 65-97.

Johnson, E.K., Westrek, E., & Nazzi, T. (2005). Language familiarity affects voice discrimination by seven-month-olds. *Proceedings of the ISCA Workshop on Plasticity in Speech Perception (PSP2005) [CD-ROM]* (pp. 227-230). London.

Jusczyk, P. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.

Jusczyk, P.W., & Aslin, R.N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology, 29*, 1-23.

Jusczyk, P.W., Cutler, A., & Redanz, N. (1993). Preference for the predominant stress patterns of English words. *Child Development*, *64*, 675-687.

Kooijman, V., Johnson, E.K., & Cutler, A. (2008). Reflections on reflections of infant word recognition. In A. Friederici & G. Theirry (Eds.), *Trends in Language Acquisition Research: Vol. 5, Early Language Development: Bridging brain and behaviour* (91-114). Amsterdam: John Benjamins.

Lew-Williams, C., & Fernald, A. (2007). Young children learning Spanish made rapid use of the grammatical gender in spoken word recognition. *Psychological Science*, *18*, 193-198.

Mandel, D.R., Jusczyk, P.W., & Pisoni, D.B. (1995). Infants' recognition of the sound patterns of their own names. *Psychological Science*, *6*, 314-317.

Mattys, S.L., & Jusczyk. P.W., Luce. P.A., & Morgan, J.L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology, 38,* 465-494.

McMurray, B., & Aslin, R.N. (2004). Anticipatory eye movements reveal infants' auditory and visual categories. *Infancy*, *6*, 203-229.

McMurray, B., & Aslin, R.N. (2005). Infants are sensitive to within-category variation in speech perception. *Cognition, 95,* B15-B26.

McMurray, B., Spivey, M., & Aslin, R.N. (2000). The perception of consonants by adults and infants: Categorical or categorized? *The Working Papers in the Language Sciences at the University of Rochester*, *2*, 215-256.

Meek, J. (2002). Basic principles of optical imaging and application to the study of infant development. *Developmental Science*, *5*, 371-380.

Mehler, J., Jusczyk, P.W., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, *29*, 143-178.

Mirman, D., Dixon, J., & Magnuson, J. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory & Language*.

Nazzi, T., Jusczyk, P.W., & Johnson, E.K. (2000). Language discrimination by English-learning 5-month-olds: Effects of rhythm and familiarity. *Journal of Memory and Language*, *43*, 1-19.

Nazzi, T., Paterson, S., & Karmiloff-Smith, A. (2003). Word segmentation by infants with Williams syndrome. *Infancy*, *4*, 251-271.

Newman, R. S., Bernstein Ratner, N., Jusczyk, A.M., Jusczyk, P. W., & Dow, K.A. (2006). Infants early ability to segment the conversational speech signal predicts later language development: A retrospective analysis. *Developmental Psychology, 42*, 643-655.

Pinto, J.P., Fernald, A., McRoberts, G.W., & Cole, S. (1999). Reliability and validity in infant auditory preference procedures. In C. Rovee-Collier, L.P. Lipsitt, & H. Hayne (Eds.), *Advances in Infancy Research* (pp). Norwood, NJ: Ablex Publishing Co.

Saffran, J.R., Aslin, R.N., & Newport, E.L. (1996). Statistical learning by 8-month-olds infants. *Science, 274*, 1926-1928.

Santelmann, L., & Jusczyk, P.W. (1998). Sensitivity to discontinuous dependencies in language learners: evidence for processing limitations. *Cognition*, *69*, 105-134.

Seidl, A., Hollich, G., & Jusczyk, P.W. (2003). Early understanding of subject and object wh-questions. *Infancy*, *4*, 423-436.

Singh, L. (2008). Influences of high and low variability on infant word recognition. *Cognition*, 106, 833-870.

Soderstrom, M., & Morgan, J. (2007). Twenty-two-month-olds discriminate fluent from disfluent adult-directed speech. *Developmental Science*, *10*, 641-653.

Stager, C.L., & Werker, J.F. (1997). Infants listen for more phonetic detail in speech perception than in word learning tasks. *Nature*, *388*, 381-382.

Swingley, D. (2003). Phonetic detail in the developing lexicon. *Language and Speech, 46,* 265-294.

Swingley, D. (2005). 11-month-olds' knowledge of how familiar words sound. *Developmental Science, 8,* 432-443.

Swingley, D., & Aslin, R.N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological Science, 13,* 480-484.

Thiessen, E.D., Hill, E.A., & Saffran, J.R. (2005). Infant-directed speech facilitates word segmentation. *Infancy*, *7*, 53-71.

Thorpe, K., & Fernald, A. (2006). Knowing what a novel word is not: two-year-olds 'listen through' ambiguous adjectives in fluent speech. *Cognition*, *100*, 389-433.

Tincoff, R., & Jusczyk, P.W. (1999). Some beginnings of word comprehension in six-month-olds. *Psychological Science*, *10*, 172-175.

Van Alphen, P., de Bree, E., Gerrits, E., de Jong, J., Wilsenach, C. & Wijnen, F. (2004). Early language development in children with a genetic risk for dyslexia. *Dyslexia* 10, 265-288.

Van Heugten, M., & Shi, R. (in press). French-learning toddlers use gender information on determiners during word recognition. *Developmental Science*.

Vihman, M., Nakai, S., DePaolis, R.A., & Halle, P. (2004). The role of accentual pattern in early lexical representation. *Journal of Memory & Language*, *50*, 336-353.

Werker, J.F., Polka, L., & Pegg, J.E. (1997). The conditioned head turn procedures as a method for assessing infant speech perception. *Early Development and Parenting,* 6, 171-178.

Wilsenach, C. (2006). *Syntactic Processing in Developmental Dyslexia and in Specific Language Impairment: A study on the acquisition of the past participle construction in Dutch*. Doctoral Dissertation, Utrecht University.

Zamuner, T.S. (2006). Sensitivity to word-final phonotactics in 9- to 16-month-old infants. *Infancy, 10*, 77-95.

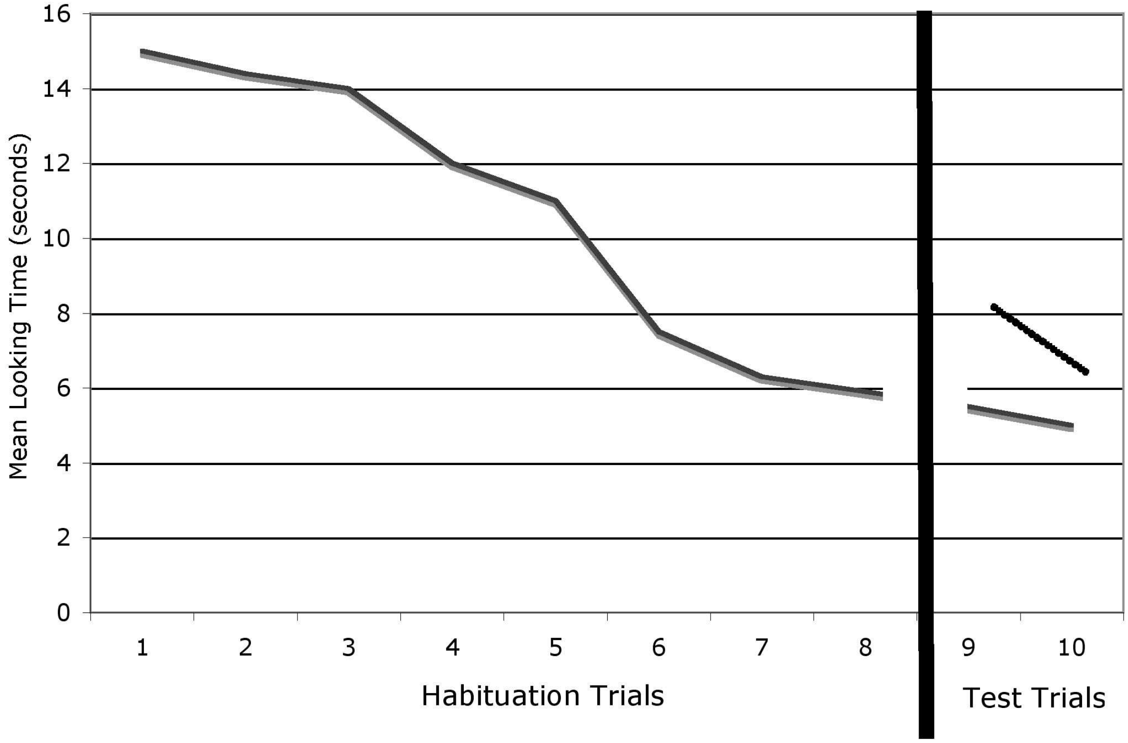

Figure 1: Trials 1 through 8 are habituation trials. Once looking time decrease by 50%, the test trials begin. In this idealized data set, the dashed line represents an outcome where the infant dishabituates, i.e. notices the difference between the habituation and test stimuli. The solid line represents the outcome where the infant fails to dishabituate, i.e. does not notice the difference between the habituation and test stimuli.

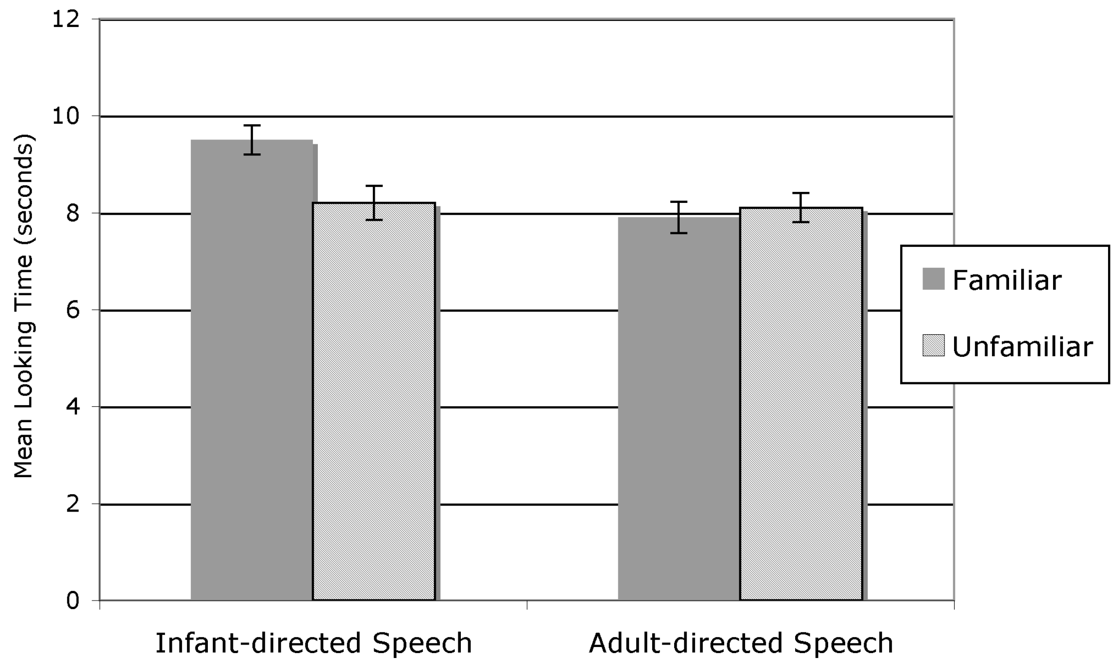

Figure 2: Looking times to test stimuli (familiar versus unfamiliar) broken down by familiarization condition (Infant- versus Adult-directed)

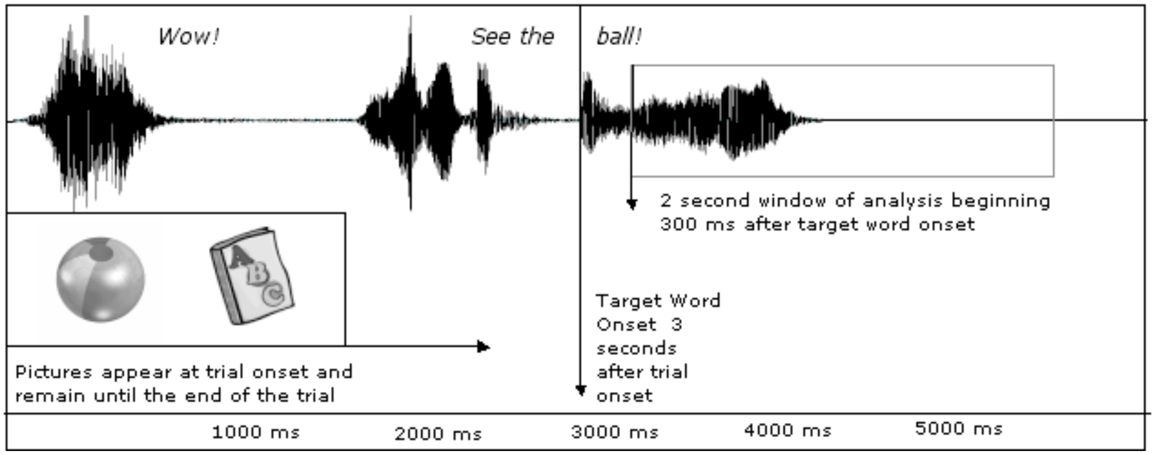

Figure 3: The timeline for a typical trial in a preferential looking study

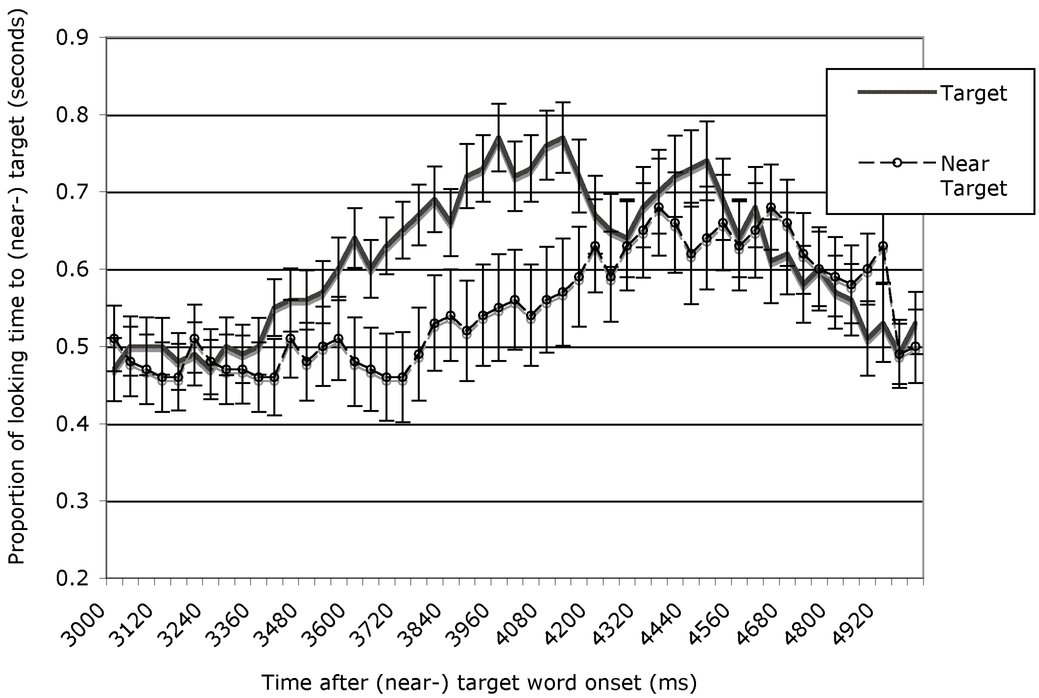

Figure 4: Timeline for fictitious study comparing infants recognition to correctly pronounced targets (e.g. *ball*) versus near targets (e.g. *pall* instead of *ball*)