The looking-while-listening procedure

Daniel Swingley

Department of Psychology, University of Pennsylvania, 3401 Walnut St. 302C, Philadelphia, PA 19104

Summary

The "looking-while-listening" or "language-guided-looking" procedure is used to closely examine children's interpretation of spoken language. The timecourse of children's eye movements toward pictures or scenes is evaluated while children hear sentences describing one of them. The procedure can be used in virtually any population, and works consistently in children from 14 months onward, with little modification necessary for testing older children or adults. Researchers have examined word recognition and sentence understanding to address a range of questions concerning language representation and performance.

Measuring performance in language acquisition

Linguists and developmental psychologists often focus on the *logical* problems of language acquisition: the discovery of grammar from apparently insufficient evidence, the learning of word meaning despite infinite possible false starts. But for parents observing their children's development, much of the fascination of language acquisition lies in the startling leaps children make from one day to the next as they reveal their thoughts: the words and expressions and sentences that suddenly rush onstage in all their glory. We sense the developing mind struggling to share its thoughts with us, and we ask, "Where on earth did that come from?" And for the most part, we don't know, or we don't know in any detail, because most of what we can determine of any child's linguistic knowledge is revealed by his or her speech—the behavior that surprised us in the first place.

Comprehension precedes production: young children who can understand multiword sentences may struggle to produce a two-word utterance. Children's corrections of their own attempts are usually improvements rather than random walks, and children can be creative as they strain to give voice to their ideas. For example, children making the transition from one-word utterances to two-word utterances often begin by using one word and one gesture to communicate two separate components of their intended meaning (Iverson & Goldin-Meadow, 2005). These observations show that when young children speak, the semantic target that they are aiming for may be clear in their minds even when the child's utterance is not understandable to others, or when the child fails to produce an utterance at all. It is commonly said that language development is effortless for children, but it's not true—children's difficulties are just different from ours. One stumbling block appears to be the inability to successfully assemble sentences out of components that children seem to *interpret* correctly, according to their language's grammar.

Precisely because comprehension precedes production, mapping out the developmental course of language acquisition requires evaluation of children's receptive knowledge of language. Researchers have risen to the task: we have ways to assess discrimination and categorization of speech sounds and sequences of sounds from birth, using habituation measures like high-amplitude sucking (Fennell, [cite-his-chapter]). We can test infants' recognition of speech sequences using contingent listening tasks like the Headturn Preference Procedure or response training tasks like Conditioned Headturn. And we can measure correlates of stimulus-driven brain activity using one of a few brain-imaging techniques (Kovelman, [cite-her-chapter]). These procedures work with infants, but for the most part, they do not depend on (or directly reveal) children's meaningful interpretation of language.

Experimental tests of language comprehension have been available for a long time: there are act-out tasks in which children manipulate toys or other objects under instruction to enact the statements of a researcher (or puppet). There is the truth-value judgment task, which asks children whether a puppet is right or wrong about some statement. These procedures certainly depend on children's interpretation of language. But they are extremely difficult to use in children younger than two to two and a half years old.

Eye movement tasks were developed to help fill the gap left by these other procedures. The first "preferential looking" studies examining infant language comprehension showed that children's gaze toward images or films was affected in sensible ways by simultaneously presented spoken language: if the child heard "Look at the doggie!" he or she tended to look at a picture of a dog more than a picture of some other familiar object (e.g., Golinkoff, Hirsh-Pasek, Cauley, & Gordon, 1987; Reznick, 1990; Thomas, Campos, Shucard, Ramsay, & Shucard, 1981). This response allows researchers to draw inferences about children's knowledge: if children increase their looking to a named image or film upon hearing it labeled, they show evidence of understanding the word. Likewise, syntactic interpretation may be probed: if children gaze at a scene that exemplifies a spoken sentence better than another scene (which might exemplify a syntactically different sentence with the same participants, for example), children provide evidence of knowing something about that syntactic structure.

This procedure has been tremendously influential in tracing the earliest development of children's knowledge of words and syntactic structures, as outlined by Piotroski and by Hirsh-Pasek (both this volume). Over the past decade and a half, preferential looking procedures have been adapted for studying not only children's knowledge of language, but also the details of children's *performance* in interpreting language on-line. It is this work that is the focus of the present chapter.

The notion of "performance" is often raised in the literature on language acquisition, traditionally as part of an effort to explain why a child didn't succeed in a task, or botched a grammatical structure, in spite of the researcher's belief that the child had already acquired the relevant grammatical knowledge. The distinction between the child's capabilities and the child's behavior is a necessary one if, like most cognitive scientists, one wishes to maintain a distinction between representation (taken to be relatively static knowledge) and process (taken to be the the implementation of this knowlege in cognition under particular circumstances). But in language acquisition, performance has not been the focus of sustained empirical attention to the extent that representation ("competence") has. This state of affairs emerged in part due to a vicious circle: the more "performance" is (often justifiably) derided as an unmotivated "fudge factor," the less it has appeared a compelling research area to work on, and one consequence is performance-based descriptions of behavior without adequate content.

This has begun to change, partly as a result of research using "looking-while-listening," "language-guided looking", or "eyetracking" methods with infants and young children. In developmental psycholinguistics, these terms all mean the same thing. Where they diverge from the traditional "preferential looking" methods is the use of characteristic features of eye movement responses to address questions about psychological processes. In particular, looking-while-listening experiments focus on the *timing* of children's responses. Response times have been an essential component of cognitive psychology from the beginning, starting with Donders and Cattell in the late 19th century and carrying on to this day, but they have had a relatively limited impact on developmental psychology simply because of the difficulty of persuading children to make an overt manual response quickly and accurately. However, very young children make rapid eye movements as they seek information, and, it turns out, they do so in response to language.

The development of looking-while-listening

This part of the story begins around 1990 in Anne Fernald's lab at Stanford. She wanted to know if the infant-directed prosodic register would make words easier for young children to understand. On each of a series of 12 trials, children at 12, 15, and 18 months of age were presented with pairs of pictures, one on each of two screens. A few seconds later, a pre-recorded speech stimulus was played, either in the infant-directed or adult-directed register, naming one of the two objects. A film of the child's face was recorded on videotape. The whole apparatus was decidedly low-tech by today's standards: the pictures were beamed onto rear-projection screens by yoked slide projectors, and the speech was presented using a reel-to-reel tape player triggered by a research assistant listening to a recording of a metronome. What distinguished Fernald's implementation, though, was her insistence on frame-by-frame, off-line coding of children's eye movements.

At the time, most labs using preferential looking methods recorded looking behavior on-line, using a button box. This is a speedy way to go—you have the child's results before you say goodbye to the mom—but it inevitably introduces error because of delays and variability in responding to the child. Off-line coding removes the response latency component to scoring and allows researchers to draw up precise standards for what counts as a fixation, when a fixation should be considered to begin and end, what to do when the child blinks, and other details that were hashed out in long late-night discussions in the Stanford lab. Originally the decision to code off-line was not based on the anticipation of measuring response latencies; it was simply a matter of getting the best possible record of behavior.

Fernald, McRoberts, and Herrera (in Fernald, McRoberts, & Swingley, 2001) found that young children indeed recognized words more reliably when spoken in the infant-directed register. They also found that target fixation was not uniform over the six second test period (starting from the offset of the sentence): among

18 month olds, performance was better from 0–2 sec after sentence offset than 2–4 sec, and worst of all at 4–6 sec. As these results were emerging, John Pinto and I joined Anne Fernald's lab. Pinto developed software that could take the lab's detailed records of eye movements and produce a matrix of gaze locations over time, showing for each trial, and at each moment in time, whether the child was fixating the target, distracter, or neither. We began exploring other possible windows of analysis: two-second windows, 1-second windows, windows starting mid-word, and so on. At some point in this exploration process I decided to compute every possible 1-second window, starting from the beginning of the trial and going to the end. A plot of the means of each overlapping window revealed a smooth curve starting at 50% and rapidly rising to about 80% before falling off. I was quite proud of this until I showed it to Pinto, who asked why I was taking a moving average rather than simply plotting the raw averages for each time-slice. Skipping the 1-second smoothing operation led to the first of our plots of the time-course of word recognition in children.

Broadly speaking, time-course analyses of young children's word recognition have yielded two main lines of research, one examining the cognitive processes involved in parsing sentences, and another examining individual differences among children. I will illustrate these research programs using selected results.

Both lines emerged from an early study using language-guided looking (Fernald, Pinto, Swingley, Weinberg, & McRoberts, 1998). On each of 8 trials, children heard sentences like "Where's the ball?" while viewing a ball on one computer screen and a shoe on the other. Children's eye movements were coded from videotape. Results were reported in three ways, in keeping with much subsequent work. First, children's proportion of fixation to the target during a 2-second window, from target-word onset, was computed. On each trial, the total number of 100-ms intervals (from 0 ms to 2000 ms) on which children were coded as fixating the target was divided by the count of target- and distracter-fixated intervals. Among

15-, 18-, and 24-month-olds mean proportions were consistently greater than the 50% that would be expected if children did not respond to the words (each picture served as target and distracter equally often in yoked pairs, so picture preferences could not lead to overall above-50% performance). Mean target fixation increased substantially with age.

Second, children's *response latencies* were computed, defined as the amount of time, from the onset of the target word, it took for children to initiate a refixation away from the distracter picture to the target, for all trials on which children had been fixating the distracter picture when the target word began. Mean response latencies declined substantially as children grew older—from a mean of 995 ms at 15 months to 679 ms at 24 months.

Third, children's proportion of target fixation over time was displayed graphically in *onset-contingent timecourse plots* (e.g., Figure 1). In such graphs, time is given on the *x*-axis. Trials are grouped according to whether the child happened to be fixating the target, or the distracter, when the target word began. This is useful because children are in different states in each case. A child looking at the target is hearing it named as the sentence unfolds; the sentence confirms that the child's focus of attention matches the utterance. Recognition is inferred when children continue fixating the target. A child looking at the distracter hears the spoken target word, and the meaning that the word evokes is inconsistent with the child's focus of attention, which is expected to provoke a refixation (Swingley & Fernald, 2002). Rather than plotting target fixation on the *y*-axis, we plot, for each unit of time, the proportion of trials on which children are, at that moment, fixating a different picture than the one they were looking at when the target word began. Thus, good performance is shown by "target-initial" trials remaining flat near zero, and "distracter-initial" trials rapidly departing from zero toward one. The advantage of plotting this way, rather than showing percent-to-target for each type of trial, is that one can see clearly when the child begins to show evidence of understanding, namely when the lines diverge. The divergence in the lines happens when defections from the target are outnumbered by rejections of the distracter picture.

For example, Figure 1 plots data from Experiment 2 in Swingley (2009), in which 14–22-month-olds responded to correctly-pronounced words (black lines on the plot) and mispronounced words (red lines). When hearing correct pronunciations, such as *book* children's refixations from the distracter (shown with a solid line) began to exceed refixations from the target (dashed line) at around 750 ms. The separation between the distracter-initial and target-intial lines is smaller than, and begins later than, what we would expect from older children or adults, but it is clear that children's responses are, on the whole, responsive to the match between the word and the fixated picture. When children heard mispronunciations, such as *dook*, children shifted away from the target and distracter pictures equally often until about 1450 ms, and the relatively small separation between target-initial and distracter-initial means reveals children's difficulty in understanding the mispronounced word.

The Fernald et al. (1998) paper led directly to a series of studies examining detailed aspects of the time-course of word recognition. In several cases these studies compared children and adults directly, and found qualitatively (and often quantitatively) similar patterns. For example, Swingley, Fernald, & Pinto (1999) compared how 24-month-olds and adults responded when hearing (e.g.) *doggie* and looking at a doll, versus hearing *doggie* and looking at a tree. At both ages, participants rejected the distracter faster when its initial sounds did not match the spoken word (e.g., the /tr/ of "tree") than when they did (e.g., the /dɔ/ of "doll"). The temporal delay in rejecting the onset-overlapping "doggie" and "doll" distracters was, at both ages, comparable to the amount of time the two spoken words overlapped, as assessed using adults' explicit judgments of fragments of the words. These results showed that children, like adults, interpret speech incrementally, updating their understanding as the words unfolded. Support for the same idea in younger children was shown by Fernald, Swingley, & Pinto (2001), and by Swingley (2009). In the latter study, children from 14 to 22 months of age responded to stimuli that were "mispronounced" at onset (e.g., *cup* as "gup") or offset (e.g., *cup* as "cub"). Relative to the onset of the target words, children's target fixation was disrupted immediately when onsets were altered, and disrupted a few hundred milliseconds later when offsets were altered, in keeping with the acoustic timing of the deviations themselves.

The principle of incremental interpretation holds for other areas of sentence parsing as well, although it is clear that children's proficiency in using linguistic information depends on mastery of the words and constructions they are hearing. By about 36 months, native Spanish learners use grammatical gender encoded in articles to help identify the following noun (Lew-Williams & Fernald, 2007); French learners have been shown to do the same at 25 months (van Heugten & Shi, 2009). On the other hand, at 30 months, English learners do not consistently use prenominal adjectives to constrain reference (Fernald, Thorpe, & Marchman, 2010). For example, confronted with a blue car and a red house, 30 month olds had only mixed success in using the adjective in sentences like "Which one's the blue car?" to guide their attention away from the red distracter picture. Given that all the children tested were capable of using *blue* and *red* appropriately in their own speech, this result points to the fact that expert language understanding is a skill in itself, over and above knowledge of word meanings and acquisition of grammatical rules.

This notion of language processing as a skill fits well with another major line of research using the looking-while-listening procedure, where quantitative details of children's responses are used to establish individual differences in speech processing. The Fernald et al. (1998) paper mentioned above explored differences in mean response latency among children of different ages; more recent work has tested whether measurement of speed and accuracy in word recognition are correlated with other cognitive assessments both contemporaneously and predictively. For example, Marchman and Fernald (2008) measured 25-month-olds' response latency to the nouns in sentences like "Where's the doggie?" or "Where's the nice cow?". Response latency made significant unique contributions to prediction of working memory scores on a standarized test given at 8 years; response latency and vocabulary size together predicted high scores on expressive language. More recent studies by the same authors reveal connections between maternal child-directed speech and children's speed of word recognition using the looking-while-listening task (Hurtado, Marchman, & Fernald, 2008). In a low-SES sample of Spanish-speaking families in California, children of more talkative mothers (assessed at 18 months) recognized words more quickly at 24 months than children whose mothers talked to them less; further, response latency at 24 months was significantly correlated with growth in vocabulary over the prior 6 months.

Both the individual difference studies and the incremental-processing studies take advantage of unique properties of the language-guided looking procedure by evaluating not only children's overall performance in linking words to referents, but also the details of the timing with which these connections are made. The procedure is also used frequently in answering questions about static representational knowledge, and in these cases the temporal dynamics of children's reponses provide an additional dependent measure but are not the focus of the research. For example, following Swingley & Aslin (2000), several studies have evaluated children's looking patterns upon hearing canonical pronunciations of words and phonologically deviant pronunciations. Children typically look to target pictures more rapidly when the words are correctly pronounced than mispronounced, just as adults do; whereas phonetic changes that are not phonologically distinctive in the language do not hinder recognition (e.g., Quam & Swingley, 2010). In these studies, the response latency measures are usually redundant with other measures of target looking (such as proportion of target fixation over a short analysis window).

Implementation

Apparatus

The procedure is not technically demanding to implement. Considering the apparatus, the most common setup is built around a "booth", a 3-sided enclosure about 1.5 m on a side. In essence, the booth is a display surrounded by visually featureless space: there is nothing to see but what's on the screen. The display itself is usually a very large video monitor (42-inch diagonal is common) or a projection system fed by a computer. Projectors allow for larger pictures and a greater amount of space between pictures. Using large, widely-spaced pictures was initially motivated by desire to maximize coding accuracy on grainy, dimly-lit videotape, and also by the reasonable supposition that larger images would capture infants' attention better. However, it is not certain that this is necessary. We have been successful in testing children ranging from < 12 months to 30 months using a single 20-inch computer monitor. If the child is only 50-60 cm from the screen, the visual angle separating pictures placed near the screen's left and right edges is adequate for judging left and right fixations, particularly using modern cameras that do a good job recording in dim light. Whether giant screens impress infants enough to lower attrition rates is an open question.

Stimulus presentation can be controlled by software designed for experiments (such as Psyscope: http://psy.ck.sissa.it/psy_cmu_edu/index.html), animations (such as Director: http://www.adobe.com/products/director), or movies (such as Quicktime: http://www.apple.com/quicktime). Software programs vary in their temporal accuracy. One cannot assume that if the program was told to play sound *x* 2000 ms after displaying pictures *y*, that this actually happens consistently. In any case, even with a consistently-presented stimulus, one still needs a way to line up the experimental events (like the onset of the spoken target word on each trial) and infant events (like an eye movement shift). Different labs have handled this in different ways. One option is to use picture-in-picture video technology to place a recording of the visual display in the corner of the recording of the child's face, so that coders can link the eye-movement time stream with the trial onset times. (In this case coders are not blind to the target picture's side of presentation, however.) Another option, which we use, is to play one audio channel of the stereo stimulus signal to the child, and embed target-word-aligned tones in the other channel. A custom-built device detects these tones, and embeds a distinctive visual signal in the videorecording of the infant. Other tones (and other visual signals) indicate other trial events. Coders can then note the timing of all infant and stimulus events in the same way, by looking at the recorded video signal.

We audiorecord each session in two ways at once: one channel records the sequence of timing beeps, and another channel records input from a microphone that captures sound in the booth. The latter permits analysis of children's vocalizations (and parents' too, if they speak), and allows confirmation that the auditory stimuli were presented as intended.

Visual and auditory stimuli

Naturally, testing language understanding using eye movements to referents of sentences requires that the sentences have picturable referents. The two main concerns are the *recognizability* of the images, and their relative *salience*. We generally use photographs rather than drawings. Some objects are quite difficult to use as referents because their forms vary across instances in children's experience. For example, we have never used "phone" (toy phone? cell phone?) or "hat" (baseball? cowboy? wool with pom-pom?). Another vexing issue is children's tendency to find certain objects much more attractive to look at than others, particularly in children under 24 months of age. Animate objects are more captivating than inanimates, so even if one searches out the most drab, mangy dog, many children will gaze longingly at it. We often try to pair animates together to offset such effects.

For both of these issues, pilot testing is the only way to be sure.

Auditory stimuli should be recorded and selected with care, particularly for studies in which phonetic properties are central. Not all talkers are up to the task, and practice is usually necessary even for a speaker with excellent vocal control. When recording stimuli for an Experiment 1, it can be useful to have the speaker read materials for potential Experiments 2 and 3 as well, because acoustic characteristics of different recording sessions vary (and our speakers have a tendency to graduate and move away). If similar stimuli are to be compared directly (e.g., a word and a deviant pronunciation of that word) these stimuli should be listed next to one another in the recording script, to mitigate effects of drift on the part of the speaker. Multiple tokens should be recorded and the best and best-matching instances selected. This process is tedious and time-consuming, but a study can only be as good as its materials.

Procedure

Once parent and child and any assorted siblings arrive in the lab, it is up to the researcher to determine when it is time to stop playing and start moving into the room with the booth. As a rule, siblings should not be in the testing room with the toddler during the procedure. Toddlers clinging to lab toys should be separated from them if possible, because toys may distract the child, and because children have an uncanny knack for using toys to obscure their eyes from the camera's view. In most cases the child can sit on the parent's lap, though if toddlers insist on sitting in the chair alone, with the mother crouching nearby, this works too, or at least works better than a tantrum.

Parents need to be prevented from biasing their children's looking. We tell parents to look downward (we can see the parent's face clearly on camera and, if necessary, pause the study between trials if we detect any peeking). A visor or opaque sunglasses may also be used. We ask parents not to speak or point to the screen, but allow that if the child becomes restless, it is okay to say, between stimulus sentences, neutral reassurances like “I’m right here” or “look up at the pictures”. If children begin struggling or crying, we pause the procedure; sometimes a session can be rescued with a brief hiatus, or a short play period. The experimental session itself only takes about five minutes.

Stimulus orders

The two main concerns in setting up trial orders are maximizing the amount of looking data each child is likely to provide, and preventing nuisance effects from interfering with interpretation of the results. Speaking very generally, most 14 to 20 month olds can participate through 30 or so trials, and older children can stay on task for 40 or perhaps even more. Increasing variety in the sentences and the pictures increases the number of trials children make it through and reduces attrition. Many researchers include filler trials to provide this variety when the experimental questions require repetitive test trials; another option is to have occasional filler clips that are not trials per se, showing interesting pictures or animations.

Once the number of trials and stimulus items has been sorted out, the positioning of the pictures (left, right) and the sequential orders need to be determined. Rather than randomizing trial orders, most researchers carefully construct 4 or 8 orders and assign children to them randomly. If pictures are repeated at all (as in most studies) they are usually yoked: for example, a dog and baby might be paired, and on 2 trials the dog is the named picture and on 2 trials the baby is. This way a simple preference for (e.g.) the dog over the baby cannot lead to above chance performance on dog/baby trials taken together. Other properties that are usually maintained if possible include balancing target side (left, right); restricting sequences of target side to a maxiumum of two or three; placing each item equally often in the first and second half of the experiment; and avoiding adjacent-trial repetition of pictures or words, among others. Putting together the first order of a study is like solving a sudoku puzzle. But other orders can be constructed from the first order, e.g. by inverting whichever factor is most theoretically important, and by rotating through the items (replacing dog/baby with duck/car, duck/car with shoe/ball, . . .).

Coding

Most of the time, it is easy to tell when a child is looking to the left or to the right. What makes coding difficult is the need to establish precisely when a left or right fixation begins and ends. Research assistants thrown into the coding task without adequate apprenticeship will produce results that correlate fairly well but that do not agree on the details. A lab using the procedure must develop consistent, verbalizable, and complete standards for what counts as fixating a picture and what does not in a given video frame. In general, these standards do not depend on perfecting the ability to determine the horizontal angle of a child's fixation; rather, they depend on learning the dynamics of eye movements over time. Children looking at an object picture typically keep their eyes in a single position for several video frames. A shift to the other picture usually begins as 2–3 frames in which the eyes move with increasing angular change (acceleration) away from this single position. On the first frame of a shift the eyes are usually still oriented to the picture, in terms of what light is hitting the fovea, but the dynamics of the movement reveal that a shift is underway. Two well-trained coders can agree on the frame at which a look begins and ends to within zero or one video frame for most looks.

Data analysis

As described above, the first study using looking-while-listening with toddlers reported results in terms of proportions of target fixation, response latencies (RT), and timecourse plots. Since that time, refinements of the proportional and RT analyses have emerged, as well as other measures.

In most studies the real statistical workhorse is the proportion-to-target analysis. Timecourse is implicated in these analyses in the selection of the window of analysis over which the proportion is computed. In our initial studies a two-second window, starting from the onset of the target word, was used. Most studies now exclude a few hundred ms from the beginning of this window on the grounds that there are neurophysiological and cognitive limits on how quickly an eye movement could possibly be generated in response to the speech signal; even among children who perform very well, the first 1/3 of a second or so shows little contingency between stimulus and response. It is important to recognize, though, that the commonplace 367–2000 ms window is *not* an "optimal" window for revealing target recognition. It extends too early. Children initially fixating the distracter take, on average, several hundred milliseconds to decide to shift. For children between 14 and 24 months, an optimal window would extend from approximately 1200 to 2200 ms, depending on the materials. Why use the earlier window, then? Because it is in the early decision making that much of the *variability* in eye movement responses can be detected. Choosing an "optimal" window for the average 18 month old, say, could obscure the difference between toddlers who respond quickly and accurately and those who are accurate in their asymptotic interpretation but need more time. Likewise, children tend to respond faster to correct pronunciations of words than mispronunciations, but performance later in the trial can be similar between conditions.

When responses to particular items are of interest, and not just overall performance on a yoked pair of items, counterbalancing alone is not sufficient protection against biases due to picture preference. For example, a child might be presented with a car and a ball and hear "car" and "ball" on separate trials. 75% car-looking upon hearing "car," and 50% ball-looking upon hearing "ball," might not mean that the child understands only the word "car." The child could simply prefer to look at cars. In principle, 50% ball-looking might reveal understanding of "ball," if hearing the word pulled the child away from his favorite picture.

This problem is usually managed by comparing looking after hearing the word against looking before hearing the word. If children enjoy looking at cars, this may be revealed to some degree by the initial portion of the trial. Subtracting target-looking proportions before the test window from proportions during the test window is a way to adjust for these picture preferences. This can be done either trial by trial, or pair by pair (i.e., by computing a "preference score" for each picture based on all of the trials in which that picture appears with its yoked partner). This procedure is not without hazard, however. Imagine an adult who likes looking at cars but otherwise fixates where he is instructed. Given 100% car fixation before word onset and 100% fixation afterward (and thus a difference score of zero) one might conclude that he shows no evidence of "car" recognition. The problem is that in good performers whose test-window looking is not affected by picture preferences, subtracting pre-test looking amounts to adding a random number. These considerations suggest that the value of doing this sort of correction depends upon the degree to which looking before and after the word is correlated.

Reponse latency (RT) is a standard measure of cognitive performance, as mentioned above. In the looking-while-listening task, RTs are typically computed only for trials on which children are fixating the distracter when the target word begins (e.g., Fernald, Swingley, & Pinto, 2001). And of course, a child only produces a reaction time when she reacts. A necessary consequence is that analyses of RT are based on fewer trials than analyses of fixation proportion. This problem can be mitigated to some degree by maximizing the number of trials in the session, but it is not unexpected when subtle effects are found in analyses of fixation proportion but not RT.

Other response measures have been developed, sometimes to evaluate specific details of performance. Swingley (2009) computed the likelihood of shifts in fixation from target to distracter and from distracter to target, not only for first shifts, but for all refixations throughout the test window. This analysis is one way to distinguish cases in which children shift a great deal but just as often to the distracter as to the target, from cases in which children shift very little—a distinction that cannot be made from the usual proportional analyses. Plunkett (e.g., Mani & Plunkett, 2007) has examined "longest look duration", which, given a relatively long window of analysis, abstracts away from the response latency component of the fixation task. A child's longest fixation might occur early or late in the analysis window, and will count the same either way. In principle, this might provide a way to equalize differences among younger and older children who tend to respond more quickly.

Future directions

The literature on word recognition and sentence processing in adults is largely composed of chronometric studies evaluating listeners' interpretation of language over time (e.g., Dahan, 2010; Tanenhaus, 2007). Many of these studies in the past 15 years have been done using eyetracking techniques that are, in all important respects, the same as the looking-while-listening procedure. Although most studies of adults, and increasingly also studies of children, make use of automatic eyetracking systems, the logic of the experiments and the nature of the listener's response is the same whether a machine or a human is coding gaze patterns. The fact that the language-guided looking task can be used effectively from the second year into adulthood reflects the naturalness and automaticity of the response, and points to the method's potential in measuring and explaining language understanding performance over the full span of development.

Further reading and resources

The best detailed introduction to the "nuts and bolts" of the looking-while-listening procedure is Fernald, Zangl, Portillo, and Marchman (2008). Discussion of the mental processes that are revealed in the task, and defense of the argument that listeners' interpretations are not artificially constrained by the response set, are provided in Dahan and Tanenhaus (2005) and Swingley and Fernald (2002). Overviews of research using the task include Swingley (2008) on infant development and phonology, and Trueswell (2008) on syntactic development. Information about software mentioned in the chapter can be found online:
for Psyscope, http://psy.ck.sissa.it/psy_cmu_edu/index.html;
for Adobe Director, http://www.adobe.com/products/director;
for Quicktime, http://www.apple.com/quicktime.

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

Dahan, D., & Tanenhaus, M. K. (2005). Looking at the rope when looking for the snake: Conceptually mediated eye movements during spoken-word recognition. Psychonomic Bulletin & Review, 12, 453-459.

Swingley, D., & Fernald, A. (2002). Recognition of words referring to present and absent objects by 24-month-olds. Journal of Memory and Language, 46, 39-56.

Swingley, D. (2008). The roots of the early vocabulary in infants' learning from speech. Current Directions in Psychological Science, 17, 308-312.

Trueswell, J. C. (2008). Using eye movements as a developmental measure within psycholinguistics. In I. Sekerina, E. M. Fernández, & H. Clahsen (Eds.), Developmental psycholinguistics: On-line methods in children's language processing (p. 73-96). Amsterdam: Benjamins.

Key terms

eyetracking, language-guided-looking, looking-while-listening: These terms all refer to procedures in which children are shown scenes or images, and hear language referring to one of the scenes or images. The details of children's fixations to the images, and in particular the timing of children's responses, are evaluated as measures of language understanding. *Eyetracking* is sometimes used to refer specifically to the use of such measures with an automated eyetracking computer.

onset-contingent timecourse plot: A graph with time on the x-axis, usually starting from the onset of the linguistic stimulus of interest, and a summary measure of children's fixations on the y-axis. Test trials are divided into those on which children initially gazed at the target image or distracter image. For distracter-initial trials the y-axis reflects moment-by-moment target fixation. For target-initial trials the y-axis reflects moment-by-moment distracter fixation. Thus the vertical separation of these lines, where the distracter-initial line rises above the target-initial line, shows the moment when children begin showing evidence of word recognition.

Mini bio

Daniel Swingley studied psychology and cognitive science at Brown, Oxford, and Stanford before taking research positions at Rochester and the MPI for Psycholinguistics. He has been at the University of Pennsylvania since 2003.

References

Dahan, D. (2010). The time course of interpretation in speech comprehension. *Current Directions in Psychological Science*, *19*, 121-126.

Fernald, A., McRoberts, G. W., & Swingley, D. (2001). Infants' developing competence in recognizing and understanding words in fluent speech. In J. Weissenborn & B. Hoehle (Eds.), *Approaches to bootstrapping: Phonological, lexical, syntactic, and neurophysiological aspects of early language acquisition* (Vol. I, p. 97-123). Amsterdam: Benjamins.

Fernald, A., Pinto, J. P., Swingley, D., Weinberg, A., & McRoberts, G. W. (1998). Rapid gains in speed of verbal processing by infants in the second year. *Psychological Science*, *9*, 228-231.

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

Fernald, A., Thorpe, K., & Marchman, V. A. (2010). Blue car, red car: developing efficiency in online interpretation of adjective-noun phrases. *Cognitive Psychology*, *60*, 190-217.

Golinkoff, R. M., Hirsh-Pasek, K., Cauley, K. M., & Gordon, L. (1987). The eyes have it: Lexical and syntactic comprehension in a new paradigm. *Journal of Child Language*, *14*, 23-45.

Heugten, M. van, & Shi, R. (2009). French-learning toddlers use gender information on determiners during word recognition. *Developmental Science*, *12*, 419-425.

Hurtado, N., Marchman, V. A., & Fernald, A. (2008). Does input influence uptake? links between maternal talk, processing speed and vocabulary size in Spanish-learning children. *Developmental Science*, *11*, F31-F39.

Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological Science*, *16*, 367-371.

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

Mani, N., & Plunkett, K. (2007). Phonological specificity of consonants and vowels in early lexical representations. *Journal of Memory and Language*, *57*, 252-272.

Marchman, V. A., & Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental Science*, *11*, F9-F16.

Quam, C., & Swingley, D. (2010). Phonological knowledge guides two-year-olds' and adults' interpretation of salient pitch contours in word learning. *Journal of Memory and Language*, *52*, 135-150.

Reznick, J. S. (1990). Visual preference as a test of infant word comprehension. *Applied Psycholinguistics*, *11*, 145-166.

Swingley, D. (2009). Onsets and codas in 1.5-year-olds' word recognition. *Journal of Memory and Language*, *60*, 252-269.

Swingley, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition*, *76*, 147-166.

Swingley, D., & Fernald, A. (2002). Recognition of words referring to present and absent objects by 24-month-olds. *Journal of Memory and Language*, *46*, 39-56.

Swingley, D., Pinto, J. P., & Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognition*, *71*, 73-108.

Tanenhaus, M. K. (2007). Spoken language comprehension: insights from eye movements. In G. Gaskell (Ed.), *Oxford handbook of psycholinguistics* (p. 309-326). Oxford: Oxford University Press.

Thomas, D. G., Campos, J. J., Shucard, D. W., Ramsay, D. S., & Shucard, J. (1981). Semantic comprehension in infancy: A signal detection analysis. *Child Development*, *52*, 798-803.

Acknowledgements

Thanks are due to Anne Fernald, Gerry McRoberts, and John Pinto, the team I joined at Stanford in the early 1990s when looking methods were already in development. The present paper was supported by NIH grant R01-HD049681.

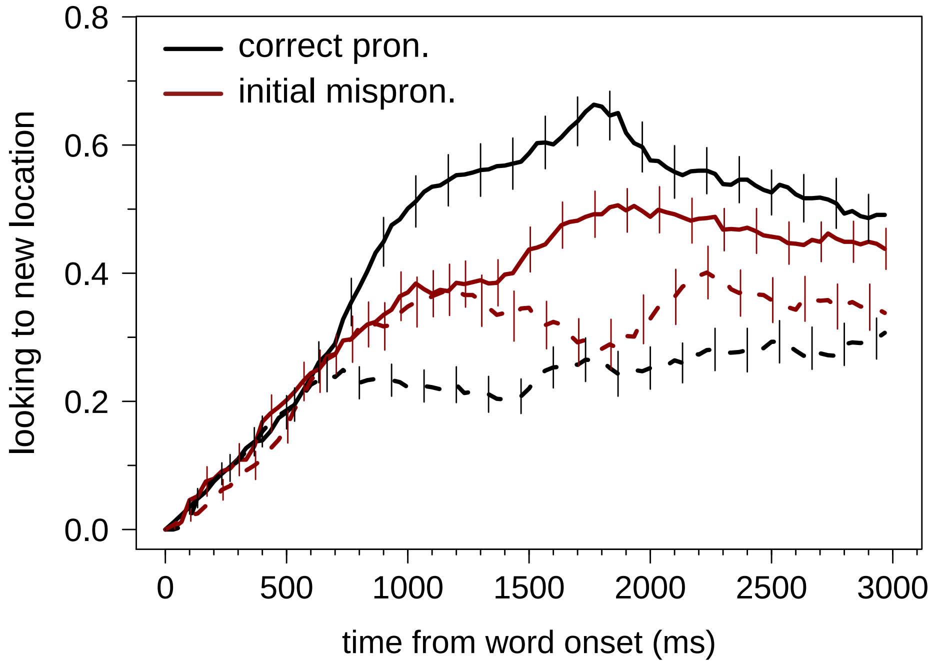

Figure 1: One-year-olds' fixation to named target pictures hearing ordinary pronunciations of words and mispronunciations of words. Time from the acoustic onset of the target word is plotted on the x-axis. The y-axis reveals the proportion of children (n=60) who, at each moment, were no longer fixating the image they had been fixating when the target word began. Solid lines indicate trials on which children happened to be fixating the distracter when the target-word began; dashed lines indicate trials on which children were already fixating the target. Black lines show responses on correct pronunciation trials, red lines mispronunciation trials. Vertical bars show standard errors of the mean computed over children.