SMART-T: A system for novel fully automated anticipatory eye-tracking paradigms

Mohinish Shukla
Brain & Cognitive Sciences
University of Rochester

Johnny Wen
Center for Visual Science
University of Rochester

Katherine S. White
Brain & Cognitive Sciences
University of Rochester,
and Dept of Psychology
University of Waterloo

Richard N. Aslin
Brain & Cognitive Sciences,
Center for Visual Science,
and Rochester Center for Brain Imaging
University of Rochester

Almost-final draft version of December 27, 2010, accepted for publication in *Behavior Research Methods*.

Anticipatory eye-movements (AEMs) are a natural and implicit measure of cognitive processing, and have been successfully used to document important cognitive capacities like learning, categorization and generalization, especially in infancy (McMurray & Aslin, 2004). Here, we describe an improved AEM paradigm to automatically assess on-line learning on a trial-by-trial basis, by analyzing eye-gaze data in each inter-trial interval of a training phase. Different measures of learning can be evaluated simultaneously. We describe the implementation of a system for designing and running a variety of such AEM paradigms. Additionally, this system is capable of a wider variety of gaze-contingent paradigms, as well as implementations of standard non-contingent paradigms. Our system, Smart-T (System for Monitoring Anticipations in Real Time with the Tobii), is a set of Matlab scripts with a graphical front-end, written using the Psychtoolbox. The system gathers eye gaze data using the commercially available Tobii eye trackers via a Matlab module, *Talk2Tobii*. We report a pilot study showing that Smart-T can detect 6-month-old infants' learning of simple predictive patterns involving the disappearance and re-appearance of multimodal stimuli.

Introduction

Eye movements to objects and locations in the visual environment are driven not only by low-level stimulus features, but also by ongoing cognitive processes. Eye movements can differ for the same visual display according to the viewer's task (Yarbus, 1967; Altmann & Kamide, 1999) and occur even in the absence of a stimulus: gaze can be directed to an empty location in space where an object has appeared in the past or to the predicted location in which a stimulus will appear in the future. Eye movements constitute a rich source of data; there are a number of potentially informative measures of looking behavior, which may map onto different aspects of processing. The analysis of eye movements is increasingly being used across a broad array of domains, from basic research on cognitive processing, to the diagnosis of pathology, to practical issues in webpage design.

In adults, eye movements have been used to study a wide variety of cognitive processes. Scanning patterns show how information is extracted from a visual scene (Loftus & Mackworth, 1978) and eye movement patterns have been used to analyze cognitive processes into their components in the laboratory and in natural tasks (e.g., Hayhoe & Ballard, 2005). Eye movement data have been particularly informative in the study of language processing. For spoken language, for example, the visual world paradigm (e.g., Cooper, 1974; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995) has capitalized on the fact that eye movements to objects in a visual scene are closely time-locked to the language input. In this paradigm, participants not only look at referents tied to the current linguistic input, but also make anticipatory

Acknowledgments: We are very grateful to Fani Deligianni for developing *Talk2Tobii* and sharing it with the scientific community. We thank Alyssa Thatcher, Holly Palmeri and Meg Schlichting for aiding in system testing, members of the Rochester Babylab for infant recruitment, and Krista Byers-Heinlein for help in brainstorming the Smart-T name. Funding for this project comes in part from the James F. McDonnell Foundation and NIH grant T32 MH19942. Smart-T can be downloaded from http://smartt.wikidot.com/

eye movements (AEMS) to potential future referents, revealing the highly predictive nature of the language processing system (e.g., Altmann & Kamide, 1999). Thus, predictive eye movements can reveal the many types of real-world and linguistic knowledge that interact during language processing. Eye movements have also been used in practical domains, such as advertising and webpage design, and to study the effects of expertise and distraction on driving. For example, predictable webpage layouts lead to more rapid and accurate eye movements to target content (Hornof & Halverson, 2002).

One significant advantage of employing looking behavior as a dependent measure is that, because eye movements can be launched automatically and without conscious awareness, they can be used to study individuals for whom more standard experimental tasks might be ill suited. Aphasic patients who have difficulty comprehending spoken language have been successfully tested on their language processing using eye-tracking paradigms that reduce task demands (Yee, Blumstein, & Sedivy, 2008). Similarly, the visual world paradigm has recently been used to assess language processing in autism, reducing the meta-linguistic demands that pose problems for this population (Brock, Norbury, Einav, & Nation, 2008). Because looking behavior is tied to a host of motor, perceptual and cognitive systems, it is also potentially useful in the diagnosis of pathology. For example, when visual targets alternate between fixed locations with fixed timing, autistic individuals show fewer anticipatory looks to the future location of the stimulus than normal controls. In contrast, in tasks in which it is necessary to inhibit saccades (e.g., delayed response task), both individuals with ADHD and schizophrenics show an increased number of intrusive anticipatory saccades (i.e., saccades to the target location during the delay period) as compared to normal controls (Sweeney, Takarae, Macmillan, Luna, & Minshew, 2004).

The same properties that make eye movement measures appropriate for studying patient populations (automaticity, lack of metalinguistic judgments) also make them ideal tools for studying infants and young children. Since the 1950s, looking methods have been invaluable for the study of infant cognition.(see Aslin, 2007, for a review of the reliability of looking measures in infancy research) Much of what we now know about the early perceptual and cognitive capacities of infants comes from looking paradigms that rely on infants' preference for familiarity or the habituation of looking over time (and a subsequent preference for novelty) (Fantz, 1961; Aslin, 2007). Looking measures have been used to assess a variety of topics, such as infants' discrimination abilities, early visual and speech processing, object knowledge, and language comprehension. The most commonly used measure is overall looking time to a particular visual stimulus or set of visual stimuli. For studies of visual processing, the visual display is the critical stimulus; for other studies, looking to a visual display in the presence of particular auditory stimuli is measured (intermodal paradigms); in still others, the visual stimulus itself is irrelevant, but looking to the visual display indirectly reflects interest in the critical auditory stimulus.

Anticipatory eye movements and categorization in infancy

In this paper, we focus on anticipatory eye movements as a measure of infants' cognitive processing. These are eye movements that are launched to the vicinity of a visual target even before its appearance. Such anticipations can be a result of the pre-existing cognitive state of the participant (e.g., expecting an object disappearing behind an occluder to re-emerge on the other side), or could be achieved through training.

When during development does the ability to make AEMs emerge? Over the first few months of life, infants' eye movement control and visual processing undergo significant changes (M. Johnson, 1990; Canfield, Smith, Brezsnyak, & Snow, 1997). Newborns show an inability to track objects by smooth pursuit, making quick saccades to track even slowly moving stimuli. At this point their saccades are simply reactive to the presence of a stimulus, not predictive. Smooth pursuit emerges by 2-3 months, marking the earliest motor program that reflects internal tracking of object location (Hofsten & Rosander, 1997). The ability to make AEMs emerges by 3-4 months: at this age, infants anticipate the location of an object that alternates between two locations at a regular time interval (Haith, Hazan, & Goodman, 1988; M. Johnson, Posner, & Rothbart, 1991). When the center of an object's trajectory is occluded, 4-month-olds quickly look to the site of reappearance, but do not anticipate (although they can learn to anticipate after training without an occluder: S. Johnson, Amso, and Slemmer, 2003, see also Hofsten, Kochukhova, and Rosander, 2007). However, by 6 months, infants make anticipatory movements prior to the object's reappearance. These developments may be due to a number of factors, such as the development of a covert attention mechanism that can shift attention to a location in advance of saccade programming (M. Johnson, 1990) or developing object representations that preserve continuity across occlusion events (S. Johnson et al., 2003). Progressions in eye movement control are mirrored by other motor programs, such as hand movements (e.g., online adjustments in hand position, orientation, etc. that anticipate the grasping of an object: Hofsten, 2004).

The categorization paradigm we describe relies not only on infants' ability to anticipate a smoothly moving object's future location, but also on the ability to link particular stimulus features to particular spatial locations. This ability emerges by 4 months: at this age, infants make AEMs to lateral locations predicted by the identity of a central cue (M. Johnson et al., 1991). Four-month-olds are twice as likely as younger infants

to make anticipations in this situation, though the rate is low even in this age group. Furthermore, if stimuli appear in both lateral locations, despite the presence of a central cue that predicts only a single location, 4-month-olds look to the correctly cued side significantly more than chance; thus, both their anticipatory looks and side choice show learning of the side-cue association. With this foundational ability in place by 4 months, it is possible to ask what features of the central stimulus infants rely on to predict future location. In other words, by varying the stimulus properties, it is possible to explore the dimensions that infants both can and naturally do use for categorization.

McMurray and Aslin (2004) developed an AEM paradigm to test infants' visual and auditory categorization. Two versions of this paradigm were implemented. In the first, non-occlusion version, infants view one of two central stimuli (e.g., red square, yellow cross). The stimulus then disappears and after a time lag (that gradually increases over the training trials) a visual*reinforcing stimulus* appears either to the left or right of the screen. The side of reinforcement depends on the identity of the central stimulus. Reinforcing stimuli (moving animals and shapes in this study) act as "rewards" for infants when they orient their gaze to that portion of the screen. Because they are highly engaging, these reinforcers motivate infants to orient their gaze predictively to the anticipated location of appearance. After training on a pre-set number of trials, infants are tested both on the original stimuli and generalization stimuli. The generalization stimuli are new stimuli that either go beyond the stimulus values presented in training or pit the two properties in competition (e.g., red cross, yellow square). In this paradigm, anticipatory looks to particular locations (before the appearance of the reinforcer) reveal which dimensions are represented as relevant for predictive behavior. In cases of competition, the generalization stimuli test whether infants prioritize property A (e.g., color) or property B (e.g., shape) in their categorization. In the second, occlusion-based paradigm, an occluder is used to motivate the temporary disappearance of the object. In this case, the reappearance of the object itself from behind the occluder serves as the reinforcer. Across a number of experiments that varied the particular visual (color, shape, orientation) or auditory (pitch, duration) features of the initial stimuli, a subset of infants aged 5-7 months were found not only to anticipate the location of the reinforcer (or the object's reappearance) during the training phase, but also to generalize to novel stimuli during the test phase, although cases of feature conflict produced different generalization patterns across infants. This type of AEM paradigm has also since been extended to investigate other questions in infant development, such as bilingual speech perception (Albareda, Pons, & Sebastian-Galles, 2008) and cognitive flexibility (Kovács & Mehler, 2009).

Goals of the current system

One limitation of most AEM paradigms is that task performance is not computed online during the experiment, but rather calculated off-line after the experiment has ended. For example, in McMurray and Aslin (2004), all infants received the same pre-set number of training and test trials, regardless of how rapidly or slowly they learned the stimulus-location associations (some infants may not have learned in the time allotted; others may have learned quickly, but become disinterested after additional training trials).

A much more powerful method is to use each infant's on-line, trial-by-trial performance as a measure of learning, which makes it possible to ensure that every infant meets a criterion of performance before entering a generalization phase. In what follows, we describe this new paradigm in greater detail. We introduce a new, flexible software system for the design, online control, and preliminary analysis of such gaze-contingent eye-tracking experiments: Smart-T (System for Monitoring Anticipations in Real Time with Tobii). We describe both the Smart-T system and the associated hardware that enable the implementation of various types of AEM paradigms that rely on automated scoring algorithms. The flexibility of the system allows the user to design other gaze contingent paradigms as well, as is detailed below. Smart-T is fairly platform- and system-independent, having been developed in Matlab ®(R2009b, The MathWorks Inc., Natick, MA) and using the Psychtoolbox (Brainard, 1997; Pelli, 1997). However, at present it relies on another Matlab module, *Talk2Tobii*, which is the critical piece of software that allows Smart-T on-line access to gaze data from compatible Tobii eye-trackers (see below). Therefore, this system is currently limited to the Tobii family of eye-trackers. Importantly, however, the modular nature of the Smart-T system means that it can easily be extended to other eye-trackers with an appropriate module that can access gaze data in real-time. We chose to use the Tobii system because it has a wide (8 inch) field of view of the eyes and it does not require participants to wear any attachment on the head (e.g., a motion sensor) – a major advantage with infant populations. In the next section we describe further specifications of the Tobii system and in the section *System Performance* below, we describe quality controls we use to ensure that the data from the eye-tracker meets appropriate accuracy requirements.

Briefly, the Smart-T system collects participant gaze data using a Tobii eye-tracker and scores each trial online, so that learning by individual participants can be assessed and each participant can be trained to a preset criterion of performance. Notice that there are two senses of "on-line". The first refers to the gaze-contingent updating of the stimulus display, and the second refers to scoring each trial for a variety of criteria, once the trial is completed. In Smart-T, the former is

the case when, for any critical point in the trial time-line, the stimulus display is paused until Smart-T detects the participant's gaze in a pre-defined on-screen observation window. For example, Smart-T can ensure that each trial begins only when the participant is looking at a pre-defined on-screen target. For the latter, various looking criteria like the percentage of total correct looking or the direction of first look can be computed during the inter-trial interval, over a pre-determined number of trials.

The anticipatory eye-tracking paradigm can be used to assess several aspects of visual or auditory categorization, including the time to learn and the robustness of the category to novel exemplars. Smart-T provides a flexible platform for rapidly designing and running AEM paradigms, but the platform is sufficiently general to allow for other kinds of looking paradigms as well (see the section *Programming AEM variants with Smart-T* below). A graphical user interface (GUI) allows for building experiments rapidly with minimal programming skills. After describing the Smart-T system in detail, we provide results from a pilot study of anticipatory eye-movement responses in 6-month-old infants.

Smart-T

Hardware

Smart-T is designed to run on the commercially available Tobii eye trackers (http://www.tobii.com/). These eye trackers are particularly well suited to special populations, including infants, since they do not require the participant to wear any special head gear, yet have a large field of view so that head position is not overly restricted. The eye tracker itself is built into the frame of a video monitor, and the participant merely sits in front and watches the display. Calibration is simple and rapid; as the participant naturally watches a visual stimulus move to various locations around the screen, proprietary software tunes a physiologically derived model of the eyes. The eye tracker can therefore compensate for head movement, and can rapidly re-acquire gaze direction if the participant looks away and back at the screen. Tobii eye trackers are available as 17-inch and, more recently, 24-inch models. The larger screen models provide a wider field of view, and are supposed to provide more robust tracking by using two cameras and automatically selecting appropriate algorithms for dark or light-colored eyes, although we have not ourselves tested these claims.

The calculation and communication of gaze data are performed by the TET (Tobii Eye Tracking) server that is either located in an external PC that communicates data over the TCP/IP port, or in the more recent versions, via firmware embedded into the frame of the display unit itself. Proprietary Windows programs (Clearview and Tobii Studio) are used for calibration and can also be used for simple stimulus presentation and analyses. Our Smart-T system has been most extensively tested with the Tobii 1750 model, wherein the TET server is located on an external PC that also runs calibration software. The 1750 gathers data at 50Hz, while newer models operate at 60Hz or 120Hz. Smart-T itself currently runs on a Mac computer (see below), and communicates with the TET server over the TCP/IP port. This second computer also presents the stimuli - sounds over connected loud-speakers and images on the Tobii display monitor.

As with any equipment, the Tobii system may not work as described by the manufacturers "out-of-the-box." Therefore, we have extensively tested our system to ensure that the relative timing has an acceptable level of accuracy for the types of paradigms used with infants. Details of these tests are described in the *System performance* section below.

Software requirements

On the PC side, we use the software provided by Tobii (Clearview or Tobii Studio) for calibrating participants. Smart-T is written entirely in Matlab and relies extensively on the Psychtoolbox for stimulus presentation and optionally requires the Matlab Statistics toolbox (http://www.mathworks.com/products/statistics/), all of which are cross-platform. However, Smart-T relies on a custom piece of software called *Talk2Tobii* to enable Matlab to interact with the TET server (see below). Since *Talk2Tobii* currently only runs on Macs with Intel processors, Smart-T currently is restricted to Intel Macs (running Mac OS 10.4 or 10.5). *Talk2Tobii* has been developed by Fani Deligianni (http://www.cbcd.bbk.ac.uk/people/affiliated/fani/talk2tobii) and is distributed as precompiled Matlab files that allow Matlab to interact with the TET server over the TCP/IP port. Due to the modular nature of the Smart-T system, it can be easily extended to other eye-trackers and platforms once the appropriate module for exchanging data between Matlab and the eye-tracker has been developed.

Outline of Smart-T

Experiments in Smart-T are programmed via a graphical interface (see Figure A1). An experiment in Smart-T is organized as a series of phases. For example, an initial training phase is typically followed by a test phase. Each phase has (a) a series of structures and (b) one or more conditions for terminating that phase (see Figure 1). Each structure counts as a single trial type, and the sequence of structures can be pre-specified, or can be randomized. The conditions that terminate a phase could be a specified number of trials or a specified duration. The real power of Smart-T, however, lies in the fact that each phase can additionally be terminated contingent on some looking criterion, scored online at the end of each trial of the phase. In this case, the trial in which this looking criterion is achieved serves as the final trial of the phase.

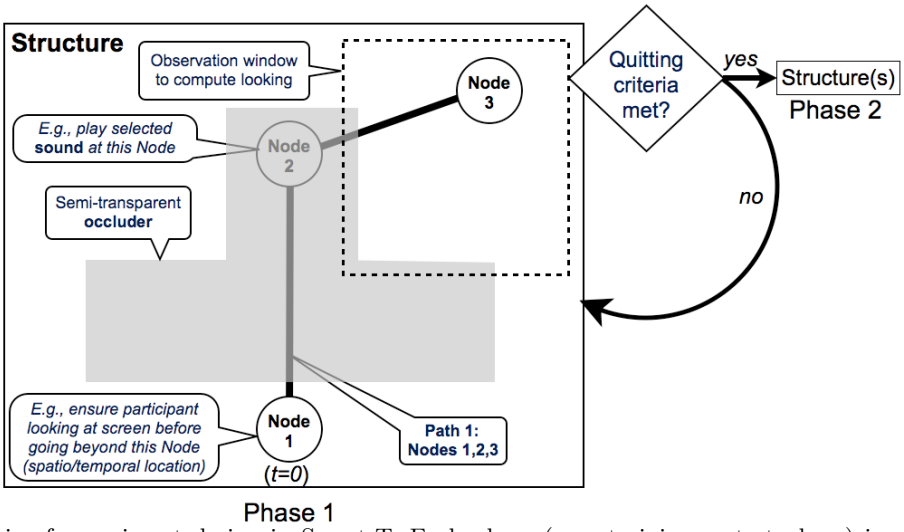

Figure 1. Schematic of experiment design in Smart-T. Each phase (e.g., training or test phase) is a series of *structures* (trial types). The on-screen layout of the structure from the first phase of a hypothetical experiment is shown: the structure contains an object (here, a white circle), an occluder, and a single path with three nodes. Nodes are the spatio-temporal points that define the path the object traverses for a given trial type. The speed of the target object between nodes is set by the experimenter, allowing for timing control (of object movement and presentation of various effects, like sounds). Nodes also allow for experimental control (e.g., gaze-contingent presentation). At the end of each trial, quitting criteria are evaluated and, if met, Smart-T proceeds to the subsequent phase.

Structures. As in other experimental design software, each experiment is conceived as a series of trials. In Smart-T, each (type of) trial is a *structure*, and each structure is envisioned as a *path*. In the kinds of AEM experiments that we take as our starting point (e.g., McMurray & Aslin, 2004), in each trial, a target object moves to one of two (or more) possible goal locations, which is why we chose to describe each trial around the intended path of the target object in that trial.

Each path is programmed as a series of (x,y) screen coordinates called nodes, along which an object can travel. As described below, nodes are spatio-temporal points of interest, which determine how individual trials unfold. The object itself is an image file; Smart-T computes the size and position of the image and uses the Psychtoolbox to animate the motion of the object as it travels along each designated path. The object itself can have arbitrarily complex borders by making certain portions of the image transparent (for image formats that support transparency). The speed and size of the object can be varied from one node to the next. Each node can thus be thought of as a spatio-temporal point of interest – for example, particular sounds (or sets of sounds) can be linked to different nodes, determining the order in which the sounds will be played. Each node is also the potential anchor point for gaze-contingent trial control; if designated as a "wait attention" node, Smart-T pauses the trial at that node until it detects a gaze point in a pre-determined area of the screen (an observation window), and loops any sound, image, or effects (e.g., looming) attached to that node. For example, designating the very first node as a "wait attention" node ensures that every trial starts only when the participant is looking at the relevant observation window, such as a central fixation point.

As described above, many AEM experiments include on-screen occluders behind which the moving object disappears before re-emerging at one of two (or more) different locations. In Smart-T, occluders are programmed through the GUI either as monochromatic polygon masks, or as arbitrary foreground images. In the former case, the transparency of the occluder can be ramped up or down across successive trials. In this way, the occluder can become opaque gradually, allowing the participant to observe the path of the moving object in early trials. In the latter case, Smart-T extracts the alpha map of the image (if the image format supports alpha channels) to provide appropriate opacity or transparency for areas of the image deemed to be occluding or non-occluding, including graded transparency levels. Finally, each structure can be assigned a background image to allow for arbitrary textured backgrounds.

Smart-T allows for a final movie event following the last node. For example, instead of the mere re-appearance of a target object, a complex "reward" movie with the object (constructed offline, prior to the experiment) can instead be played. Again, the size and location of this movie can be easily set through the GUI.

In sum, each structure (trial type) has a visual object traveling along a predetermined path interpolated between spatio-temporal points of interest (nodes), behind optional occluders and in front of optional backgrounds. Key events like disappearances and re-appearances, sound effects, and gaze-contingent pauses are achieved

by attaching them to the various nodes. Timing control is achieved by varying the distance between nodes, the relative speed of the object between nodes, and by introducing loom effects (see example 1 in *Programming AEM variants with Smart-T* below). For illustrative purposes, the appendix outlines how a simple experiment is built in Smart-T.

Phases. A sequence of structures constitutes a phase, and each experiment can be made up of several distinct phases. For example, an experiment might consist of an initial training phase followed by a test phase. The order and frequency of appearance of each structure within a phase can be randomized or it can be fixed. The frequency of appearance is determined by the relative numeric weight assigned to each structure. For example, two structures with weights 7 and 3 will ensure that, on average, the first structure will appear 70% of the time, while the second will appear 30% of the time. Smart-T can be forced to ensure that this is exactly the case, e.g., over each block of 10 trials (in this example).

One of the main powers of Smart-T lies in the fact that moving from one phase to the next can be programmed based on several criteria. In the simplest case, this is a fixed number of trials. However, Smart-T can also compute various indices that define a criterion of learning. The user specifies a series of rectangular observation windows via the GUI, and, for each relevant structure (trial type), the set of correct and incorrect observation windows for that structure. The spatio-temporal period of interest, for example the period of occlusion, is programmed by specifying the beginning and end nodes that mark the critical period. Smart-T uses this information to compute the looking pattern during each trial, aggregated over the last 'n' trials, where 'n' is specified by the user as part of the various criteria.

We have implemented three indices, which can be invoked in any Boolean combination. The first index is the total looking in the correct windows compared to the total looking in correct and incorrect windows over the last 'n' trials. The second measure is the direction of first look - the observation window in which the first (possibly anticipatory) gaze lands in the last 'n' trials. The final, more stringent measure, is a t-test between the looks to the correct and incorrect windows computed over the last 'n' trials. In addition, quality control parameters ensure that only trials in which sufficient reliable data is gathered are used for computing criterion. For example, trials in which the infant was not looking at the display for more than a pre-specified amount of time or trials in which the quality of the data from the eye tracker is poor, can be excluded from the on-line analyses.

Testing the experiment. Smart-T includes a debug mode for testing the setup of the experiment. In the GUI, when 'Connect Tobii' is unchecked (see Figure A1), the user can run the entire experiment without the Mac being connected to a Tobii. Checking 'Debug' in the GUI allows the user to use the mouse to simulate eye gaze; Smart-T draws a yellow dot on the screen at the location of the mouse pointer to represent 'eye gaze'. In addition, Smart-T draws all the observation windows and paths defined for that experiment. When the experiment is run in this mode, the nodes of the path for the current trial are highlighted. The x-y coordinates of the on-screen gaze/cursor position are also presented to help fine-tune the placement of the various nodes and observation windows. For phases with quitting criteria, current values corresponding to each of the selected quitting criteria are also displayed on the screen at the end of each trial. By moving the 'gaze' (mouse) to the various observation windows, the user can verify that the experiment is correctly set up to compute the specified criteria for each phase.

Programming AEM variants with Smart-T. Although primarily designed to facilitate AEM experiments, which typically involve objects moving behind occluders (or disappearing and reappearing), the flexible nature of the Smart-T system makes it easy to program other simple gaze-contingent experiments by manipulating how structures are set up. Some examples include:

1. Preference/looking time studies: Turn off all occluders and use Smart-T to present a single audio, visual or multimodal stimulus from a pre-determined set, per trial. For example, in each trial the user sets the size of the visual object at the first, central node to zero, and waits for the infant to direct attention to the center of the screen, consisting of a fixation target on the background image. Contingent upon detecting gaze in an observation window centered around the central fixation target, the second node is triggered. The visual object can be made to stay on-screen by attaching to the node a loom of size 100% (i.e., a constant size), for as long as required. In the simplest case, for example, looking time to two different visual stimuli could be compared across trials.

2. Dynamic spatial indexing studies (e.g., Richardson & Kirkham, 2004) with multiple target locations can be implemented by varying the background image, which can consist of pre-drawn target locations (boxes). The object size is set to zero for the first few nodes, and in one of these nodes the auditory cue plays. Subsequently, the visual target is set to 100% at a node centered in one of the target boxes. Anticipatory responses can be coded between the node at which the sound plays and the node at which the object "appears".

3. Arbitrary trials can be programmed as movies, and by attaching movies to the different structures as "reward" movies, each trial can consist of just a single node where Smart-T waits for the participant to look at the screen, and then plays the "reward" movie. This allows implementing preference studies with movies instead of still images. Arbitrary movies can also be interspersed as attention-getting trials.

4. By carefully placing objects on background images, select portions of a more complex visual scene can be made to move contingent on them being looked at. For example, we are currently using a visual scene paradigm in which, in each trial, one of two objects on a table is the target, and, contingent upon being looked at, it looms, moves, and a background voice "speaks" the name of the object. This is achieved by having a background image with the target missing; the object is then placed (with loom and moving effects) over the background image at nodes corresponding to the desired locations.

Data overview

The smallest measurement unit in the Tobii system is a single frame of gaze data. Since the Tobii 1750 collects data at 50 Hz, each frame has a temporal 'width' of 20 ms. Smart-T queries the Tobii eye-tracking (TET) server at a faster rate (60 Hz), and updates its internal data frame every time it differs from the previous data frame. Since each data frame from the TET server is time-stamped, each data frame internal to Smart-T is unique. Smart-T's internal data frame is a combination of eye gaze data from the TET server and data pertaining to the current status of the experiment. This includes the time-stamp from the Mac, the turn-around time required for querying the TET server for each data frame, the current information about experiment phase, trial, structure, path and node, the current x-y coordinates of the moving object, as well as the presence/absence of gaze in the various observation windows (a Boolean variable). In the *System performance* section below, we demonstrate the timing quality of the data obtained in the course of the experiment reported below. In brief, the temporal "width" of 20 ms is obtained more than 97% of the time (i.e., frame drop-out or double-counting is very rare).

The gaze data from individual data frames is used by the Tobii software to determine fixations, by employing a variety of fixation filters. However, in order to move away from having to determine individual fixations, we have chosen to use the raw gaze data as our primary variable of interest (cf. McMurray & Aslin, 2004). So, for example, in computing the proportion of looking to one observation window over another (see the *Phases* section above), Smart-T counts up the total number of valid data frames during the appropriate nodes in the two windows of interest. This measure therefore does not rely on an arbitrary determination of what constitutes a single fixation (eliminating times when the eye is moving by some criterion) and count the number of fixations, but is simply a measure of the total time that gaze position is located within each of the two observation windows during the time interval of interest. These numbers are also used to compute the t-test, since for each trial Smart-T records the number of data frames in which gaze is located in each of the relevant observation windows. Similarly, first look is computed based on the first data frame when gaze enters one of the two windows of interest.

Each experiment, designed through the GUI, is stored as an editable text file. This allows users to keep a record of the precise experimental settings and also allows for easy modification of an experiment using standard text editing functions. In order to minimize memory requirements, Smart-T computes and saves data on the fly to the disk into a temporary directory, retaining in RAM only the minimum data required to determine if criterion is reached for the end of a phase. Therefore, most of the computer RAM is kept free to support accurate stimulus presentation. In addition to the eye gaze data, Smart-T also records timing data. In particular, it stores the time-stamp of each data request to the TET server (via *Talk2Tobii*), a time-stamp of each response received, and the time-stamp of the data received from the TET server (see the reported experiment below for an analysis of timing data). At the end of each experiment the stored data is collected into a single Matlab (.mat) file with a unique filename that reflects the date and time when the experiment was run. While Smart-T stores a complete record of the entire experiment as a Matlab file, a Matlab helper function extracts the most important details into a comma-separated (.csv) file that can be read directly by text editing or spreadsheet programs, and also directly into Matlab. The helper function also generates a small text file that summarizes the course of the experiment – in particular, if and how each phase was terminated. Finally, a second Matlab helper function plots the output of the experiment in a series of figures, one for each phase of the experiment. Each figure displays WxP subplots that show eye gaze data for the 'W' different observation windows for each of the 'P' different paths in that phase. Each subplot shows the eye gaze data during the course of the entire path, with the nodes marked as they occur in time. For each path, the eye gaze is plotted as the proportion of all trials when the gaze rested in that particular observation window at each point in time. An example of such a plot is shown in Figure 2.

In the next section, we describe a pilot experiment with 6-month-olds that demonstrates Smart-T in action.

Anticipatory looking in 6-month-olds: A pilot study

As described in the Introduction, by six months of age, infants are capable of making anticipatory looks to the expected location where an occluded object will reappear. In this pilot study, we presented 6-month-olds with a training phase in which multi-modal 'objects' that were the combination of an image and a spoken syllable predicted the side on which the object re-appeared. In a subsequent generalization phase, training stimuli were presented intermingled with novel stimuli wherein the syllables were spoken by a different talker. Smart-T was used to record eye gaze data throughout the exper-

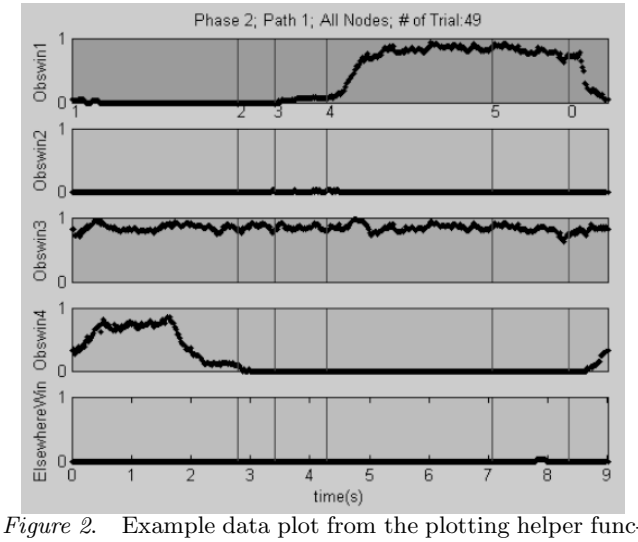

Figure 2. Example data plot from the plotting helper function showing the proportion of fixations at each time-point for all the observation windows over all (n=49) trials for a single participant. At the onset of the trial, most looks are in observation window 4 (Obswin4), but starting at node 4, the proportion of looks in window 1 (Obswin1) increases. Note that Obswin3 represents an observation window that covers the entire screen. Obswin1 has the node numbers marked on the x-axis. The final node is labeled '0' and marks the end of the trial; the time points beyond it represent the inter-trial interval. The final subplot (ElsewhereWin) shows gaze points that are not in any of the other observation windows, and its x-axis is the time axis (in seconds) for all subplots.

iment, to score the trials in the training phase, and to quit the training phase when a specified learning criterion was met so that the infant could move on to the generalization phase.

Materials & methods

Stimuli. Each object was a multimodal stimulus comprising an image (Elmo or Cookie Monster from the children's show Sesame Street) and a spoken syllable ('ta' or 'ga'). Both images were 258 pixels in extent and subtended approximately 4° of visual angle as their starting size. In the training phase, Elmo was consistently paired with the syllable 'ta' in a female voice, while Cookie Monster was consistently paired with the syllable 'ga', also in a female voice. In the "old" trials of the generalization phase, these combinations of syllable, gender, and image were maintained. In the new trials of the generalization phase, the image-syllable pairing was maintained, however the syllables were now spoken in a male voice. Three syllables of each type were recorded by each talker, and these were edited to be matched for duration and loudness.

Participants. Participants were eleven full-term infants with no known hearing or vision deficits. Infants were around 6 months of age (mean 177 days), and two were girls. Of these infants, only one did not provide sufficient eye gaze data to be included in the behavioral analysis (nevertheless, we could still use his data to estimate the communication performance between Smart-T and the eye-tracker, as described in the *System Performance* section below). Informed consent was obtained from the caregiver of each infant. All procedures and protocols were approved by the University of Rochester Research Subjects Review Board.

Procedure. Infants were seated on a caregiver's lap about 60cm away from the 17-inch Tobii screen. The Tobii system was calibrated using Clearview by presenting small circles that moved to one of 5 locations on the screen and "shrank" to attract the infant's gaze. The caregiver wore sound attenuating headphones over which they heard masking music, and wore a soft felt visor to help block out the Tobii screen so as not to bias their infant. Caregivers were instructed to pay as little attention to the stimuli as they could.

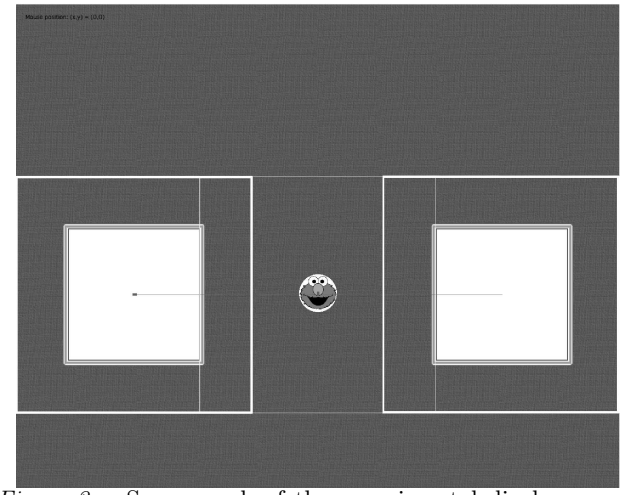

Figure 3. Screen grab of the experimental display screen in debug mode. In each trial, the object image started at the central location, where it waited for a gaze point in the central window to launch the trial. At this point, a single syllable was played and the object then shrank in size and disappeared. It appeared 1500 ms later in one of the two target windows, depending on the image+syllable combination. White rectangles mark the three (overlapping) observation windows, while the line represents the paths along which the object travels, to the left and the right of the central location.

The experimenter started the experiment by pressing the 'S' key, which initiates the first trial. In each trial of the first (training) phase, the image of Elmo or Cookie Monster appeared at the central location and loomed (from 100% to 150% in size over 800 ms), accompanied by playful sound effects to orient the infant towards the screen (see Figure 3). When Smart-T detected that the infant's gaze was directed to the central observation window, a syllable token was presented in a female voice, and the visual object then shrank to disappear at the central location. Subsequently, the 'disappeared' object

traveled over a period of 1500 ms to one of the on-screen rectangular windows, 304 x 304 pixels (˜7.5°) in size, located symmetrically about the vertical axis, their centers 90 pixels away from the monitor midline, and 103 pixels below the horizontal midline. It then reappeared at twice the original size, loomed rapidly (twice over 300 ms), and finally loomed once more over 1 sec, while a second syllable token of the same type played as reinforcement. For all infants, Elmo was paired with the syllable 'ta', while Cookie Monster was paired with the syllable 'ga'. Elmo reappeared in the left window, while Cookie Monster reappeared in the right window.

Learning was assessed for the anticipation period, the 1500 ms period during which the object was invisible. As described above (*Data overview* section), anticipations were computed by counting the number of valid data frames in each of the two target windows. The rectangular target windows were set larger than the on-screen windows, and extended 500 pixels inwards from the right or left edge of the screen, subtending a 12.3° visual angle in the horizontal and vertical directions. For example, if out of the 75 frames that make up the 1500 ms anticipatory period, 50 were in the correct window, 20 in the incorrect window, and 5 were invalid, Smart-T would count this as a 71.4% bias in the correct direction. In the current experiment, this bias was averaged over the last 5 trials of the training phase to get a more robust estimate of the anticipatory response. If this averaged response, computed automatically by Smart-T, exceeded a threshold (65% in our case), or a t-test between the five (paired) values of valid gaze points in correct vs. incorrect windows was significant, Smart-T judged that performance as successful learning. Finally, if the first gaze data frame was in the correct window in at least 3 of the 5 previous trials, it was also taken as evidence of successful learning.

In the generalization phase, we presented infants with eight blocks of four trials each – two 'old' trials that were identical to the training trials and two 'new' trials. The new trials had the same character+syllable+side association, but now the syllable was spoken in a male voice. Importantly, for the new trials, the object never re-emerged. The lack of reinforcement for these trials meant that infants were not "told" how to categorize the novel stimuli; rather, their response to these stimuli depended on their ability to generalize the learned character+syllable combination to a new voice. The first two trials of the test phase were always the two new trials. The mixture of old and new trials during the generalization phase was designed to reduce the likelihood that infants would forget the original relationship between the two cueing stimuli (Elmo+/ta/ or Cookie Monster+/ga/) and the location of object reappearance.

Results

Figure 4 shows the number of training trials completed by each of the 10 included infants for whom reliable gaze data was obtained. Only 1 of these 10 infants did not reach any of the pre-set criteria, and instead completed the 32, maximum allowed training trials.

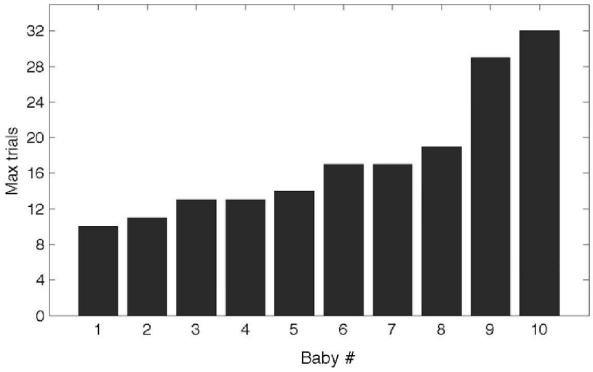

Figure 4. Number of training trials (max=32) for the infants in the pilot study, sorted by rank. All but one reached a desired looking criterion. Infants 1, 7 and 9 passed both on total looking and on the first-look criteria, infants 3, 4 and 5 passed both on total looking and the t-test criteria, and infants 2, 5 and 8 passed on only the total looking criterion.

Figure 5 shows the overall results from the training and generalization phases for the nine infants that reached criterion in the training phase. The bars represent the proportion correct looking across these nine infants for the last five trials of the training phase (first bar) and the 'old' and 'new' trials from the generalization phase (second and third bars, respectively). For these infants, two-tailed, one-sample t-tests revealed that the proportion correct looking was highly significant for the last five trials of the training phase ($t(8)=3.95$, $p=0.004$). For the generalization phase, proportion correct looking was significant for the 'new' trials ($t(8)=2.34$, $p=0.047$), but did not reach significance for the 'old' trials ($t(8)=1.72$, $p=0.12$).

Discussion

In this pilot study, we demonstrated that 6-month-old infants can learn to predict the appearance of a multimodal (cartoon character+spoken syllable) stimulus, and that they can generalize this learning when the acoustic properties of the syllable stimuli change. Why did the proportion of correct looking fail to reach significance for the old trials in the testing phase? One possibility is that, since the first two trials of the test phase were the new trials, infants immediately generalized to these novel stimuli, but were disrupted by the switch back to the previously heard voice. A second possibility is that, having gone through the training with the old voice, they were less motivated to respond in the test phase to that same voice.

Given the nature of our stimuli, the results are consistent with several possible ways of encoding the relevant stimulus dimensions. That is, our learning and generalization results could be a result of infants encoding

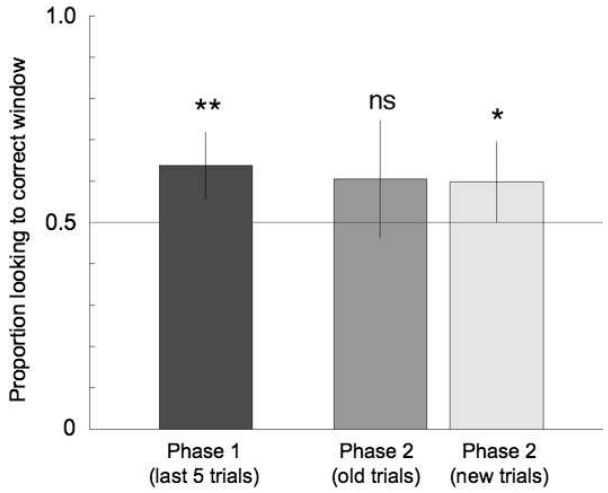

Figure 5. Results from the pilot experiment with 6-month-olds. The bars show the mean 'correctness' score (looking to the correct window / total looking to both windows), averaged across the nine infants that reached criterion in training. The first bar is the mean score for the last five trials in the training phase, the second is the mean score for the 'old' trials in the generalization phase, while the third bar is the mean score for the 'new' trials in the generalization phase.

the stimuli as: (1) abstract multimodal visual-auditory gestalts consisting of a cartoon character plus a syllable, ignoring the surface properties in both modalities (2) abstract auditory stimuli (syllable type), ignoring the visual dimension completely or (3) along the visual dimension alone, ignoring the accompanying sounds. Further experiments are needed to understand what dimensions infants use to categorize multimodal stimuli such as these. For example, generalization trials in which the visual and auditory dimensions are swapped would reveal which dimension infants rely on to predict the upcoming appearance of a multimodal object.

The major technical advancement over prior experimental paradigms is the fully automated, on-line coding and evaluation of learning. As can be seen from Figure 4, infants took between 10 and 30 trials to reach pre-determined criteria that indicate learning. Smart-T can compensate for these differences in learning rate by ensuring that infants proceed to the test/generalization phase when they show evidence of learning.

System Performance

In this section we describe the performance of the setup as estimated from the pilot experiment. First, we describe the accuracy of the Tobii eye-tracking itself. Calibrations in the Tobii system currently only return graphical estimates of calibration success or failure. To visualize the accuracy of eye tracking, we plotted the gaze data across the entire experiment for the 10 infants who were successfully tracked. Figure 6 shows the distribution of gaze points in the x-y plane of the screen (top subplot) and the distribution along just the x-axis (lower subplot). The figure demonstrates that gaze points are distributed in three main regions, corresponding to the left and right target windows and the central starting point. The concentric rings in each region in the upper subplot show the density of gaze points within these regions (also reflected in the three bold Gaussians in the lower subplot; see caption for details.) The plot also shows that there is some variability, presumably due to differences in the quality of calibrations. As estimated from the location of the central peak, this difference is less than 55 pixels (˜1.5°), which is substantially less than the size of the observation windows (500 x 500 pixels, or ˜12.3° of visual angle).

In order to quantify the quality of the calibration, we looked at both the accuracy (the mean absolute distance of an individual gaze point from the intended target) and precision (the variability of the distribution of gaze points) of the gaze points. We extracted gaze points along the x-axis that correspond to the central peak, which is expected to be at the screen mid-point of 640 pixels. The location of the central peak was computed by examining the second derivative of a low-pass filtered distribution of gaze along the x-axis with a 20-pixel-wide first-order filter. The extents of the central peak were computed as the nearest 'valleys'. The mean peak location across all 10 infants was 650.1(±23.5) pixels, ranging from 628 to 682 pixels. Therefore, infants on average showed an accuracy of 10 pixels (that is, they were ˜10 pixels away from the intended location), with a precision of about 100 pixels (95% confidence). Notice that the precision, which is a measure of the variability of the gaze points around the central peak, is affected by both the quality of the calibration and the natural variations in gaze position as infants fixate the central target. Therefore, these are conservative, upper limits for the precision. Such a natural imprecision in gaze position can also contribute in some small part to the measured accuracy. Nevertheless, these values are smaller than the size of the target objects and the observation windows. The bold Gaussian curve in the lower subplot in Figure 6 gives a pictorial representation of this accuracy and precision.

As mentioned earlier, for each data frame, Smart-T records the Mac time-stamp when it queried the TET server and when it received a reply, and the PC time-stamp contained in the reply from the TET server.

Since all computer clocks run at slightly different frequencies, we first estimated how much lag was experienced by the system due to the discrepancies in the clocks on the two computers. We found that the Mac clock was a little faster than the PC clock, and there was a slight variation for each of the experimental runs for the 11 participants. Nevertheless, this time difference was small, amounting to a discrepancy of 5.88 ms per minute (3.0 ms standard deviation). Of course, this discrepancy is expected to vary for each pair of computers.

Second, we asked how long Smart-T took to receive

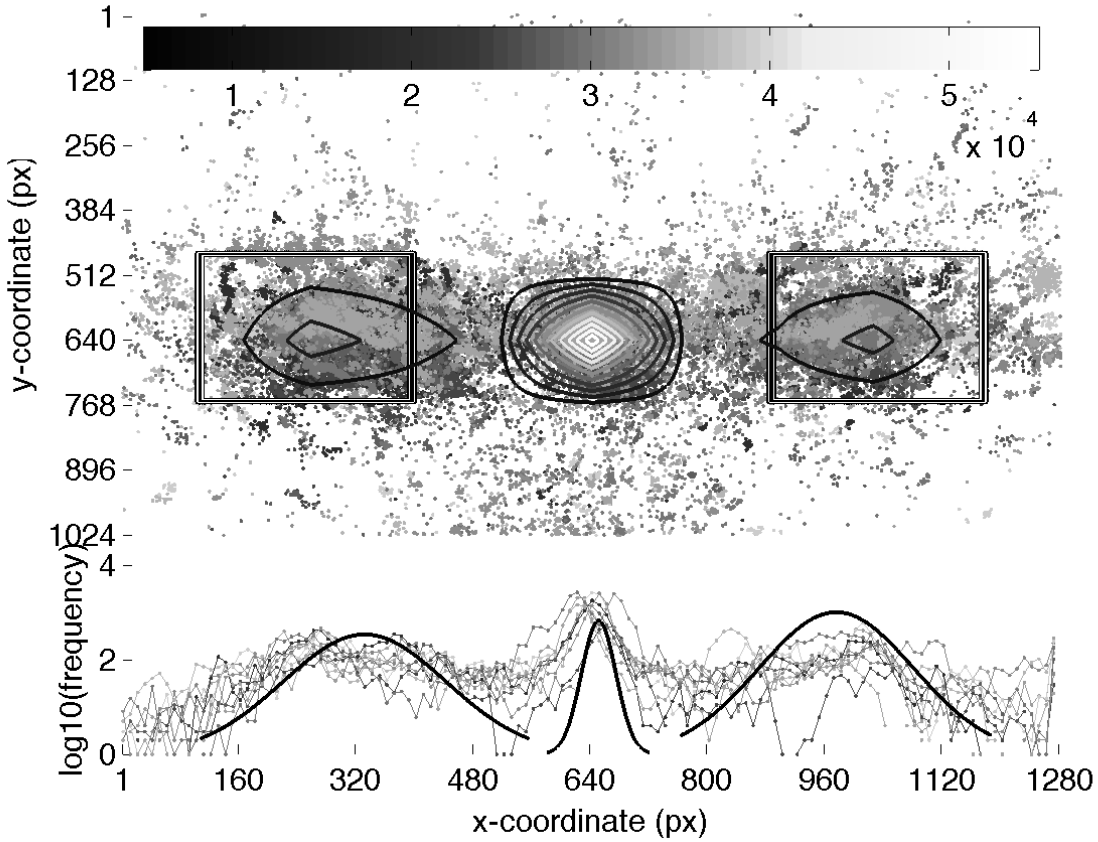

Figure 6. Distribution of gaze points across the entire experiment for all 10 infants who were successfully tracked. The upper subplot shows a scattergram of x-y gaze points for all infants, with the two rectangular observation windows and an overlaid contour map. The contour lines show a central, high peak (lighter shades are higher counts, as shown in the color scale bar at the top), and two shallower peaks, mostly within the two lateral observation windows. The lower subplot shows the distribution of gaze points along just the x-axis (on a log10 frequency scale). The central bold Gaussian in the lower subplot represents the estimated distribution of gaze along the x-axis for the central peak (±3 S.D.), and reflects the quality of the calibration. The two lateral Gaussians represent the estimated distribution of gaze along the x-axis for the wider, lateral peaks (±2 S.D.).

a response from the TET server once it requested data. This turn-around time was very small, never exceeding 0.2 ms throughout the course of the experiment for all 11 infants. Third, we asked how often the system missed a data frame from the TET server. As mentioned earlier, the Tobii 1750 gathers eye gaze data at 50 Hz. Therefore, we expect that the 'width' of each sample (the time from one data frame to the next) is 20 ms. Figure 7 shows a histogram of all sample widths across the experimental runs from all infants (note the logarithmic y-axis in the main graph). As can be seen from the graph, there is some variation in sample widths, with the system sometimes missing data frames (corresponding to the peaks at multiples of 20 ms). However, data loss is extremely rare - no more than 5 consecutive data frames were ever lost, and data loss (of any number of frames) occurs less than 3% of the time (see inset figure).

These timing data still leave one unresolved issue - is there a discrepancy between when the system reports that the eye gaze has moved and when the eye itself has actually moved? In order to estimate this discrepancy, we tested the system with a high-speed camera. In a simple setup, we presented an adult participant with an image that moved between two marked locations, one on the bottom-left and the other on the top-right of the Tobii monitor. The participant was instructed to look at the window that the object currently occupied, and to move her eyes as quickly as possible when the object moved to the other window.

Using a mirror located next to the Tobii monitor, we simultaneously video recorded both the appearance/disappearance and the location of the object on the Tobii monitor and the gaze direction and changes in gaze direction of the participant. A high speed camera (the Casio Exilim EX-F1) captured video at 300 frames per second. Using the appearance/disappearance of the

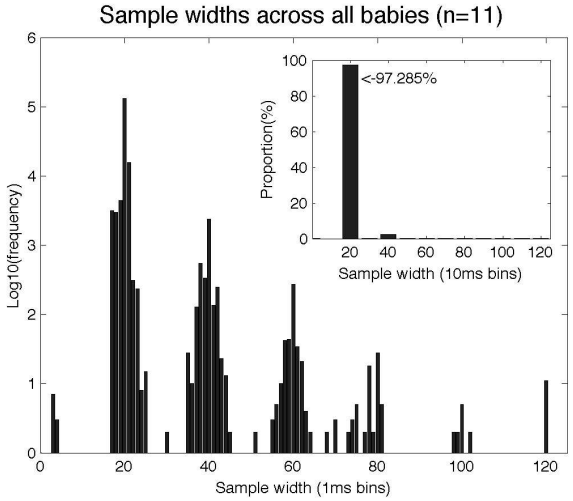

Figure 7. Timing data showing the distribution of sample "widths" – the time between one data frame and the next. The expected value of ˜20 ms is obtained >97% of the time.

object as a proxy for the timing of the nodes in Smart-T, we were able to estimate the time of shifts in gaze independently via Smart-T and via the high-speed camera.

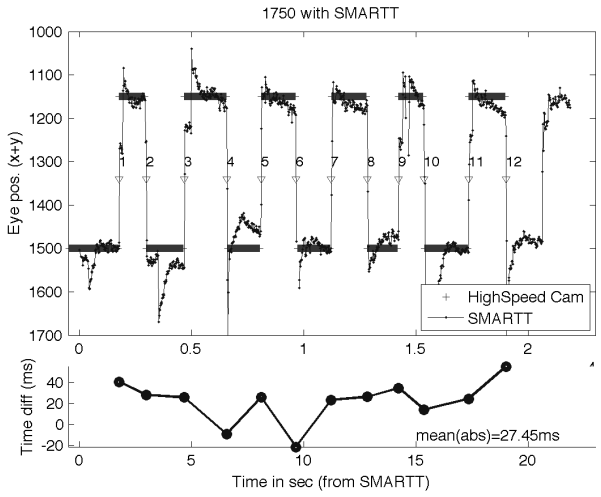

Figure 8. High-speed camera timing test results. The top subplot shows the eye-position as coded from the high-speed camera and as reported by Smart-T. A Matlab script determined the mid-point of the transition from one position to the next (inverted arrowheads). The lower subplot shows the timing difference between the transition points according to the high-speed camera and according to Smart-T.

Figure 8 shows the relation between the eye gaze position as estimated from Smart-T and from the high-speed camera. Numbered, downward-pointing arrowheads mark the (automatically detected) changes in gaze position as detected by Smart-T. As the lower sub-figure shows, there is a small variation between the two (a mean absolute timing difference of 27.45 ms). These figures depend on the hardware and software running on the computers – a Mac for Smart-T and a PC for the eye-tracking server. What we show is that, under the right conditions, Smart-T can provide eye-gaze data that is accurate to under 100 ms, more than 95% of the time: more than 95% of the samples are collected at the expected time (Fig. 7), and the range of delay between the position of the eye as detected by the high-speed camera and by Smart-T is less than 100 ms (Fig. 8). Finally, we add a caveat that, given the timing performance, the system is currently not suitable for eye-tracking paradigms that rely on saccade data.

Conclusions and future directions

In this paper, we describe a system for developing and running fully automated eye-tracking paradigms. This system uses the Tobii eye-trackers in conjunction with custom Matlab software built around the *Talk2Tobii* module for communicating with the eye tracker. In a pilot study we demonstrate the feasibility of using this system to examine learning in 6-month-old infants. Given the growing interest in the Tobii system (e.g., "Tracking infants' eye gaze patterns using the Tobii system" (2010); Breakfast roundtable at the XVI-Ith Biennial International Conference on Infant Studies, Baltimore, MD) and in fully automated eye-tracking paradigms, Smart-T provides a first solution for utilizing their potential. The application of Smart-T to other eye-tracking systems is relatively straightforward, requiring only the development of modules equivalent to *Talk2Tobii*, that can be integrated with the eye-trackers for on-line data collection.

References

Albareda, B., Pons, F., & Sebastian-Galles, N. (2008). The acquisition of phoneme categories in bilingual infants: New data from a new paradigm. *Poster presented at the ICIS, Vancouver*.

Altmann, G., & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition*, *73*(3), 247-264.

Aslin, R. (2007). What's in a look? *Developmental Science*, *10*, 48-53.

Brainard, D. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433-436.

Brock, J., Norbury, C., Einav, S., & Nation, K. (2008). Do individuals with autism process words in context? Evidence from language-mediated eye-movements. *Cognition*, *108*(3), 896-904.

Canfield, R., Smith, E., Brezsnyak, M., & Snow, K. (1997). Information processing through the first year of life: a longitudinal study using the visual expectation paradigm. *Monogr Soc Res Child Dev*, *62*(2), 1-145.

Cooper, R. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, *6*, 84-107.

Fantz, R. (1961). The origin of form perception. *Sci Am*, *204*, 66-72.

Haith, M., Hazan, C., & Goodman, G. (1988). Expectation and anticipation of dynamic visual events by 3.5-month-old babies. *Child Dev*, *59*(2), 467-479.

Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends Cogn Sci*, *9*(4), 188-194.

Hofsten, C. von. (2004). An action perspective on motor development. *Trends Cogn Sci*, *8*(6), 266-272.

Hofsten, C. von, Kochukhova, O., & Rosander, K. (2007). Predictive tracking over occlusion by 4-month-old infants. *Developmental Science*, *10*, 625-640.

Hofsten, C. von, & Rosander, K. (1997). Development of smooth pursuit tracking in young infants. *Vision Res*, *37*, 1799-1810.

Hornof, A., & Halverson, T. (2002). Cleaning up systematic error in eye-tracking data by using required fixation locations. *Behav Res Methods Instrum Comput*, *34*(4), 592-604.

Johnson, M. (1990). Cortical maturation and the development of visual attention in early infancy. *J Cogn Neurosci*, *2*, 81-95.

Johnson, M., Posner, M., & Rothbart, M. (1991). Components of visual orienting in early infancy: Contingency learning, anticipatory looking, and disengaging. *J Cogn Neurosci*, *3*, 335-344.

Johnson, S., Amso, D., & Slemmer, J. (2003). Development of object concepts in infancy: Evidence for early learning in an eye-tracking paradigm. *Proc Natl Acad Sci U S A*, *100*(18), 10568-10573.

Kovács, Á.M., & Mehler, J. (2009). Cognitive gains in 7-month-old bilingual infants. *Proc Natl Acad Sci U S A*, *106*(16), 6556-6560.

Loftus, G., & Mackworth, N. (1978). Cognitive determinants of fixation location during picture viewing. *J Exp Psychol Hum Percept Perform*, *4*(4), 565-572.

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

Pelli, D. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437-442.

Richardson, D., & Kirkham, N. (2004). Multi-modal events and moving locations: Eye movements of adults and 6-month-olds reveal dynamic spatial indexing. *J Exp Psychol General*, *133*, 46-62.

Sweeney, J., Takarae, Y., Macmillan, C., Luna, B., & Minshew, N. (2004). Eye movements in neurodevelopmental disorders. *Curr Opin Neurol*, *17*(1), 37-42.

Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of visual and linguistic information during spoken language comprehension. *Science*, *268*, 1632-1634.

Yarbus, A. (1967). *Eye movements and vision.* New York: Plenum Press.

Yee, E., Blumstein, S., & Sedivy, J. (2008). Lexical-semantic activation in Broca's and Wernicke's aphasia: evidence from eye movements. *J Cogn Neurosci*, *20*(4), 592-612.

Appendix
Designing a simple experiment in Smart-T

In this appendix, we outline the set up of a simple anticipatory eye-tracking study using Smart-T. In this hypothetical study, we would like to see if infants can make simple predictive responses to two distinct multimodal stimuli. Say the goal is to test if infants can learn that a central red square accompanied by a speech syllable ('ba') disappears and re-appears in a left on-screen window while a central blue square accompanied by a piano C note disappears and re-appears in a right on-screen window.

Required media files

1. Image of a red circle, with parts of the image beyond the boundaries of the circle set to 100% transparency (e.g., in the .png format).
2. Image of a blue square, as above.
3. A background image with a uniform color or a texture, and two distinct rectangular "windows" symmetrically placed on either side of the screen center (see Figure 3). Ideally, this image is the (pixel) size of the Tobii screen, e.g., 1280x1024 for the Tobii 1750. These rectangles are the on-screen target windows in which the objects will reappear.
4. A wav file of the syllable 'ba'.
5. A wav file of a piano note in middle C. Ideally, both sounds are edited to the same duration.

Observation windows

1. A small central window encompassing the central, starting location of the objects.
2. A window encompassing the left re-appearance location.
3. A window encompassing the right re-appearance location.
4. The entire screen.

Paths

We only require two paths, one to the left and one to the right, starting from an origin point at the horizontal center of the screen, and in line with the horizontal line joining the centers of the two re-appearance windows. For our purposes, each will be a series of 7 nodes to accommodate the various effects as follows:

1. First node, one pixel below the origin point, object size=1, loom effect (e.g., grows to 150% and back to 100% over 0.5 sec), designated as a "wait attention" node. Set correct observation window to window 1 (central window). Smart-T will pause at this node, waiting for an eye-gaze in this observation window at the beginning of each trial.
2. Second node at the origin point, object size=1, with a 100%, pseudo-loom (i.e., with no change in size) for the duration of the sound files.

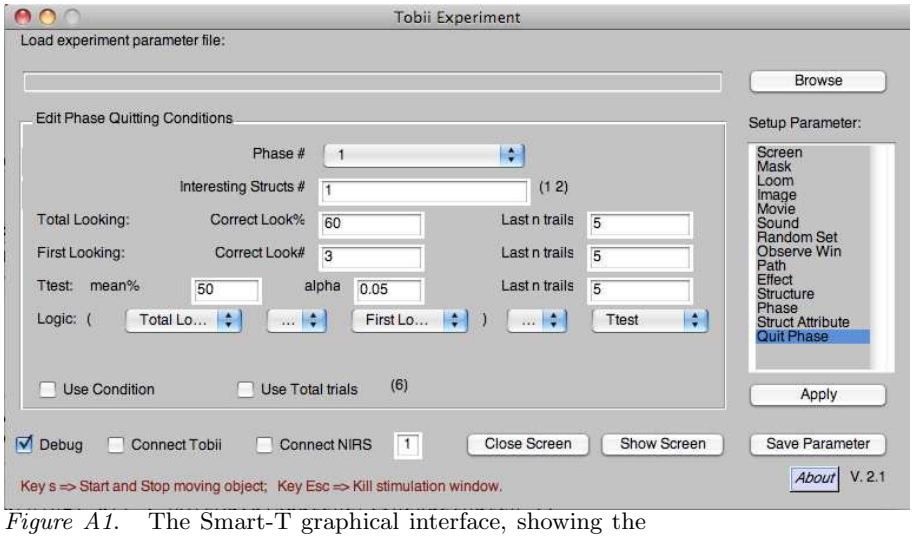

Figure A1. The Smart-T graphical interface, showing the panel that allows the user to specify the criteria that must be met if a phase is to be terminated contingent on looking behavior. See text for details.