

TRAINING IN ANTICIPATORY LOOKING EXPERIMENTS WITH ADULT PARTICIPANTS

Johannes Bjerva, Ellen Marklund & Francisco Lacerda

Department of Linguistics, Stockholm University, Sweden
bjerva@ling.su.se; ellen@ling.su.se; frasse@ling.su.se

ABSTRACT

The amount of training necessary to trigger anticipatory looking was investigated in adults (n=16) using a simple testing paradigm, in order to create a baseline for studies on infants' language acquisition. Participants were presented with training containing implicit associations between two syllables (/da/ and /ga/) and visual events displayed on different areas on the screen. The training series were periodically interrupted by test trials where a syllable was presented but no visual event was displayed. Significantly altered looking behaviour, as measured by participants' first gaze fixation latency towards the Non-target area (where the visual event should not be expected), was found after 28-36 training trials.

Keywords: anticipatory looking, methodology, eye-tracking

1. INTRODUCTION

Methodological aspects are of critical importance in experimental research on infants' language acquisition. Studies often focus on pre-linguistic infants and their abilities, and in order to obtain interpretable answers to one's research questions, the experiment paradigm needs to be adapted for infants. The Anticipatory Looking Paradigm (ALP) relies not only on the infant's tendency to rapidly form audio-visual associations, e.g. [9], but also on anticipatory eye-movements produced revealing the infant's current perception of the underlying audio-visual associations. However, the amount of information that the infant needs in order to derive such underlying audio-visual associations and consistently show anticipatory looking behaviour is not certain. The present study aims to investigate this issue by measuring the amount of exposure required by adult subjects in order to achieve stable anticipatory looking behaviour, in a test paradigm similar to those used to test infants.

2. BACKGROUND

A necessary aspect of an ALP experiment is that the participants make anticipatory eye-movements, i.e. look somewhere in anticipation, before a stimulus has been presented. At the age of four months, infants show anticipatory looking behaviour when trained to expect a visual stimulus on one of two screens depending on which attractor-stimulus had been previously presented on a third, centred, screen [4]. Adults have been shown to look at potential visual targets (as determined by the linguistic context) before any target word has been presented [1, 5].

The spontaneous looking behaviour towards a visual stimulus when hearing an associated auditory stimulus is also essential for ALP experiments. Infants have been shown to do so, e.g. [10]. In a similar vein, adults' eye-movements have been shown to reflect the linguistic input; upon hearing a word describing one of several visible objects, they spontaneously look at the named object, e.g. [2, 3, 11].

The ability to visuo-spatially index auditory information (i.e. associate an auditory stimulus with a visual stimulus appearing in a specific location) is critical to experiments employing ALP. This ability has been demonstrated in 6-month-old infants; after being presented with different sounds co-occurring with images in different locations, they were able to predict the location when presented with the sound only [7, 8]. Similarly, adults have been shown to visuo-spatially index auditory information [7]. Participants were presented with different facts while faces appeared in one of four locations. When quizzed on the facts, they consistently looked toward the area in which the face had appeared when the related fact was originally presented.

In summary, participants in ALP experiments need to be able to associate a visual stimulus in a certain location with an auditory stimulus, be inclined to look at the visual stimulus they associate with what they hear, and look towards it

even if it does not appear (anticipate its appearance). Having established that both infants and adults seem to fit the profile of an ALP experiment participant, the next methodological issue appears: how much training do participants of different ages require to form audio-visual associations and consistently show anticipatory looking behaviour in response to speech sounds matched with locations on a screen?

The aim of the current study is to investigate the amount of training necessary for adult participants to reliably associate auditory speech stimuli (the syllables /da/ and /ga/) to visual events occurring in arbitrary but consistent areas on a screen. Participants' eye-movements were recorded while they were presented with a series of training trials, and test trials were periodically inserted in order to test for anticipatory looking behaviour.

Although there are obvious and immense differences in cognitive development between adults and infants, the present experiment is expected to provide some indication of a training threshold, below which participants can not be expected to respond to the audio-visual connection with anticipatory looking.

3. METHOD

The experiment was designed for infants, but the participants of the present study were adults. They were exposed to a series of trials in which the presentation of syllables was systematically paired with an event occurring in specific areas of the screen. Periodically, the participants were tested on their ability to associate the two, as measured by their anticipatory looking behaviour.

3.1. Participants

The participants were 16 Swedish-speaking adults (mean age 26.3 yrs, range 19-63 yrs), 5 of which were male and 10 female. All were students at Stockholm University and received a cinema ticket as compensation for participating in the study.

3.2. Stimuli

The stimuli consisted of several short film sequences (each lasting 5 seconds) in which syllables were presented and images appeared in specific locations on the screen.

The images were cartoon-like colourful depictions of animate objects (cats, dogs, etc.), and were created using MS Paint, Adobe Photoshop 7.0 and Adobe Photoshop CS4. A total of 19 images were used in the study, one as an attention-getter and the rest in the training sequences.

The speech material was recorded in an anechoic chamber, using a Brüel & Kjær condenser microphone and pre-amplifier set, and software Adobe Audition 1.5. A female native speaker of Swedish read several instances of the syllables /da/ and /ga/, and two exemplars with relatively matching acoustic properties (as measured by a portable sound level meter and in Praat 5.2.01) were selected as stimuli for the study (Table 1).

Table 1: Acoustic properties of the stimuli syllables.

| | /da/ | /ga/ |
|--|-------|-------|
| <i>Mean intensity (dB_{SPL})</i> | 55.0 | 54.2 |
| <i>Mean f0 (Hz)</i> | 151.7 | 152.4 |
| <i>Duration (s)</i> | 0.573 | 0.561 |
| <i>Occlusion duration (s)</i> | 0.213 | 0.192 |

The images and the speech material were used to create the film sequences in Adobe Premiere Pro CS4. In the training sequences two white boxes rotated on a black background with an attractor image centred on the screen. Prior to syllable onset, the attractor twirled and disappeared. The syllable was then presented, and a reinforcement image twirled into existence in one of the boxes (Figure 1, bottom). The attractor was the same in all sequences, but the image appearing in the boxes varied with each trial. Test sequences were identical to training sequences, except that no image appeared in either box at syllable onset (Figure 1, top). Each film sequence constituted a trial in the experiment.

The experiment consisted of 36 training trials and 10 test trials. First, a test trial was presented, followed by 4 training trials. This pattern was repeated throughout the experiment, up to the 10th test trial. The test trials were presented in fixed positions in the trial series, while all training sequences were randomised. There were 4 different versions of the experiment, balancing for syllable-location combination and test trial order (Table 2).

Figure 1: Trial setup. An attractor image was presented at the beginning of each trial, disappearing after 1.21 s. At 2.02 s a syllable was presented. In training trials, a reinforcement image appeared at 2.12 s (bottom), while in test trials no image appeared (top). Total trial duration was 5 s.

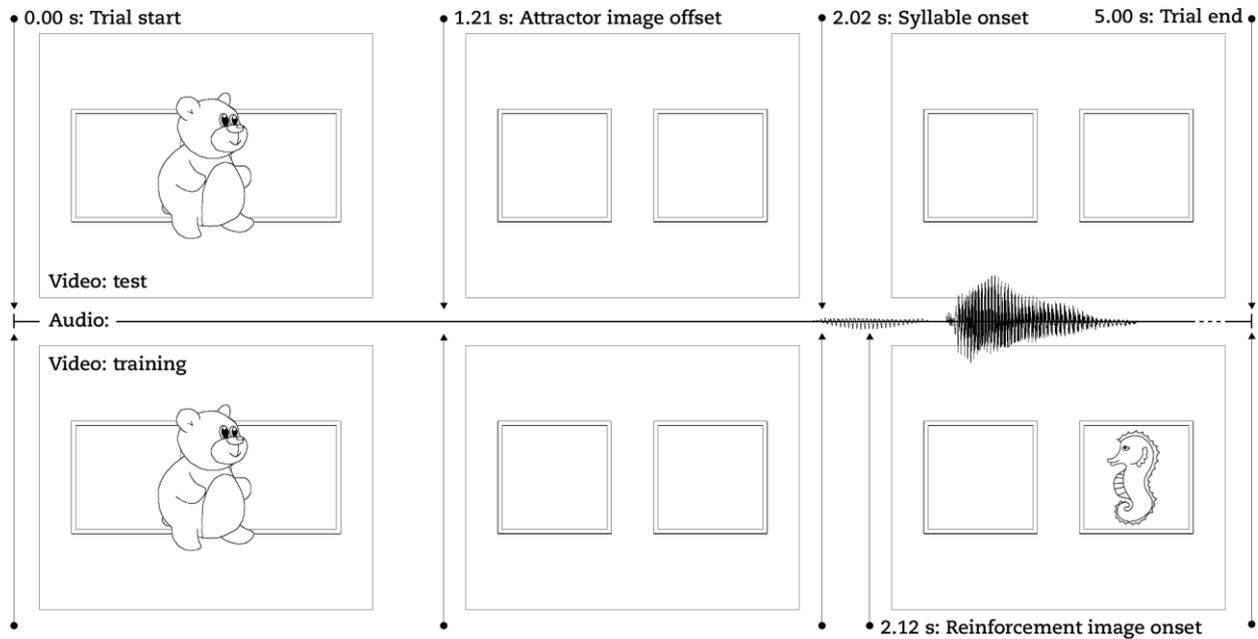


Table 2: Balancing for syllable-box pairing and test order. In versions 1 and 3, the reinforcement image appeared in the left box (L) when the syllable /da/ was presented and in the right box (R) when the syllable /ga/ was presented. Versions 2 and 4 had the opposite pairing. In experiment versions 1 and 2, the syllable in the first test was /da/, while in versions 3 and 4 it was /ga/. The test syllables then alternated throughout the experiment.

| Version | 1 | 2 | 3 | 4 |
|-----------------------------|----------------------|----------------------|----------------------|----------------------|
| <i>Syllable/box pairing</i> | /da/ - L /ga/ - R | /da/ - R /ga/ - L | /da/ - L /ga/ - R | /da/ - R /ga/ - L |
| <i>Test 1</i> | /da/ - L | /da/ - R | /ga/ - R | /ga/ - L |
| <i>Test 2</i> | /ga/ - R | /ga/ - L | /da/ - L | /da/ - R |
| <i>Test 3</i> | /da/ - L | /da/ - R | /ga/ - R | /ga/ - L |
| ... | ... | ... | ... | ... |
| <i>Test 10</i> | /ga/ - R | /ga/ - L | /da/ - L | /da/ - R |

3.3. Procedure

Participants were seated in a sound-attenuated room in front of a Tobii T120 eye-tracking monitor and a set of Creative Inspire T5400 loudspeakers. After a short calibration of the sound level and the eye-tracking system, the experiment was initiated and lasted for approximately 4 minutes. The experimenters controlled the eye-tracking system (using Tobii Studio 2.2.7) and monitored the experiment from an adjacent control room. The participants were given no information about the

study, other than that their eye movements would be tracked by the monitor.

3.4. Data preparation and measurements

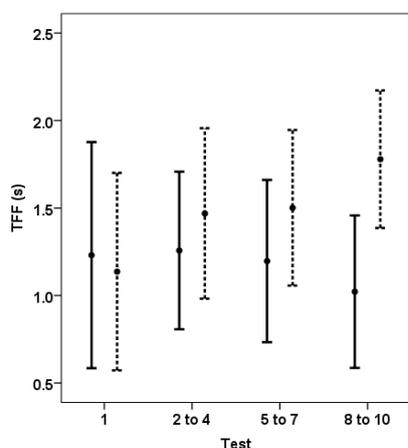
Data preparation (defining time windows for analysis and areas of interest on the screen) was performed in Tobii Studio 2.2.7.

The time to first fixation after syllable onset (TFF) towards the Target or Non-target areas was used to measure anticipatory looking. In particular, the latency of fixations towards the Non-target areas were considered to reveal the subject's expectation that the visual object should have appeared in the Target area. TFF towards the Non-target area is also convenient because TFFs on the Target area have a lower bound (TFF approaching 0), once the subject starts showing correct anticipatory looking behaviour. The first test trial served as baseline, and the remaining 9 were divided into 3 even blocks, averaging the results from 3 tests within each block.

4. RESULTS

A repeated measures ANOVA of the test by test contrasts of the latency in TFF on Non-target area revealed a significant increase in this latency (compared to baseline) for the last block of tests ($F(1,15)=5.840$, $p<0.029$), see Figure 2. All statistical analyses were performed in SPSS Statistics 19.

Figure 2: Error bars of the mean time to first fixation (y-axis, CI=95%) towards Target (solid line) and Non-target (dashed line) after 0, 4-12, 16-24 and 28-36 training trials (x-axis, tests 1, 2-4, 5-7 and 8-10 respectively).



5. DISCUSSION

The present results suggest that adults are able to derive underlying audio-visual associations and respond to them with their looking behaviour after 28-36 training trials.

Several participants who reported not having looked on Target during the tests were shown to have done so by their gaze data, indicating that any visuo-spatial indexing of auditory information that occurred did so implicitly. Furthermore, infants have previously been shown to succeed in grasping relatively simple correspondences where adults looked for more complicated associations and were unable to perform the task [6]. This suggests that infants can be expected to successfully derive the associations as well as show consistent anticipatory looking responses when prompted.

Preliminary results from the ongoing corresponding infant study, show that between 20 and 46 training trials are sufficient. The difference in TFF latency between the baseline and test number 6 (after 20 training trials) is not significant, while the difference between the baseline and test number 10 (after 36 training trials) is significant ($F(1,21)=4.611$, $p<0.044$). The results indicate that infants and adults need similar amounts of training in order to consistently show anticipatory looking behaviour.

6. ACKNOWLEDGEMENTS

The study was funded by the Faculty of Humanities at Stockholm University, and the Bank of Sweden Tercentenary Foundation (K2003-0867). The authors would like to thank Sophie

Gasson and Fredrik Myr for helpful comments on the manuscript, Jenny Ekström and Johan Engdahl for help with data collection, and Anna Ericsson, Hillevi Hägglöf and Klara Marklund for help with stimuli preparation.

7. REFERENCES

- [1] Altmann, G.T.M., Kamide, Y. 2007. The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language* 57, 502-518.
- [2] Cooper, R.M. 1974. The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology* 6, 84-107.
- [3] Huettig, F., Altmann, G.T.M. 2005. Word meaning and the control of eye fixation: semantic competitor effects and the visual world paradigm. *Cognition* 96, B23-B32.
- [4] Johnson, M.H., Posner, M.I., Rothbart, M.K. 1991. Components of visual orienting in early infancy: Contingency learning, anticipatory looking, and disengaging. *Journal of Cognitive Neuroscience* 3, 335-344.
- [5] Kukona, A., Fang, S.Y., Aicher, K.A., Chen, H., Magnuson, J.S. 2011. In press. The time course of anticipatory constraint integration. *Cognition*.
- [6] Mattsson, L. 2009. *Prototype of Infant Hearing Test Using Eye Tracking*. Master of Science, Industrial Engineering and Management, KTH, Stockholm.
- [7] Richardson, D.C., Kirkham, N.Z. 2004. Multimodal events and moving locations: Eye movements of adults and 6-month-olds reveal dynamic spatial indexing. *Journal of Experimental Psychology: General* 133, 46-62.
- [8] Shukla, M., Wen, J., White, K.S., Aslin, R.N. 2011. SMART-T: A system for novel fully automated anticipatory eye-tracking paradigms. *Behavior Research Methods Online First* 1-15.
- [9] Slater, A., Quinn, P.C., Brown, E., Hayes, R. 1999. Intermodal perception at birth: Intersensory redundancy guides newborn infants' learning of arbitrary auditory-visual pairings. *Developmental Science* 2, 333-338.
- [10] Spelke, E. 1976. Infants' intermodal perception of events. *Cognitive Psychology* 8, 553-560.
- [11] Tanenhaus, M.K., Spivey-Knowlton, M.J., Eberhard, K.M., Sedivy, J.C. 1995. Integration of visual and linguistic information in spoken language comprehension. *Science* 268, 1632-1634.