

Eye-Tracking

Michael K. Tanenhaus

University of Rochester, Rochester, New York, USA

Michael J. Spivey-Knowlton

Cornell University, Ithaca, New York, USA

Participants following spoken instructions to touch or move either real objects or objects on a computer screen make saccadic eye movements (to the objects) that are closely time-locked to relevant information in the speech stream. Monitoring eye movements using a head-mounted eye-camera allows one to use the locations and latencies of fixations to examine spoken word recognition during continuous speech in natural contexts. Preliminary work using this paradigm provides striking evidence for the continuous and incremental nature of comprehension, as well as clear effects of visual context on the earliest moments of linguistic processing. We review the eye-movement paradigm and refer to recent experiments applying the paradigm to issues of spoken word recognition (e.g. lexical competitor effects), syntactic processing (e.g. the interaction of referential context and ambiguity resolution), reference resolution (disambiguating temporarily ambiguous referential phrases), focus (modulating the salience of certain objects via contrastive stress), as well as issues in cross-modality integration that are central to evaluating the modularity hypothesis.

Issues Addressed

1. The on-line interaction between spoken language and visual processing.
2. The role of real-world context in spoken language comprehension.
3. Word recognition as a by-product of language comprehension.

Requests for reprints should be addressed to Michael K. Tanenhaus, Department of Brain and Cognitive Sciences, Meliora Hall, River Campus, The University of Rochester, Rochester, NY 14627, USA. E-mail: mtan@psych.rochester.edu

This research was supported in part by NIH Grant No. HD-27206 and National Resource Grant No. P41-RR09282. We wish to acknowledge the central role that our collaborators Kathleen Eberhard and Julie Sedivy played in developing this paradigm with us.

First Uses

Cooper (1974) recorded people's eye movements to drawings, using a fixed-head eye-tracker, while they listened to a story. Tanenhaus, Spivey-Knowlton, Eberhard and Sedivy (1995, 1996) recorded people's eye movements to real objects while they carried out spoken instructions to touch or move the objects.

Description

The subject wears a lightweight helmet with an infrared camera that records eye-in-head position and a colour camera that records the subject's field of view. Gaze position, superimposed (as crosshairs) on the field of view, in addition to the audio track, are recorded on a Hi-8 VCR with frame-accurate rendering, thus allowing frame-by-frame synchronisation (and playback) of audio and video.

Stimuli

Subjects receive instructions ("live" or pre-recorded) to look at, touch or move real objects in a workspace (e.g. a horizontal table or upright board), or pictures of objects on a computer screen (using a mouse).

Dependent Variables

1. Latency of a saccadic eye movement to a particular object after a critical point in the speech stream.
2. Probability of fixating a particular object as a function of time.
3. Sequence of eye movements to objects.

Independent Variables

1. Presence or absence of particular objects in the visual display.
2. Features and/or locations of objects in the visual display.
3. Point in the speech stream where a referring expression becomes unique with respect to the potential set of referents defined by the relevant visual world.
4. Prosodic and acoustic manipulations in the speech stream.
5. Duration of preview of the visual display.
6. Accessibility of objects in the display.

Analysis Issues

1. Analysis of data by video (30 Hz) allows only 33 msec temporal resolution. However, cross-indexing the video frames with the eye-camera's data stream (of 60+ Hz) can provide finer-grain temporal resolution. Newer cameras provide 200 Hz resolution.
2. Saccadic eye movements typically take about 200 msec to program. Therefore, the time of target identification can be approximated at 200 msec before the launch of the saccade.

Effects Found with Paradigm

1. Visually mediated lexical competitor (cohort) effects; that is, the latency and accuracy of eye movements to a target object (e.g. an eye movement to a candle following the instruction, "Pick up the candle") are affected by the presence of an object with overlapping initial phonemes (e.g. candy) and objects that rhyme with the target (e.g. handle).

Shown by: Spivey-Knowlton, Tanenhaus, Eberhard and Sedivy (1995); Allopenna, Magnuson and Tanenhaus (1996).

2. Reference is established incrementally, shortly after the earliest possible point at which the linguistic input provides sufficient information to disambiguate a referent from among the possible set of alternatives.

Shown by: Cooper (1974); Eberhard, Spivey-Knowlton, Sedivy and Tanenhaus (1995); Tanenhaus et al. (1996).

3. Visual context mediates syntactic ambiguity resolution; that is, the on-line syntactic processing of the spoken input is immediately affected by referentially relevant information in the visual input.

Shown by: Tanenhaus et al. (1995).

4. Contrastive stress affects salience of objects in a visual display.

Shown by: Sedivy et al. (1995).

Design Issues

1. Filler instructions and displays must be carefully constructed so as to minimise predictability of conditions and prevent induction of strategies.
2. A central fixation target must be used to control eye position at the start of a trial.
3. At the start of a trial, critical objects in the display should be equidistant from the central fixation target.

Validity

1. Effects of incremental reference assignment (Cooper, 1974; Eberhard et al., 1995) and effects of contrastive focus (Sedivy et al., 1995) have been replicated in multiple experiments.
2. The influence of visual context on lexical competitor effects (Spivey-Knowlton et al., 1995) and on syntactic ambiguity resolution (Tanenhaus et al., 1995) has been replicated with both live and pre-recorded instructions.
3. Work examining the specific nature of lexical competition in spoken word recognition, and details about the uptake of acoustic information in word recognition (e.g. co-articulatory effects and segmentation issues), are in the preliminary stages (cf. Allopenna et al., 1996).

Advantages

1. Spoken language processing can be studied both with real-time precision and with a non-invasive task—that is, a task that does not require interrupting the speech stream or having the subject make an overt decision about the speech itself.
2. This methodology allows word recognition to be studied in a realistic environment where words are embedded in connected, meaningful language.
3. A largely ignored, yet rich, source of information for language comprehension, the real-world visual context, can be studied for its role in on-line spoken language processing, allowing more powerful tests of hypotheses about modularity and information encapsulation.
4. Spoken language processing can be studied under pragmatically natural conditions, allowing one to examine both attentional and intentional effects.
5. The same response measure can be used to study word recognition and higher-level language processing.

Potential Artifacts

Predictability of conditions or induction of unnatural strategies can cause misleading results.

Problems

Only language that makes reference to concrete objects can be used. More work needs to be done to assess task strategies and to assess whether informative results can be obtained without reaching, as in Cooper's (1974)

original studies. If a reaching task is required, this will place limitations on the paradigm. If not, the paradigm can be used to address a broad range of psycholinguistic issues.

Uses with Other Populations

None so far.

Other Comments

Detailed exploration of this methodology has only recently begun (despite the early work of Cooper, 1974). Thus all results are somewhat preliminary and the additional advantages/disadvantages of the methodology are likely to be discovered as more work is conducted. Future work will explore issues such as fine-grained uptake of acoustic information during word recognition, coordination of information in lexical and higher-level language processing, eye movements related to visual memory retrieval, use of a stored mental model versus use of the immediate visual input, the process of perspective taking, conversational interactions, eye movements to objects without an explicit reaching task, and the potential for developmental studies with children, as well as neuropsychological studies with brain-damaged patients.

References

- Allopenna, P.D., Magnuson, J.S., & Tanenhaus, M.K. (1996). Watching spoken language perception: Using eye movements to track lexical access. In G.W. Cottrell (Ed.), *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society*, p. 723. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Cooper, R. (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.
- Eberhard, K.M., Spivey-Knowlton, M.J., Sedivy, J.C., & Tanenhaus, M.K. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research*, 24, 409–436.
- Sedivy, J.C., Tanenhaus, M.K., Spivey-Knowlton, M.J., Eberhard, K.M., & Carlson, G. (1995). Using intonationally-marked presuppositional information in on-line language processing: Evidence from eye-movements to a visual model. In J.D. Moore & J.F. Lehman (Eds), *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society*, pp. 375–380. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Spivey-Knowlton, M.J., Tanenhaus, M.K., Eberhard, K.M., & Sedivy, J.C. (1995). Eye-movements accompanying language and action in a visual context: Evidence against modularity. In J.D. Moore & J.F. Lehman (Eds), *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society*, pp. 25–30. Mahwah, NJ: Lawrence Erlbaum Associates Inc.
- Tanenhaus, M.K., Spivey-Knowlton, M.J., Eberhard, K.M., & Sedivy, J.C. (1995). Integration of visual and linguistic information during spoken language comprehension. *Science*, 268, 1632–1634.

Tanenhaus, M.K., Spivey-Knowlton, M.J., Eberhard, K.M., & Sedivy, J.C. (1996). Using eye movements to study spoken language comprehension: Evidence for visually mediated incremental interpretation. In T. Inui & J. McClelland (Eds), *Attention and performance XVI: Information integration in perception and communication*, pp. 457-478. Cambridge, MA: MIT Press.

Special Issue of
Language and Cognitive Process

**A GUIDE TO SPOKEN WORD
RECOGNITION PARADIGMS**

Edited by

François Grosjean
University of Neuchâtel, Switzerland

Uli H. Frauenfelder
University of Geneva, Switzerland



Psychology Press
a member of the Taylor & Francis group
