Cogn Process (2012) 13 (Suppl 1):S193–S197
DOI 10.1007/s10339-012-0476-6

SHORT REPORT

Processing motion implied in language: eye-movement differences during aspect comprehension

Stephanie Huette · Bodo Winter · Teenie Matlock · Michael Spivey

Published online: 5 August 2012

Abstract Previous research on language comprehension has used the eyes as a window into processing. However, these methods are entirely reliant upon using visual or orthographic stimuli that map onto the linguistic stimuli being used. The potential danger of this method is that the pictures used may not perfectly match the internal aspects of language processing. Thus, a method was developed in which participants listened to stories while wearing a head-mounted eyetracker. Preliminary results demonstrate that this method is uniquely suited to measure responses to stimuli in the absence of visual stimulation.

Keywords Spoken language processing · Eyetracking · Grammatical aspect

Introduction

In recent years, language processing research has begun to pay attention to the incremental nature of speech processing: sounds come in as a time-linear and continuous signal that must be processed continuously, even as new input arrives (Rayner and Clifton 2009). The result of incremental processing is time pressure, as well as the need to be sensitive to speech at a very fine temporal resolution. Because of this, it is extremely crucial that a measurement is sensitive to the time scale of the phenomena. The eyes are constantly moving, and saccades tend to take place before any discrete decisions have been made. Since it is a very low-cost behavior, the eyes will focus on visual stimuli that are even just briefly relevant or mentioned in the speech stream. This is arguably a window into processing.

Many studies have shown that eye movements are closely time-locked to the speech signal (e.g., Altmann and Kamide 1999; Dahan et al. 2001; Eberhard et al. 1995). These studies generally use a visual display and spoken speech files. For example, in Altmann and Kamide (1999), the scene might contain a boy, a toy train, a ball, a toy racecar, and a cake, and participants will hear something such as "The boy will eat the cake". Even without an explicit task, the eye movements of listeners will closely follow the picture that matches the lexical item being accessed, and even make anticipatory fixations. Similarly, Altmann (2004) displayed visual stimuli, and after changing to a blank white screen, participants would hear a speech file. Importantly, eye movements were locked to where the images *have been* on the screen, indicating spatial information was preserved in the meaning of the pictures and was interfacing with the lexical and semantic information. Thus, a visual world need not be present in order for eye movements to be informative about the comprehension of a sentence or a story. Rather, memory can be accessed in order to build and form meaning from percepts and concepts that have been previously processed (see also Spivey and Geng 2001).

Previous research using the visual world paradigm posits that this method detects language processing via an interaction between language and vision at a conceptual level of processing (Salverda et al. 2011). This kind of spoken language processing is particularly action-oriented, often involving a task where a participant uses a computer mouse to actively click on a picture. This engagement of action goals as integral to the comprehension scenario necessarily gives us a view of language processing that

S. Huette (✉) · B. Winter · T. Matlock · M. Spivey
Cognitive and Information Sciences, University of California, Merced, 5200 North Lake Road, Merced, CA 95343, USA
e-mail: shuette@ucmerced.edu

does not easily include instances where one might not have such a contextually constrained motoric action goal, e.g., when passively listening to a lecture or debating a topic with a friend (see Paxton and Dale under review). Perhaps, the most fascinating and puzzling example of this is when we listen to a story. Here, the goal of listening may simply be entertainment, but this in itself does not have a clear goal for an immediate behavioral outcome. The present paradigm is designed to investigate processing during storytelling, where there is no behavioral response required of the participants.

Another possible objection to the visual world paradigm is the pictures, and tasks used may not accurately reflect language processing in the wild. Instead, perhaps, language processing only proceeds in a richly interactive manner (sensitive to subtle contextual manipulations) when immediate visual input is present, and with specific, concrete behavioral goals such as following explicit instructions to click images on a computer screen. One possibility is that language processing is very different when there are no concurrent visual stimuli and no goals driving it. Could it be that language about things that are not visibly present in the environment is processed in a more abstract, symbolic, and non-interactive fashion? Alternatively, perhaps, the processing interactions between visual, motor, and auditory systems that dominate visually contextualized language use are also present during de-contextualized language use (such as idle listening or talking on the telephone). Is language processing still a richly interactive and embodied cognitive process even when it is not necessary to map speech sounds onto visual stimuli or to access meaning for the purpose of performing some explicit cognitive task?

The goal of this study is to investigate subtle grammatical effects without constraining the visual semantics to pictures in front of them. This is the case with many communicative contexts, such as listening to a story, attending a lecture, or listening to conversations between characters in a movie. Previous research has used implicit measures such as heart rate, respiration, galvanic skin response, and pupil dilation as continuous indices of fluctuations in cognitive processing (for examples see: Konvalinka et al. 2011; Mulder and Mulder 1981; Harris et al. 2003; Hyönäa et al. 1995). However, these measures often have substantial temporal delays, and they lack a spatial component. Thus, saccadic eye movements may be uniquely suited to investigating fluctuations in cognitive processing with stimulus conditions that lack a clear behavioral goal.

The predictions here are simply that the grammar of language allows for shades of meaning that modify how we process information, and these modifications may be detectable even without forcing participants to perform tasks that unavoidably direct their attention to these modifications. For example, *I was walking to the store yesterday* implies different expectations about events than *I walked to the store yesterday*. The former sentence (with past progressive, *was* VERB+*ing*) places emphasis on the event of walking, perhaps, also encouraging prediction about the events to take place after the walking as well. The latter sentence (with simple past, VERB+*ed*) places emphasis on the store, highlighting the end point of the event. These two different grammatical forms are known to differentially influence the way we process and report information (Ferretti et al. 2007; Madden and Zwaan 2003; Matlock 2011; Fausey and Matlock 2011).

Because of the constant interaction between visual and linguistic information processing in cognition, we hypothesize that this connection has been learned at a very implicit level within the grammar. That is to say, when grammar places emphasis on the end location of an event (e.g., simple past), attention should be focused on a static location. This will result in a tighter distribution of eye movements. Conversely, when grammar places emphasis on the ongoingness of an event (e.g., past progressive), attention should be oriented toward action and change over time. Visual and motor motion is typically associated with ongoing events, and this attention on change over time may cause a wider distribution of eye movements, even during an undirected goal-free language processing scenario.

Method

Thirty-one right-handed native English speakers (UC Merced undergraduates) participated in the past progressive condition (*was* VERB+*ing*) and thirty-two in the simple past condition (VERB+*ed*). Fixations and saccades were monitored and recorded with an Eyelink II eyetracking system, in accordance with IRB standards. Data were recorded at 500 Hz. To prevent participants from strategizing or developing clear covert goals, they were first given a visual masking task and then told the target block of experimental trials was merely intended to help in forgetting these pictures. They were instructed to remain seated and keep their eyes on the monitor to avoid having to recalibrate the eyetracking equipment. This type of setting is similar to when one is listening to a lecture or a talk, where many of the words being used by the speaker lack an immediate visual referent, and the behavioral goal is too distant or too abstract to be directly tracked or measured.

The stories were the same across the two groups of participants, except for grammatical aspect (*was* VERB+*ing* vs VERB+*ed*), and consisted of three to four sentences each. For example, "Peter was on the road last week. He (drove/was driving) to Florida to see his friend. He (stayed/was staying) in cheap motels and (ate/was eating) in diners. He (used/was using) his mother's credit card." A male native speaker of American English recorded the stories. These sound files were played over headphones during viewing of the blank white screen. There was a slight difference between the total duration of the past progressive stories (186.9 s) and the simple past stories (174.4 s), but the mean difference in duration was not significant ($p > .05$). Data recorded include the fixation location in pixel coordinates along with time stamps of each fixation (x, y, t).

Results

Coordinates outside the range of the screen were discarded before analysis. Data were aggregated over all trials and subjects by condition. To adequately capture the density of fixations in the central region of the screen (where most fixations took place), the probability density function (PDF) was computed using bivariate kernel density estimation with a Gaussian filter. The simple past condition (Fig. 1) shows high kurtosis (i.e., sharp peakedness in the distribution) indicating greater density of fixations in the central region of the screen. The past progressive (Fig. 2) shows a more spread out distribution, indicating people were moving their eyes around more than in the simple past condition. In order to visualize these differences, the bins and counts from histograms of the raw data were computed for each condition. Then each bin's count from the past progressive was subtracted from the simple past condition. This computation gives us an idea and visualization of where the differences in fixations were located (Fig. 3). This figure demonstrates that the difference between the two conditions is mainly in the duration and number of fixations to the central region of the screen.

Discussion

Speech processing was investigated via eye movements in the absence of a visual scene. This study characterized raw spatial information contained in the location of eye fixations as a function of grammar. It was found that when there was an implicit grammatical emphasis on end location (*John went to the grocery store*), eye fixations were located more in roughly the center of the screen, in a tighter distribution than when the emphasis was on the event and the continuous unfolding of motion (*John was going to the grocery store*). This is similar to response measures that gauge general arousal (such as galvanic skin response, heart rate, pupil dilation, etc.), but it is useful for investigating fine-grained spatial differences that unfold in real time.

The extension of this abstracted form of motion processing into the body, seen in this study in the eye movements of participants, has implications for theories of

Fig. 1 Aggregate PDF of eye fixations, simple past condition

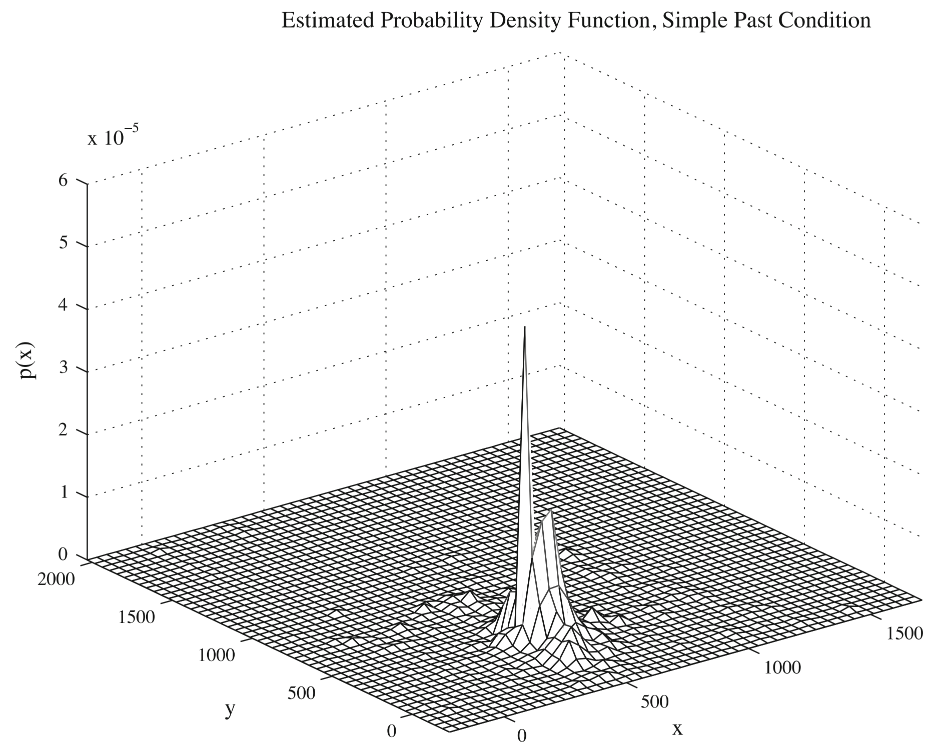
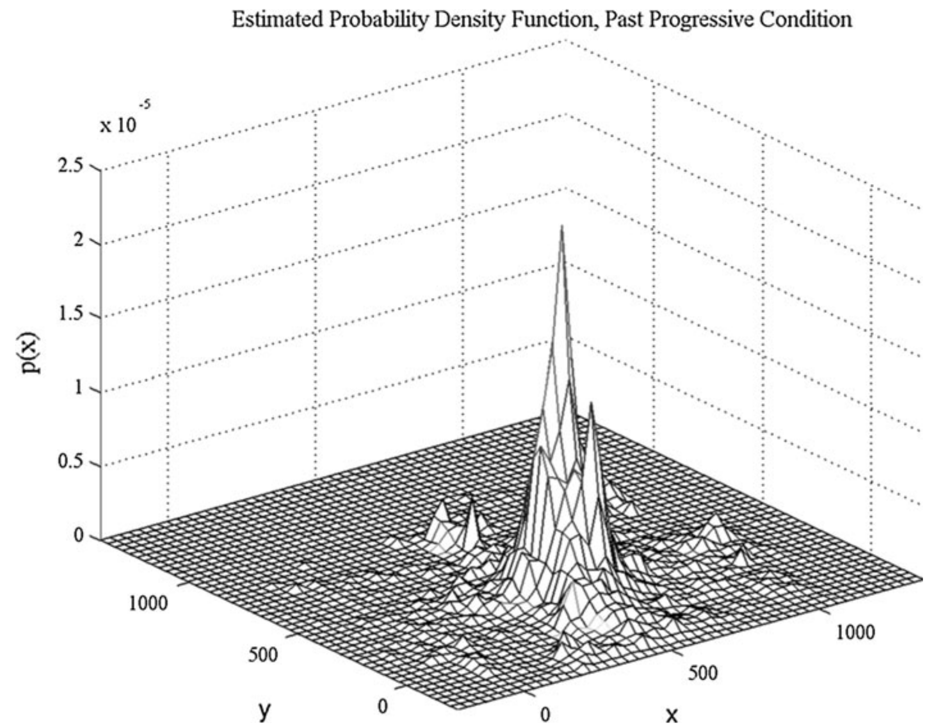


Fig. 2 Aggregate PDF of eye fixations, past progressive condition



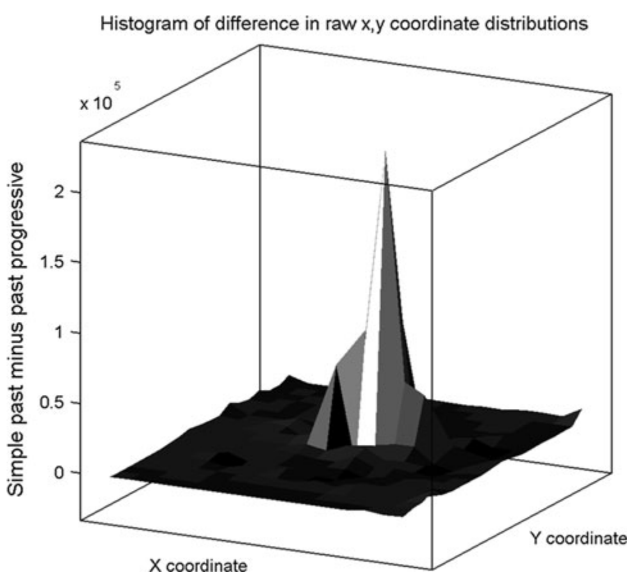


Fig. 3 Difference between conditions—the positive spike in the middle indicates participants in the simple past condition fixated more and for a longer duration on the central portion of the screen

embodiment. This is a first step in showing a connection between more abstract levels of language processing being grounded in body movements. Future studies need to demonstrate not only an influence of language on the body, but also the body influencing the comprehension of these kinds of sentences. This demonstrates an inextricable importance of the body for comprehension processes.

Experimental contexts must elicit a goal-directed response unless the measure is designed to pick up on behavior that occurs constantly. These include measurements such as breathing rate, heart rate, sucking rate (pacifiers with infants), and galvanic skin response (skin conductance). The current method is similar in that the eyes are constantly in motion (just as the heart is always beating and the skin is always slightly conductive), and thus, small deviations in perceptual and cognitive processing easily perturb these systems, making it possible to infer perceptual and cognitive content from these measures. Further, the inherent spatial nature of eye movements allows for many kinds of analyses of the location of fixations and saccades, and their relative timing. These analyses are based on an experimental paradigm that minimizes task demands and more adequately mirrors real-world passive listening situations.

Conflict of interest This supplement was not sponsored by outside commercial interests. It was funded entirely byECONA, Via dei Marsi, 78, 00185 Roma, Italy

References

Altmann G (2004) Language-mediated eye movements in the absence of a visual world: the 'blank screen paradigm'. Cognition 93:B79–B87

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

Dahan D, Magnuson J, Tanenhaus MK (2001) Time course of frequency effects in spoken-word recognition: evidence from eye movements. Cogn Psychol 42:317–367

Eberhard KM, Spivey-Knowlton MJ, Sedivy JC, Tanenhaus MK (1995) Eye movements as a window into real-time spoken language comprehension in natural contexts. J Psycholinguist Res 24:409–436

Fausey CM, Matlock T (2011) Can grammar win elections? Polit Psychol 32:563–574

Ferretti TR, Kutas M, McRae K (2007) Verb aspect and the activation of event knowledge. J Exp Psychol Learn Mem Cogn 33:182–196

Harris CL, Ayçiçegi A, Gleason JB (2003) Taboo words and reprimands elicit greater autonomic reactivity in a first language than in a second language. Appl Psycholinguist 24:561–579

Hyönäa J, Tommola J, Alaja AM (1995) Pupil dilation as a measure of processing load in simultaneous interpretation and other language tasks. Q J Exp Psychol A Hum Exp Psychol 48:598–612

Konvalinka I, Xygalatas D, Bulbulia J, Schjødt U, Jegindø EM, Wallot S, van Orden G, Roepstorff A (2011) Synchronized arousal between performers and related spectators in a fire-walking ritual. Proc Nat Acad Sci 108:8514–8519

Madden CJ, Zwaan RA (2003) How does verb aspect constrain event representations? Mem Cogn 31:663–672

Matlock T (2011) The conceptual motivation of aspect. In: Panther K, Radden G (eds) Motivation in grammar and the lexicon. John Benjamins, Amsterdam, pp 133–147

Mulder G, Mulder LJM (1981) Information processing and cardio-vascular control. Psychophysiology 18:392–402

Paxton A, Dale R (under review). Argument disrupts interpersonal alignment

Rayner K, Clifton C (2009) Language processing in reading and speech perception is fast and incremental: implication for event-related potential research. Biol Psychol 80:4–9

Salverda AP, Brown M, Tanenhaus MK (2011) A goal-based perspective on eye movements in visual-world studies. Acta Psychol 137:172–180

Spivey MJ, Geng J (2001) Oculomotor mechanisms activated by imagery and memory: eye movements to absent objects. Psychol Res 24:235–241