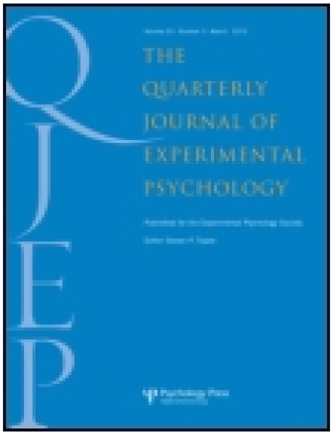
This article was downloaded by: [University of California Davis]
On: 15 May 2015, At: 15:44
Publisher: Routledge
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



The Quarterly Journal of Experimental Psychology
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/pqje20

Pupillometry reveals processing load during spoken language comprehension

Paul E. Engelhardt [a] , Fernanda Ferreira [a] & Elena G. Patsenko [b]
[a] University of Edinburgh , Edinburgh, UK
[b] Michigan State University , East Lansing, MI, USA
Published online: 15 Dec 2009.

To cite this article: Paul E. Engelhardt , Fernanda Ferreira & Elena G. Patsenko (2010) Pupillometry reveals processing load during spoken language comprehension, The Quarterly Journal of Experimental Psychology, 63:4, 639-645, DOI: 10.1080/17470210903469864

To link to this article: http://dx.doi.org/10.1080/17470210903469864

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the “Content”) contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at http://www.tandfonline.com/page/terms-and-conditions

THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY
2010, 63 (4), 639–645

Psychology Press
Taylor & Francis Group

Rapid Communication

Pupillometry reveals processing load during spoken language comprehension

Paul E. Engelhardt and Fernanda Ferreira
University of Edinburgh, Edinburgh, UK

Elena G. Patsenko
Michigan State University, East Lansing, MI, USA

Downloaded by [University of California Davis] at 15:44 15 May 2015

This study investigated processing effort by measuring peoples' pupil diameter as they listened to sentences containing a temporary syntactic ambiguity. In the first experiment, we manipulated prosody. The results showed that when prosodic structure conflicted with syntactic structure, pupil diameter reliably increased. In the second experiment, we manipulated both prosody and visual context. The results showed that when visual context was consistent with the correct interpretation, prosody had very little effect on processing effort. However, when visual context was inconsistent with the correct interpretation, prosody had a large effect on processing effort. The interaction between visual context and prosody shows that visual context has an effect on online processing and that it can modulate the influence of linguistic sources of information, such as prosody. Pupillometry is a sensitive measure of processing effort during spoken language comprehension.

Keywords: Syntactic ambiguity resolution; Prosody; Pupillometry; Visual context; Processing effort

Successful language comprehension requires the quick and accurate integration of multiple sources of information. One of the primary ways researchers have investigated this issue is by examining the processing of sentences with locally competing syntactic analyses. Such sentences, also called garden paths, contain a temporary ambiguity in which a listener is often led into an incorrect interpretation and then later material signals that an error has been made. Experimental manipulations then assess whether a particular information source affects how structural analyses are built and revised. Two information sources are critical in real-world environments: prosody and visual context.

Correspondence should be addressed to Paul E. Engelhardt, Department of Psychology, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ. E-mail: Paul.Engelhardt@ed.ac.uk

This work was supported by National Institutes of Health (NIH) Grant MH65310 and National Science Foundation Integrative Graduate Education and Research Traineeship (NSF IGERT) Grant DGE-0114378. Study protocols were approved by Michigan State University's Human Research Protection Program and conformed to APA ethical standards. The authors would like to thank Kelly Spegel, Emily France, and Timothy Bond for their help creating the stimuli and collecting the data. We would also like to thank Anouschka Bergmann and the Ohio State Linguistics Department for their assistance with the Tones and Break Indices (ToBI) analysis of our stimuli. Portions of this work were presented at the 29th Annual Conference of the Cognitive Science Society.

© 2009 The Experimental Psychology Society

http://www.psypress.com/qjep
DOI:10.1080/17470210903469864

Downloaded by [University of California Davis] at 15:44 15 May 2015

Prosody refers to the stress, timing, and intonation of an utterance and tends to align with syntactic structure (Selkirk, 1984). More specifically, prominent prosodic breaks, in terms of pauses and pitch accents, tend to occur at major syntactic breaks (Ferreira, 1993). These facts have led to a body of work investigating the question of whether prosodic information can be exploited by listeners to resolve syntactic ambiguity (Pynte & Prieur, 1996; Speer, Kjelgaard, & Dobroth, 1996). For example, Kjelgaard and Speer (1999) investigated the processing of early closure ambiguities, such as *When Roger leaves the house is dark*. They found that an appropriately placed intonational phrase boundary (i.e., between *leaves* and *the*) speeded comprehension, suggesting that prosodic boundaries allow a hearer to group syntactic constituents, thereby informing structural decisions (see also Ferreira, 2007; Nagel, Shapiro, Tuller, & Nawy, 1996; Price, Ostendorf, Shattcuk-Hufnagel, & Fong, 1991; Warren, Grabe, & Nolan, 1995).

Visual context refers to visual information that is available as a person processes language. Work investigating the role of visual context on parsing has primarily focused on prepositional phrase attachment ambiguities, such as *Put the apple on the towel in the box*. The key manipulation is whether there are one or two referents (i.e., apples) in the environment. The results from several studies have shown that when the display contains a single apple people will often look to an empty towel, indicating that they initially interpreted *on the towel* as a location (Engelhardt, Bailey, & Ferreira, 2006). However, when there are two apples in the display, participants rarely make eye movements to an empty towel (Spivey, Tanenhaus, Eberhard, & Sedivy, 2002; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). This difference in fixation patterns has been taken as evidence that visual information can influence parsing.

In the current study, we were interested in examining the role of prosody and visual context on the processing effort required to comprehend a garden path sentence. One challenge has been that it is difficult to measure online processing during spoken-language comprehension (cf. Steinhauer, Alter, & Frederici, 1999). However, pupil diameter has been shown to be a reliable physiological measure of processing effort for tasks such as short-term memory, mental arithmetic, and pitch discrimination (Hess & Polt, 1964; Kahenman, 1973), but it has rarely been used as a measure of language processing (cf. Just & Carpenter, 1993; Schluroff, 1982). In this study, we used pupillometry to measure processing effort as participants listened to spoken garden path sentences containing prosodic cues and while examining photographs presented on a computer screen. The main research question that we addressed was whether one source of information or the other would determine processing effort, or whether prosody and visual context would both be integrated online.

EXPERIMENT 1

The first experiment examined prosody. We created two types of prosody for sentences such as Example 1. Sentences with "cooperative" prosody contained a prosodic break between *cleaned* and *the dog*. For Example 1, there was a boundary tone on *cleaned*, and it was followed by a short pause (Beckman & Hirschberg, 1994). Sentences with "conflicting" prosody were spoken as if *the dog* were the object of the verb *cleaned* (i.e., with no break in the pitch and no pause). Two levels of cooperative prosody were created by varying the length of the pause following the intransitive verb; the lengths were 200 ms and 400 ms. After listening to the sentence, participants were required to answer a comprehension question (see Examples 2 and 3). If prosody influences online processing effort and the final interpretation, then we expect an increase in pupil diameter and more incorrect responses to comprehension questions with conflicting prosody than to those with cooperative prosody.

1. While the woman cleaned (#) the dog that was big and brown stood in the yard.
2. Did the woman clean the dog? ("No" question)
3. Did the dog stand in the yard? ("Yes" question)

Downloaded by [University of California Davis] at 15:44 15 May 2015

Method

Participants

A total of 18 native speakers of English with normal or corrected-to-normal vision were recruited from the Michigan State University subject pool.

Materials and apparatus

Stimulus materials consisted of 30 sentences containing optionally transitive subordinate verbs (Christianson, Hollingworth, Halliwell, & Ferreira, 2001). The sentences were recorded at a normal speaking rate by a female native speaker of English. For each sentence, two separate versions were recorded as described above. These utterances were placed into six lists that were counterbalanced with question type. Lists were rotated in a Latin square design, so that each participant heard each sentence only once.

Pupil diameter was monitored with an Eyelink II at 500 Hz. Stimulus presentation was programmed using SR research Experiment Builder software. The eye tracker and a 19″ CRT display monitor (refresh rate of 140 Hz) were interfaced with a 3-GHz Pentium 4 PC. Pupil data were analysed in a 1.2-s time window beginning 200 ms past the onset of the disambiguating word (i.e., *stood*). The 200-ms offset was to allow some time for word recognition. The rationale for examining a 1.2-s time window was based on a previous reading study by Just and Carpenter (1993). They observed peak pupil response 1.2 s following the location in an ambiguous sentence where processing demand increased.

Blinks were filtered, and missing data were replaced using linear interpolation. Pupil diameter was analysed only for trials in which the comprehension question was answered correctly. These correct trials were then averaged together, which resulted in six vectors for each participant. The data were then normalized by calculating the grand mean of the six vectors and then dividing each individual data point by the grand mean. A simple regression, with time as an independent variable and pupil diameter as a dependent variable, was used to determine the slope of the change in pupil size for each participant in each condition.

Design and procedure

The design was 2×3 (question type×prosody type) repeated measures. Questions queried either the main clause or the garden path misinterpretation (e.g., that the woman cleaned the dog). Prosody was conflicting, cooperative (200 ms), or cooperative (400 ms). Conflicting prosody sentences did not have a break between the intransitive verb and the ambiguous noun phrase (see "#" in Example 1). Cooperative prosody was created by placing a prosodic break between clauses.

Participants were instructed that they would hear a sentence, and afterwards they would have to answer a comprehension question about the sentence. Their task was to look at a fixation cross during the sentence and then to answer the comprehension question as quickly and accurately as possible. Participants completed 3 practice trials, 30 experimental trials, and 70 filler trials. Fillers included 25 main subordinate sentences, 25 coordination structures, and 20 sentences involving ellipsis. The entire session lasted approximately 40 min. Analyses were conducted with subjects (F_1) and items (F_2) as random factors.

Results

Prior to the analysis, fixation location was checked to ensure that participants were viewing the fixation cross as they listened to the sentences.[1] Comprehension accuracy showed that both main effects were significant: question type, $F_1(1, 17) = 4.34$, $p = .05$; $F_2(1, 29) = 5.49$, $p < .05$, and prosody type, $F_1(2, 34) = 5.01$, $p < .05$; $F_2(2, 29) = 6.37$, $p < .05$ (see Figure 1). The interaction was also significant, $F_1(2, 34) = 4.75$, $p < .05$; $F_2(2, 58) = 3.95$, $p < .05$. Simple effects showed that the conflicting prosody/no question condition was significantly worse than the conflicting prosody/yes question condition, $t(17) = -2.83$,

[1]For both experiments, we screened the data for outliers. Means of more than ± 3.5 standard deviations were replaced by the mean value for that condition. This affected <2% of the pupil data. There were no outliers in comprehension accuracy.

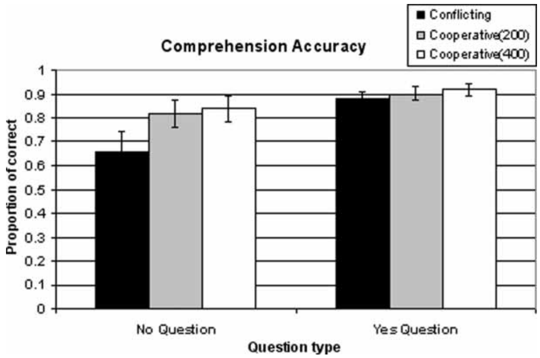


Figure 1. *Comprehension accuracy showing the proportion of trials answered correctly in Experiment 1. Error bars show the standard error of the mean.*

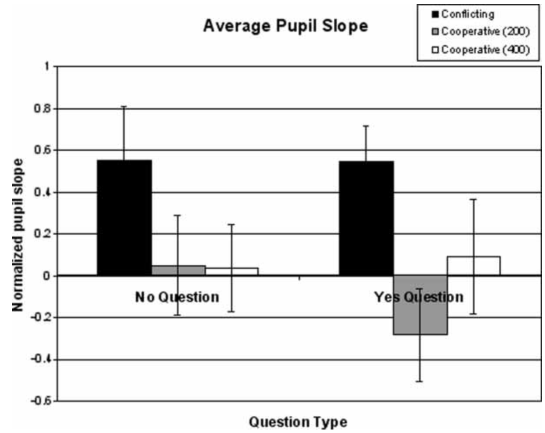


Figure 2. *Pupil slope in a 1.2-s time window beginning 200 ms past the onset of the disambiguating word (e.g.,* stood*) in Experiment 1. Error bars show the standard error of the mean.*

$p < .05$. Neither of the paired comparisons with cooperative prosody was significant. These results show that conflicting prosody leads to a much higher likelihood of getting the garden path misinterpretation.

Pupil slope was analysed with a 2×3 (Question Type×Prosody Type) repeated measures analysis of variance (ANOVA). The results showed only a significant main effect of prosody, $F_1(2, 34) = 4.21$, $p < .05$; $F_2(2, 58) = 5.35$, $p < .05$ (see Figure 2).[2] Simple effects tests revealed that conflicting prosody had a significantly greater slope than both the 200-ms, $t(17) = 2.65$, $p < .05$, and 400-ms, $t(17) = 2.54$, $p < .05$, cooperative prosody conditions. The two cooperative prosody conditions were not different from one another ($p > .05$). These results show that when syntactic structure and prosodic structure conflict, pupil diameter reliably increases, whereas when the prosody and syntax align, pupil slope is flat or slightly negative.

EXPERIMENT 2

The second experiment examined both prosody and visual context. The same sound files were used, so there were again three levels of prosody. Visual context had two levels: It was either consistent or inconsistent with the ultimately correct interpretation. Inconsistent visual context depicted the garden path misinterpretation, which, for Example 1, was a picture of a woman cleaning a dog. Consistent visual context depicted a woman cleaning something that was not a dog (e.g., a window). The main goal of this study was to determine how prosody and visual context affect processing load and the final meaning obtained for spoken garden path sentences.

Method

Participants

A total of 18 native speakers of English with normal or corrected-to-normal vision were recruited from the Michigan State University subject pool.

Materials and apparatus

The two visual contexts for each sentence were selected to be similar in luminance, and the average picture size was kept relatively small (~5 cm×~7 cm) so the majority of the computer screen was white. This minimized luminance differences both within and between pictures. Pictures appeared in the centre of the computer

[2]Question type did not produce a main effect, and it did not interact with prosody type ($p > .05$). This is expected since the comprehension question occurs after the sentence is over and therefore cannot affect the pupil response during the sentence.

Downloaded by [University of California Davis] at 15:44 15 May 2015

screen simultaneously with sentence onset. For the filler trials, half of the pictures were related to the sentence, and half were not. The apparatus and analysis procedures were the same as those in Experiment 1.

Design and procedure

The design was a 2 × 3 (Visual Context × Prosody Type) repeated measures. Visual context was either consistent or inconsistent. Prosody was either conflicting, or 200-ms or 400-ms cooperative. For this experiment, participants only saw the "No" question (see Example 2). Participants were instructed that they would be presented with a spoken sentence and a picture, which may or may not be related to the sentence. Their task was to answer the comprehension question quickly and accurately based on the information given in the sentence. Each session consisted of 3 practice trials, 30 experimental trials, and 70 filler trials. The entire experimental session lasted ~40 min. Analyses were conducted with subjects (F_1) and items (F_2) as random factors.

Results

Prior to the analysis, fixation location was checked to ensure that participants were viewing the picture as they listened to the sentence. Comprehension accuracy showed only a main effect of prosody, $F_1(2, 34) = 13.41$, $p < .01$; $F_2(2, 58) = 23.72$, $p < .01$ (see Figure 3). Therefore, we collapsed across visual context, and paired comparisons showed that the conflicting condition was significantly worse than both the 200-ms, $t(17) = 3.83$, $p < .01$, and 400-ms, $t(17) = 4.31$, $p < .01$. cooperative prosody conditions. These results show, similar to the first experiment, that conflicting prosody results in significantly more garden path misinterpretations.

Pupil slope showed a main effect of prosody, $F_1(2, 34) = 3.81$, $p < .05$; $F_2(2, 58) = 8.94$, $p < .05$, and a significant interaction, $F_1(2, 34) = 3.19$, $p = .05$; $F_2(2, 58) = 3.23$, $p < .05$ (see Figure 4). This interaction was driven primarily by differences with the inconsistent pictures. The 200-ms cooperative prosody condition did not

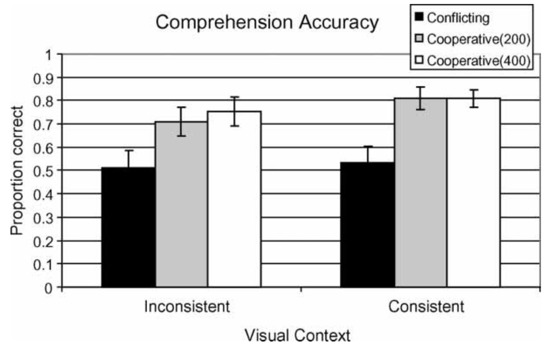


Figure 3. *Comprehension accuracy showing the proportion of trials answered correctly in Experiment 2. Error bars show the standard error of the mean.*

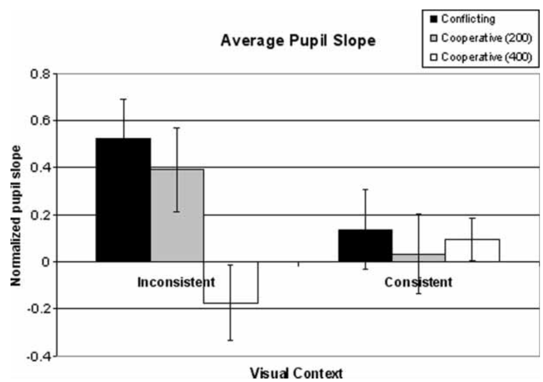


Figure 4. *Pupil slope in a 1.2-s time window beginning 200 ms past the onset of the disambiguating word (e.g.,* stood*) in Experiment 2. Error bars show the standard error of the mean.*

differ from the conflicting prosody condition ($p > .05$). However, both differed from 400-ms cooperative prosody, conflicting, $t(17) = 3.48$, $p < .05$, and 200-ms, $t(17) = 2.18$, $p < .05$. In contrast, there were no differences between the three levels of prosody with the consistent picture. These results show that the visual context affects the online processing effort.

DISCUSSION

The first experiment showed that conflicting prosody resulted in a significant increase in pupil

Downloaded by [University of California Davis] at 15:44 15 May 2015

Downloaded by [University of California Davis] at 15:44 15 May 2015

diameter compared to the two cooperative prosody conditions. Similarly, comprehension accuracy revealed that participants were more likely to get the garden path misinterpretation with conflicting prosody than with cooperative prosody. The second experiment showed that comprehension accuracy was affected only by prosody; misinterpretations were more frequent when the prosody suggested the wrong analysis. In contrast, the pupil measure revealed that the ease with which correct interpretations were computed varies depending on both prosody and visual context. When the visual context was consistent with the sentence's meaning, prosody had little effect on processing effort. However, when the visual context contradicted the sentence's meaning, prosody had a large effect. Moreover, not only does conflicting prosody cause processing load to increase, but cooperative prosody has a graded effect: Processing effort with the 200-ms pause was higher than that with the 400-ms pause. The graded effect of cooperative prosody was not observed in comprehension accuracy and was different from the pattern observed in the first experiment.

Therefore, one of the key results from this study was the dissociation between online and offline performance measures when the 200-ms cooperative prosody was paired with the inconsistent visual context. Pupil diameter indicated a high processing load, but the increase in processing effort did not result in more ultimately incorrect interpretations. One possible explanation is that the inconsistent visual context made the temporarily incorrect interpretation more salient as the sentence was heard, but offline the increase in processing difficulty resulted in only slightly worse comprehension performance (i.e., 71% vs. 74% correct). To further examine the competition between information sources in this condition, we examined a histogram plot of the pupil data to determine whether there was evidence of a bimodal distribution (see Figure 5). That is, were there some trials in which the prosody resoundingly prevented the garden path and other trials that showed more competition between the visual context and the prosody. However, there was no evidence of bimodality in any of the

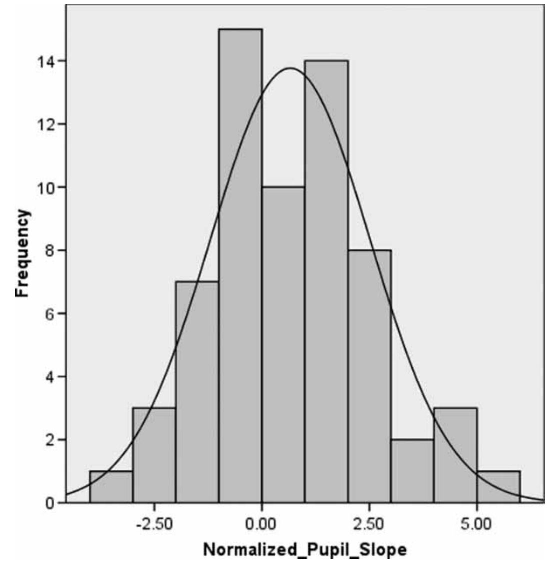


Figure 5. *Frequency distribution of the pupil slope data in the 200-ms cooperative prosody condition with inconsistent visual context.*

six conditions, which suggests a more uniform and graded competition between information sources.

This work is noteworthy for two additional reasons. The first is that this study used pupillometry as a measure of processing effort for spoken garden path sentences. This has not been done previously. Second, this study revealed that it is possible to measure ambiguity resolution in spoken language comprehension in a graded fashion. That is, pupil diameter is not a discrete measure like comprehension accuracy or saccadic eye movements (see also Farmer, Cargill, Hindy, Dale, & Spivey, 2007, for graded measures of language processing using mouse tracking). Therefore, the online measure reported here provides unique and more fine-grained information about online processing effort/difficulty.

This study shows that visual context does affect online processing effort during syntactic ambiguity resolution. The interaction of visual and prosodic information in Experiment 2, which was not observed in comprehension accuracy, indicates that pupil size is a more sensitive measure of language processing. Pupillometry reveals that visual context has a major effect on processing load during comprehension, and indeed it seems to modulate the influence of prosody.

Original manuscript received 23 May 2009
Accepted revision received 13 August 2009
First published online 15 December 2009

REFERENCES

Beckman, M. E., & Hirschberg, J. (1994). *The ToBI annotation conventions*. Columbus, OH: Ohio State University.

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, *42,* 368–407.

Engelhardt, P. E., Bailey, K. G. D., & Ferreira, F. (2006). Do speakers and listeners observe the Gricean maxim of quantity. *Journal of Memory and Language*, *54,* 554–573.

Farmer, T. A., Cargill, S. A., Hindy, N. C., Dale, R., & Spivey, M. J. (2007). Tracking the continuity of language comprehension: Computer mouse trajectories suggest parallel syntactic processing. *Cognitive Science*, *31,* 889–909.

Ferreira, F. (1993). The creation of prosody during sentence processing. *Psychological Review*, *100,* 233–253.

Ferreira, F. (2007). Prosody and performance in language production. *Language and Cognitive Processes*, *22,* 1151–1177.

Hess, E. H., & Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem solving. *Science*, *143,* 1190–1192.

Just, M. A., & Carpenter, P. A. (1993). The intensity dimension of thought: Pupillometric indices of sentence processing. *Canadian Journal of Experimental Psychology*, *47,* 310–339.

Kahenman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.

Kjelgaard, M. M., & Speer, S. R. (1999). Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language*, *40,* 153–194.

Nagel, N. H., Shapiro, L. P., Tuller, B., & Nawy, R. (1996). Prosodic influences on the resolution of temporary ambiguity during on-line sentence processing. *Journal of Psycholinguistic Research*, *25,* 319–344.

Price, P. J., Ostendorf, M., Shattuck-Hufnagel, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America*, *90*(6), 2956–2970.

Pynte, J., & Prieur, B. (1996). Prosodic breaks and attachment decisions in sentence parsing. *Language and Cognitive Processes*, *11,* 165–191.

Schluroff, M. (1982). Pupil responses to grammatical complexity of sentences. *Brain and Language*, *27,* 322–344.

Selkirk, E. O. (1984). *Phonology and syntax: The relation between sound and structure*. Cambridge, MA: MIT Press.

Speer, S. R., Kjelgaard, M. M., & Dobroth, K. M. (1996). The influence of prosodic structure on the resolution of temporary syntactic closure ambiguities. *Journal of Psycholinguistic Research*, *25,* 249–270.

Spivey, M. J., Tanenhaus, M. K., Eberhard, K. M., & Sedivy, J. C. (2002). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology*, *45,* 447–481.

Steinhaur, K., Alter, K., & Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience*, *2,* 191–196.

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.

Warren, P., Grabe, E., & Nolan, F. (1995). Prosody, phonology, and parsing in closure ambiguities. *Language and Cognitive Processes*, *10,* 457–486.

Downloaded by [University of California Davis] at 15:44 15 May 2015