Title
Pupillometric Indices of Visual and Prosodic Information on Spoken Language Comprehension

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 29(29)

ISSN
1069-7977

Authors
Engelhardt, Paul E.
Patsenko, Elena G.
Ferreira, Fernanda

Publication Date
2007

Peer reviewed

Pupillometric Indices of Visual and Prosodic Information on Spoken Language Comprehension

Paul E. Engelhardt (paul@eyelab.msu.edu)
Elena G. Patsenko (patsenka@msu.edu)
Department of Psychology, Michigan State University
East Lansing MI 48824-1116 USA

Fernanda Ferreira (fernanda.ferreira@ed.ac.uk)
Department of Psychology, 7 George Square
Edinburgh, UK EH3 6NN

Abstract

Two experiments investigated the effects of prosody and visual context on syntactic ambiguity resolution. The results from the first experiment showed increased processing difficulty when prosodic structure conflicted with syntactic structure, consistent with previous reports. In the second, both prosody and visual context were manipulated. The results showed that conflicting visual context cancelled the effects of cooperative prosody, suggesting that visual context has a greater effect on ambiguity resolution than does prosody. These results are consistent with models of sentence processing, which assume that both top-down and bottom-up information can affect online processing.

Keywords: syntactic ambiguity resolution; sentence processing; prosody; pupil diameter; visual context.

Introduction

The integration of different sources of information continues to be a central topic in the field of natural language comprehension. One of the primary ways that researchers have addressed this question is by examining sentences that contain locally competing syntactic analyses. Many of the seminal studies in this area were conducted using written materials and the dependent measure was fixation location and duration. In the current experiments, we examined the effects of prosody and visual context on the processing of spoken garden-path sentences. Garden-path sentences contain a temporary syntactic ambiguity in which a listener is led into an incorrect interpretation and then later material signals that an error has been made. Fox Tree and Meijer (2000) looked at effects of prosody and discourse-level linguistic context on the final interpretation of ambiguous sentences. Their results indicated that discourse context had a greater effect on ambiguity resolution than did prosody. The goal of the current experiments was similar, except that we examined the online processing of both prosody and visual context in order to more accurately specify the time course in which these information sources affect ambiguity resolution.

Prosody is the stress, timing, and intonation of an utterance, and is related to syntactic structure (Selkirk, 1984). More specifically, prominent prosodic breaks tend to occur at major syntactic breaks (Ferreira, 1993).[1] This fact naturally leads to the prediction that prosody should influence initial syntactic parsing, and the results from several studies confirm this prediction (Kjelgaard & Speer, 1999; Nagel, Shapiro, Tuller, & Nawy, 1996; Price, Ostendorf, Shattuck-Hufnagel, & Fong, 1991; Speer, Kjelgaard, & Dobroth, 1996). However, other studies have shown little or no effect of prosody (Allbritton et al., 1996). The data are mixed but the bulk of the evidence indicates that prosodic boundaries should allow a hearer to perceptually group syntactic constituents, thereby informing immediate parsing decisions.

The main issue addressed in Experiment 1 was the time course in which prosody influences the processing of a garden-path sentence. Restricted processing models assume that initial parsing is based on a limited amount of information. Speer and colleagues (1996) argued that prosodic structure determines initial syntactic constituency (see also Frazier, 1987). Marcus and Hindle (1990) in contrast, argued that prosody affects parsing only after a syntactically-based initial parse.

Unrestricted or interactive processing models assume that any type of information can immediately influence processing (Spivey & Tanenhaus, 1998). Therefore, prosody or visual context which either supports or refutes a garden-path interpretation should affect initial syntactic decisions. Support for such models has come from referential context effects, primarily involving the presuppositions on the felicitous use of modified noun phrases (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). The general finding is that participants are more likely to pursue a modifier interpretation of an ambiguous prepositional phrase when the visual context contains multiple referents.

[1] The most reliable prosodic cues to syntactic structure are amplitude and durational cues (Ferreira, 1993). The most effective use of these cues is tasks where two participants interact freely (Kraljic & Brennan, 2005; Schafer, Speer, Warren, & White, 2000).

In sum, restricted processing models assume that the initial parse should be impervious to top-down information. Unrestricted models assume that any type of relevant information can be used by the sentence processor. The strongest evidence for these assumptions is an effect of linguistic discourse or visual context biasing the parser to a more complex interpretation. However, recent pragmatically-based arguments have questioned the interpretation of these data (Engelhardt, Bailey, & Ferreira, 2006). Therefore, the primary purpose of the current study was to further investigate these theoretical controversies in the sentence processing literature.

Experiment 1

In the first experiment prosody either conflicted with or was cooperative with respect to the correct syntactic structure (see example 1). Sentences with cooperative prosody were spoken with an intermediate prosodic break between *cleaned* and *the dog* (see Figure 1). Notice that in Figures 1 and 2 the highlighted area shows the intransitive verb (i.e. *cleaned*). Sentences with conflicting prosody were spoken as if *the dog* were the direct object of the verb *cleaned* (see Figure 2). If prosody influences comprehension, then we expect more incorrect responses to questions like example 2 with conflicting prosody compared with cooperative prosody. The dependent measures were comprehension accuracy and pupil diameter.

(1) While the woman cleaned the dog that was big and brown stood in the yard.
(2) Did the woman clean the dog? (subordinate)
(3) Did the dog stand in the yard? (matrix)

Pupil diameter has been shown to be a reliable index of processing intensity during syntactic ambiguity resolution. Just and Carpenter (1993) investigated the processing of subject and object relative sentences. Their results showed a reliable increase in pupil diameter peaking 1.2 sec following the location in a sentence where processing demands increased. The purpose of the first experiment was to determine if an appropriately placed intermediate prosodic break would influence the processing and interpretation of sentences like example 1.

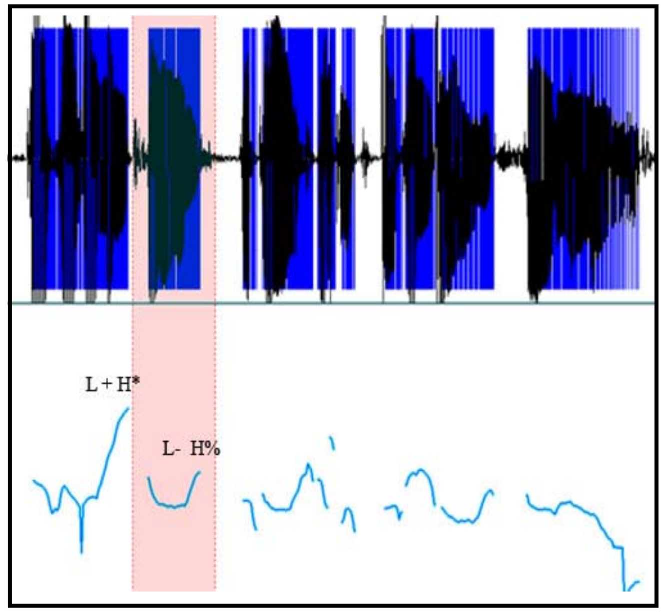

Figure 1: Wave form for cooperative (200) prosody.

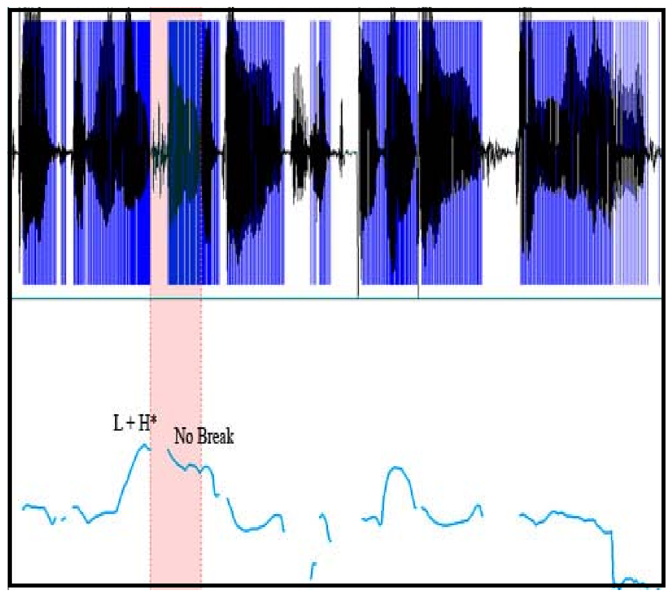

Figure 2: Wave form for conflicting prosody.

Methods

Participants

Eighteen participants from the Michigan State University subject pool were recruited to participate. All participants were native speakers of English.

Materials and Apparatus

Stimulus materials consisted of thirty sentences containing optionally transitive subordinate verbs (Christianson, Hollingworth, Halliwell, & Ferreira, 2001).

Pupil diameter was monitored with an Eyelink II at 500 Hz. Pupil data were analyzed in a 1200 ms window beginning 200 ms past the onset of the disambiguating word to allow for word recognition to have taken place. Blinks were filtered and missing data were replaced using linear interpolation. Pupil diameter was analyzed for correct trials only[2], and a simple regression using time as an independent variable and pupil diameter as a dependent variable was

[2]Correct trials indicate that participants fully reanalyzed the sentence, indicating they got the correct interpretation.

used to determine pupil diameter slope for each participant in each condition.

Design and Procedure

The design was 2 (Question type) x 3 (Prosody type). Questions queried either the subordinate or matrix clause. Prosody was either conflicting or cooperative. Conflicting prosody sentences had no break between the ambiguous noun phrase and the intransitive verb. Cooperative prosody was created in a separate utterance by placing an intermediate prosodic break between clauses. Two levels of Cooperative prosody were created by varying the length of the pause following the intransitive verb, lengths were 200 or 400 ms. Participants were instructed that they would hear a sentence followed by a comprehension question. Their task was to look at a fixation cross during the sentence, and then to answer the comprehension question. Participants completed three practice trials and 100 regular session trials. Fillers included 25 main-subordinate sentences, and the entire session lasted ~ 40 min. Analyses were conducted with subjects (F1) and items (F2) 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. Comprehension accuracy showed that both main effects were significant [Question $F1(1, 17) = 4.34$, $p = .053$; $F2(1, 29) = 5.49$, $p < .05$ & Prosody $F1(2, 34) = 5.01$, $p < .05$; $F2(2, 29) = 6.37$, $p < .05$], and so was the interaction $F1(2, 34) = 4.75$, $p < .05$; $F2(2, 58) = 3.95$, $p < .05$ (see Figure 3). Simple effects showed that the conflicting prosody/subordinate question condition was significantly worse $t(17) = -2.83$, $p < .05$ than the conflicting prosody/matrix question condition. Neither of the paired comparisons with the cooperative prosody were significant. These results show that conflicting prosody leads to more incorrect garden-path interpretations.

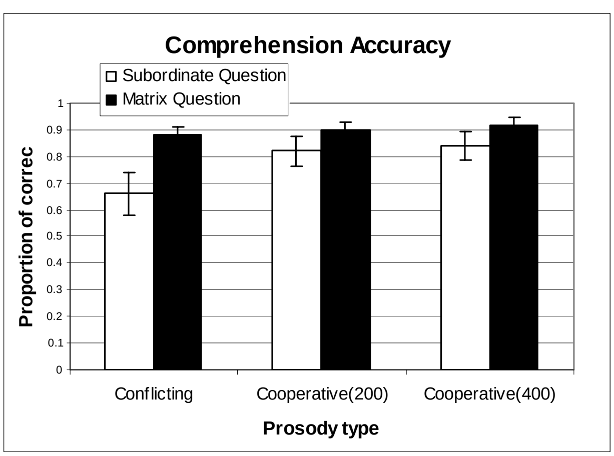

Figure 3: Comprehension accuracy for Experiment 1.

Pupil diameter slope was analyzed as a one-way ANOVA with three levels. The results showed a significant main effect of prosody $F1(2, 34) = 4.21$, $p < .05$; $F2(2, 58) = 5.35$, $p < .05$ (see Figure 4). Simple effects tests revealed that the conflicting prosody condition had a significantly greater slope than both of the cooperative prosody conditions $t(17) = 2.54$, $p < .05$ and $t(17) = 2.65$, $p < .05$. The cooperative prosody conditions were not different from one another. These results show that when syntactic structure and prosodic structure conflict, pupil diameter reliably increases, whereas when the prosody and syntax coincide, pupil slope is negative.

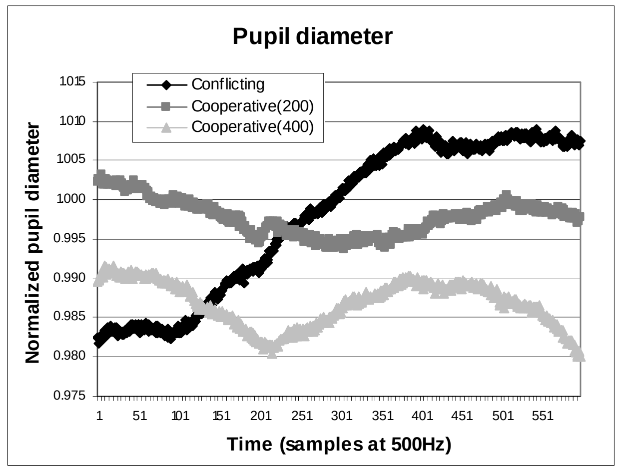

Figure 4: Pupil response beginning 200ms past onset of disambiguating word (i.e. *stood*, in example 1).

Returning to the models reviewed above, we can conclude that conflicting prosody leads to processing difficulty beginning shortly after the onset of the disambiguating word compared with both of the cooperative prosody conditions. This finding supports the assumption that cooperative prosodic structure can be utilized by the sentence processor to inform syntactic decisions. This result is consistent with models of sentence processing in which bottom-up information can affect initial syntactic parsing (Frazier, 1987; Speer et al., 1996; Spivey & Tanenhaus, 1998).

Experiment 2

In the second experiment, we manipulated both bottom-up and top-down information sources. The same sound files from Experiment 1 were used, so there were three levels of prosody. Visual context had two levels. Inconsistent visual context depicted the misinterpretation. For example (1), inconsistent visual context showed a picture of a woman cleaning a dog. Consistent visual context depicted a woman cleaning something that was not a dog. If visual information can influence parsing as predicted by unrestricted processing models, then we should see an effect of visual context on pupil diameter. However, if initial parsing is based strictly on bottom-up information, then we should find only a main effect of prosody and no effect of visual context.

Methods

Participants

Eighteen participants from the Michigan State University subject pool were recruited to participate. All were native speakers of English.

Materials

The materials were the same as Experiment 1, except that pictures matching two interpretations of the sentence were created.[3]

Design and Procedure

The design was 2 (Visual context) x 3 (Prosody type). Visual context depicted either the misinterpretation or a consistent interpretation. Prosody was either Conflicting, Cooperative (200), or Cooperative (400). Pictures appeared in the center of the computer screen at the same time as the sentence began. The same procedures were used as in Experiment 1, except participants were told that they would be presented with a spoken sentence and a picture which may or may not be consistent with the sentence. Their task was to answer the comprehension question on the basis of the information presented in the sentence.

Results

Fixation location was again examined prior to statistical analyses in order to ensure that participants were fixating the pictures as they listened to the sentences. Comprehension accuracy showed a significant main effect of Prosody $F1(2, 34) = 13.41$, $p < .01$; $F2(2, 58) = 23.72$, $p < .01$, and there was a marginal effect of Visual context $F1(1, 17) = 3.56$, $p = .08$; $F2(1, 29) = 4.17$, $p = .05$. The interaction was not significant (see Figure 5). The simple effects showed only one marginal effect of visual context. The 200 ms Cooperative prosody condition with inconsistent visual context resulted in slightly more misinterpretations than did consistent visual context. Prosody showed a robust main effect. Visual context, in contrast showed only a marginal effect, and here the results suggested that inconsistent visual context results in slightly worse performance across the board.

Pupil slope showed a main effect of Prosody type $F1(2, 34) = 3.81$, $p < .05$; $F2(2, 58) = 7.96$, $p < .05$, and a significant interaction $F1(2, 34) = 3.19$, $p = .05$; $F2(2, 58) = 2.52$, $p = .089$. The main effect of Visual context was not significant. Examining the results in Figures 6, 7, and 8 reveals that the interaction is largely driven by performance with inconsistent Visual context. Here the 200ms Cooperative prosody was not different from the Conflicting prosody $t(17) = .54$, $p > .05$, but both were significantly different from the 400ms Cooperative prosody [200ms $t(17) = 3.48$, $p < .05$ & Conflicting $t(17) = 2.18$, $p < .05$]. This suggests that inconsistent Visual context overrides the disambiguating effect of 200ms Cooperative prosody. Another interesting result was that consistent Visual context seemed to reduce processing effort with the Conflicting prosody, as evidenced by the flattening out of the pupil slope, suggesting that reanalysis is completed ~600 ms after disambiguation. Pupil diameter indicates greater processing effort with inconsistent Visual context in both the Conflicting and 200ms Cooperative prosody conditions. These data suggest that visual information influences processing with both Conflicting and Cooperative prosody.

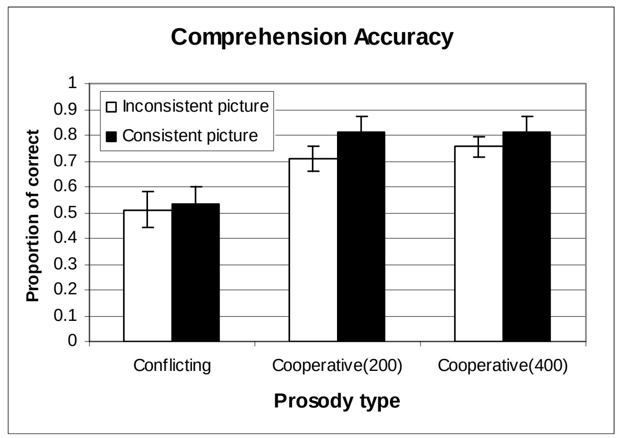

Figure 5: Comprehension results for Experiment 2.

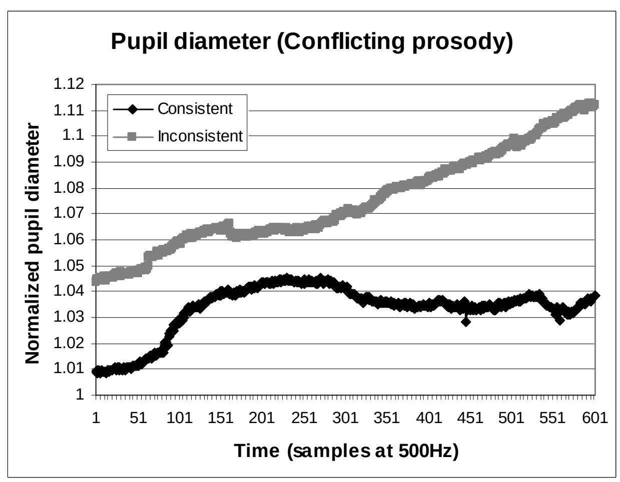

Figure 6: Pupil response beginning 200ms past onset of disambiguating word with conflicting prosody.

[3] Pictures were selected to be similar in luminance, and the average picture size was relatively small (5cm x 7cm), so the majority (~75%) of the computer screen was white.

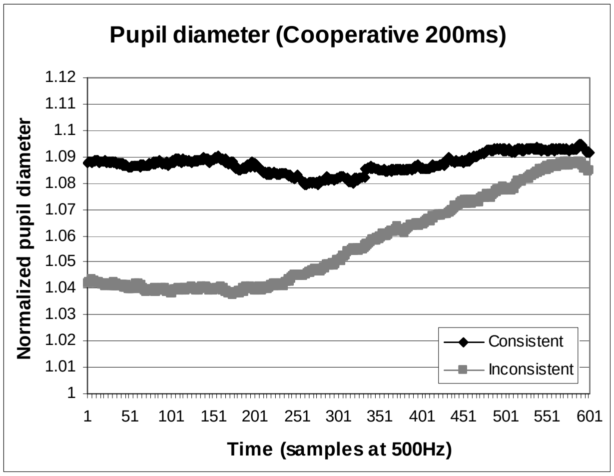

Figure 7: Pupil response beginning 200ms past onset of disambiguating word with cooperative 200 prosody.

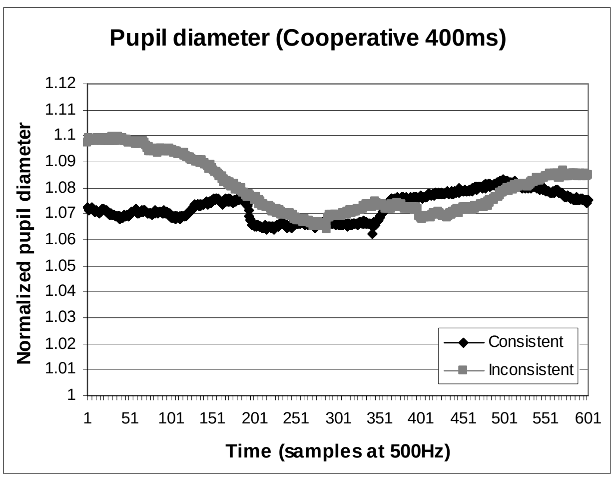

Figure 8: Pupil response beginning 200ms past onset of disambiguating word with cooperative 400 prosody.

General Discussion

The results from the first experiment showed that when prosody conflicts with syntactic structure, participants are more likely to misinterpret the sentence. More importantly, pupil diameter showed a clear positive slope in the disambiguating region, which suggests a higher processing load with Conflicting prosody, as expected. This finding supports models of sentence processing in which bottom-up information can affect initial syntactic processing (Frazier, 1987; Speer et al., 1996). The results from the second experiment showed that when the visual context supports the misinterpretation, participants are more likely to be garden-pathed. Here inconsistent Visual context appears to hurt performance in the 200ms Cooperative prosody condition. This result is supportive of unrestricted processing models. However, visual contexts were presented at the beginning of the utterance, therefore there was sufficient opportunity for the visual context to activate conceptual-level representations, which could have in turn primed the incorrect proposition that the woman cleaned the dog (see Henderson & Ferreira, 2004 for a review of the interface between vision and language). This explanation is all the more plausible considering that gist extraction from a scene occurs in less than 100ms (Castellanos & Henderson, submitted). In addition, there is some evidence to suggest that the processing system engages in some degree of syntactic prediction (Lau, Stroud, Plesch, & Phillips, 2006; Staub & Clifton, 2006). Given these facts it is not possible to definitely conclude that these data support unrestricted processing models. In current work, we are varying the onset of the visual context in order to determine the exact time course in which visual information exerts its effect.

These results are noteworthy for two additional theoretical reasons and one methodological reason. The first theoretical reason is that these findings corroborate the implicit prosody hypothesis (Fodor, 1998). This hypothesis assumes that readers project a prosodic contour onto sentences when reading. Many psycholinguistic studies have invoked this assumption to explain parsing preferences. Support for the implicit prosody hypothesis was observed in Experiment 1, where auditory sentences with conflicting prosody produced similar comprehension scores as previous research has shown with written materials. Therefore, our data indirectly support the assumption that readers impose the incorrect prosodic structure on garden-path sentences.

The second theoretical contribution concerns the conclusions made by Carlson et al. (2001). These authors argued that prosodic boundaries are not interpreted with respect to the absolute size of any particular prosodic boundary. Instead they argued that prosodic structure is interpreted based on the relative size of the boundary with respect to the global prosodic structure. Our results are contrary to these arguments because pupil diameter showed dramatic differences between the 200ms and 400ms Cooperative prosody conditions in Experiment 2, which suggests that the absolute duration of a boundary can affect online processing.

Finally, the methodological importance of this work concerns the use of pupil diameter as a dependent measure. Online investigations of spoken language comprehension have traditionally been limited to cross-modal lexical decision. The Visual World Paradigm was revolutionary in its ability to demonstrate distinct interpretations of ambiguous utterances based on fixation location (Tanenhaus et al., 1995). The current work which used pupil diameter shows similar promise for testing the competing models of sentence processing. If pupil diameter proves reliable across different paradigms and in different contexts, then psycholinguists have one more implicit online measure of processing difficulty than those used previously (cf. Just & Carpenter, 1993).

In conclusion, the current experiments showed that when no visual context was present prosody is a robust disambiguator of syntactic ambiguity. These results were evident in both the online and offline measures. The second experiment showed that Inconsistent visual context impaired performance in the 200ms Cooperative prosody condition. This result suggests that visual context can affect initial parsing, which would be consistent with unrestricted models of sentence processing. Further research will investigate the exact time-course of integrating visual, prosodic, and syntactic information.

Acknowledgements

This work was supported by NIH grant MH65310, and NSF IGERT grant DGE-0114378. 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 ToBI analysis of our stimuli.

References

Allbritton, D. W., McKoon, G., & Ratcliff, R. (1996). Relaibility of prosodic cues for resolving syntactic ambiguity. *Journal of Experimental Psychology: Learning, Memory, & Cognition , 22,* 714-735.

Carlson, K., Clifton, C., Jr., & Frazier, L. (2001). Prosodic boundaries in adjunct attachment. *Journal of Memory and Language, 45,* 58–81.

Castelhano, M. S., & Henderson, J. M. (submitted). The influence of color and structure on perception of scene gist.

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.

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

Fodor, J. D. (1998). Learning to parse. *Journal of Psycholinguistic Research, 27,* 285-319.

Fox Tree, J. E., & Meijer, P. J. A. (2000). Untrained Speakers' use of prosody in syntactic disambiguation and listeners' interpretations. *Psychological Research, 63,* 1–13.

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 601-681). Hillsdale, NJ: Erlbaum.

Henderson, J. M., & Ferreira, F. (2004). *The interface of language, vision, and action: Eye movements and the visual world.* New York: Psychology Press.

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.

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.

Kraljic, T., & Brennan, S. E. (2005). Prosodic disambiguation of syntactic structure: For the speaker or for the addressee? *Cognitive Psychology, 50,* 194-231.

Marcus, M., & Hindle, D. (1990). Description theory and intonation boundaries. In G. Altmann (Ed.), Cognitive models of speech processing: Computational and psycholinguistic perspectives. Cambridge, MA: MIT Press.

Lau, F., Stroud, C., Plesch, S., & Phillips, C. (2006). The role of structural prediction in rapid syntactic analysis. *Brain and Language, 98,* 74-88.

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.

Schafer, A. J., Speer, S. R., Warren, P., & White, S. D. (2000). Intonational disambiguation in sentence production and comprehension. *Journal of Psycholinguistic Research, 29,* 169-182.

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. (1998). Syntactic ambiguity resolution in discourse: Modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 24,* 1521-1543.

Staub, A., & Clifton, C. (2006). Syntactic prediction in language comprehension: Evidence from *either… or. Journal of Experimental Psychology: Learning, Memory, & Cognition, 32,* 425-436.

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.