USING PUPILLOMETRY TO ASSESS PROSODIC ALIGNMENT IN LANGUAGE COMPREHENSION

Jesse A. Harris; Sun-Ah Jun

University of California, Los Angeles
jharris@humnet.ucla; jun@humnet.ucla.edu

ABSTRACT

We show that the pupillometry method, which continuously captures minute changes in pupil size, provides a natural, real-time measure sensitive to mismatches between prosodic and syntactic grouping. Pupillometry is increasingly used as a measure of language comprehension, finding increased pupil size for syntactically complex sentences (e.g., [5]), violations of expected meter [18], and inadequate or misleading pitch accent [25].

Using materials from Kjelgaard & Speer [9], we found that improper prosodic boundary location increases pupil size, which increased differentially when the syntactic structure was independently hard to parse. The findings suggest that prosodic misalignment created additional cognitive load during auditory sentence comprehension, confirming that prosodic boundaries are immediately incorporated into structural analysis. We discuss implications for theories of prosody-syntax alignment in real-time sentence parsing, along with methodological and statistical innovations afforded by this method.

Keywords: Prosodic alignment; boundary placement; prosodic processing; pupillometry

1. INTRODUCTION

1.1. Prosodic boundaries and alignment

In spoken speech, the degree of juncture between each word is different, reflecting the organizational structure of the utterance, known as *prosodic structure*, the hierarchically organized groupings of words. These groupings are prosodic units such as an Intonational Phrase (IP; a major prosodic unit) or Intermediate Phrase (ip; a minor prosodic unit) [3, 7]. Since prosodic grouping is influenced by many factors such as syntactic structure, information structure, rhythm, length of a phrase, and even speech rate [21], there can be multiple prosodic boundaries in an utterance. Though the prosodic structure is not isomorphic to syntactic structure, the boundary of a major prosodic unit often matches or aligned with the boundary of a major syntactic constituent [21, 7, 20]. In English, there are multiple acoustic cues marking an IP boundary. The end of an IP is realized by pre-boundary lengthening and changes in pitch (called a *boundary tone*, e.g., H% for ending with a high pitch, L% for ending with a low pitch), and possibly followed by a pause [3, 12]. Studies on the role of prosodic phrasing in syntactic disambiguation have shown that speakers use a prosodic boundary to deliver their intended syntactic structure, and major prosodic boundaries are robustly perceived by listeners [15].

Even though not all prosodoic boundaries are obligatory, listeners are nonetheless highly sensitive to their location in disambiguation, as syntactic clauses and major intonational phrases (IP) tend to be closely aligned with each other. For example, prosodic boundaries are known to disambiguate globally (1) and temporarily (2) ambiguous structures [9, 17, 22]. In (1), even though the prepositional phrase *from Alabama* is structurally ambiguous, the location of a boundary (%) can disambiguate whether it modifies the noun (1a) or the verb (1b), indicated by underlining.

(1) a. Paula phoned % <u>her friend</u> from Alabama
b. Paula <u>phoned</u> her friend % from Alabama

Similary, boundary placement can disambiguate temporarily ambiguous sentences. In (2), the noun *the house* is temporarily ambiguous between the subject of a new clause (2a: *Early closure*) and the object of the preceding verb (2b: *Late closure*). A Late closure interpretation is generally preferred in the absence of prosodic or orthographic cues. There are many explanations of this preference, including the idea that treating *the house* as an object is initially syntactically less complex, and must be reinterpreted as a subject upon encountering the disambiguating matrix verb *is* [6].

(2) When Roger leaves <u>the house</u> ...
a. When Roger leaves % the house is dark

b. When Roger leaves the house % it's dark

How do prosodic boundaries affect parsing decisions? In a series of mainly offline experiments, Kjelgaard & Speer [9] found that interpreting a sentence was facilitated when a prosodic boundary aligned with the syntactic clause boundary (Cooperating prosody), and that interpretation was hindered when the prosody and syntax were misaligned (Non Cooperating prosody) compared to a neutral prosody baseline condition. Disambiguating boundaries have also been shown to elicit unique electrophysiological signatures in electromagnetic (EEG) studies [23]. In addition, visual world paradigm (VWP) studies have shown that listeners use prosodic boundaries to anticipate upcoming structure [22].

However, each of these methods has potential drawbacks: offline measures are too coarse to assess real time processing; EEG measures are comparatively expensive and time-consuming; and gaze patterns within the VWP are highly contingent on the images or scenes presented to the subject, and may therefore encourage task-specific strategies [16]. Though no measure is perfect, the pupillometry method outlined below offers a promising middle ground: an easy to administer, comparatively inexpensive, but online and passive measurement.

1.2. Pupillometry method

Pupillometry records continuous changes in pupil diameter associated with a stimulus, typically reaching its peak between 700 and 1200ms after stimulus onset [13]. Pupil size naturally fluctuates involuntarily, and dilates in response to (i) the presentation of an external stimulus (e.g., changes in luminance, arousing imagery, difficult arithmetic problems), or (ii) neural inhibition from either the parasympathetic oculoumotor system or the locus coeruleus of the noradrenergic system [24]. Pupil size increases as a function of cognitive load or mental effort, and has been implicated in many tasks involving memory and attention [13].

Pupil dilation has recently been explored as a measure of language comprehension (see [19] for review), finding increased pupil size for structurally complex sentences [5, 4, 8], semantic anomalies [4], lexical frequency or increased emotional valence [10], violations of expected meter [18], and inadequate or misleading pitch accent [25]. Though still sparse, the current literature strongly suggests that growth in pupil size indexes increased effort and attention associated with difficult to interpret material along various linguistic dimensions.

Engelhardt et al. [5] measured pupil dilation on Early closure sentences like (3), manipulating the presence and size of a prosodic break at the clause junction (between *cleaned* and *the dog*). Pupil dilation after the temporarily ambiguous noun (*the dog*) increased in conditions that lacked the appropriate prosodic boundary, suggesting that a prosodic boundary aligned with the syntactic clause eased the interpretation of syntactically difficult structures.

(3) While the woman cleaned (%) the dog that was big and brown <u>stood</u> in the yard.

The present experiment complements Engelhardt et al.'s study by making several methodological and statistical improvements. First, we added a Late closure baseline, to assess the effect of Non Cooperating prosody on sentences that do not typically produce processing difficulty. Second, we recorded pupil dilation immediately after the prosodic phrase containing disambiguating information. Importantly, the pupil was recorded during silence, so acoustic differences between sentences could not affect pupil size. Third, we used a statistical measure that is more appropriate for continuous measures with complex functional forms than is possible with the ANOVAs or linear regression models that were in standard use at the time. In particular, we used a growth curve analysis to model complex changes in pupil dilation over time.

2. EXPERIMENT

2.1. Participants

Thirty-four native college-aged speakers of American English were recruited from a Psychology Subject Pool at UCLA. No participant reported any history of hearing loss or language disorder. Experimental sessions lasted no more than 30 minutes, and participants were compensated with course credit.

2.2. Materials and method

Subjects were instructed to look at a fixation cross while listening to a sentence until it disappeared (3 seconds post-sentence). Their heads were stabilized and the size of their pupils was measured by a high-resolution eye tracker. Individual variation in pupil size was normalized by subtracting pupil size from a 500ms pre-sentence baseline recorded on a series of # marks, and then calculating the percent change at each point over the entire trial per subject. Auditory materials were played over loud speakers in a sound isolated room.

Sixteen quartets, modified from [9], were constructed, in a 2x2 design crossing Prosody (Cooperating, Non Cooperating), Closure syntax (Early, Late). A sample item is shown in Table 1 with the ToBI tone labels [2] representing the prosodic boundaries and nuclear pitch accents found in experimental items. The 16 experimental items were presented in counterbalanced and individually randomized order, interspersed with 36 sentences from unrelated experiments, and 26 non-experimental filler sentences, for a total of 78 sentences per experimental session. Comprehension questions were presented after half of the items to ensure that subjects were paying attention, but were not analyzed further.

2.3. Results

To avoid assuming a linear form or an arbitrary time window for analysis, the data were fit to a growth curve model [14], a powerful statistical approach, which can be used to quantify continuous changes in pupillary response [11]. The best fitting model was a third-order (cubic) orthogonal polynomial with fixed effects of Prosody, Closure, and their interaction on all time terms, and by-subject and by-item random slopes [1], as shown in Table 2. Experimental predictor variables received deviation coding with Cooperating prosody and Late closure conditions specified as the baselines for their respective factors.

Table 2: Linear mixed effects regression model of a growth curve analysis with three orthogonal polynomial terms (Linear; Quadratic; Cubic). Significant terms marked as * rows.

Parameter	Estimate	SE	t-value
(Intercept)	-0.08	0.04	-1.94 +
Linear	-16.94	0.31	-55.55 *
Quadratic	1.58	0.31	5.19 *
Cubic	2.07	0.31	6.78 *
NonCoop	0.01	0.04	0.18
Early	-0.01	0.04	-0.22
Linear:NonCoop	1.86	0.31	6.11 *
Quadratic:NonCoop	-1.84	0.31	-6.03 *
Cubic:NonCoop	0.16	0.31	0.51
Linear:Early	-1.64	0.31	-5.37 *
Quadratic:Early	-1.03	0.31	-3.37 *
Cubic:Early	0.91	0.31	2.96 *
NonCoop:Early	0.00	0.04	-0.11
Linear:NonCoop:Early	-2.13	0.31	-6.99 *
Quadratic:NonCoop:Early	-0.10	0.31	-0.32
Cubic:NonCoop:Early	0.89	0.31	2.93 *

We follow [11] in the interpretation of pupil dilation in terms of the growth curve components: larger *linear* terms indicate steeper deflection from a flat line; more negative *quadratic* terms indicate increased non-linearity; and more positive *cubic* terms reflect the steepness of the rise and fall of the pupillary response; see Figure 1.

We summarize the major points from the statistical model from Table 2. In cases of Non Cooperating prosody, there were steeper slopes (an increased linear polynomial) with a more non-linear shape (identified in terms of the quadratic component). For the Closure condition, Early closure sentences elicited steeper slopes with earlier, more acute peak amplitudes (effects in all three polynomials) than Late closure counterparts. Finally, we observed an interaction between the two conditions: The effect of Non Cooperating prosody was greater for Early Closure with steeper, more acute peak (the linear and cubic components).

In all, the results show that the conditions elicited distinct pupillary responses from our subjects. We assume that increased peak and slope index increased cognitive load. There were processing penalties for both misleading prosody and syntactically complex sentences. These conditions interacted, so that there was an increased penalty for Non Cooperating Early closure sentences, in which the misleading prosodic boundary was compatible with the Late closure interpretation, possibly increasing the cost of misinterpretation.

Figure 1: Pupil dilation after sentence offset

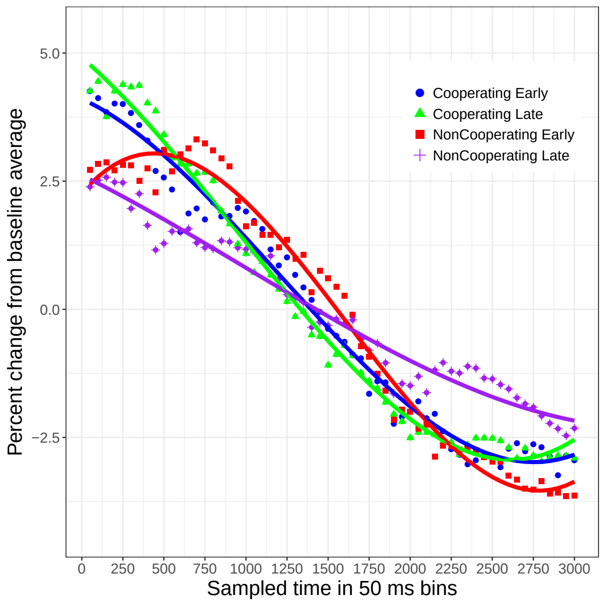
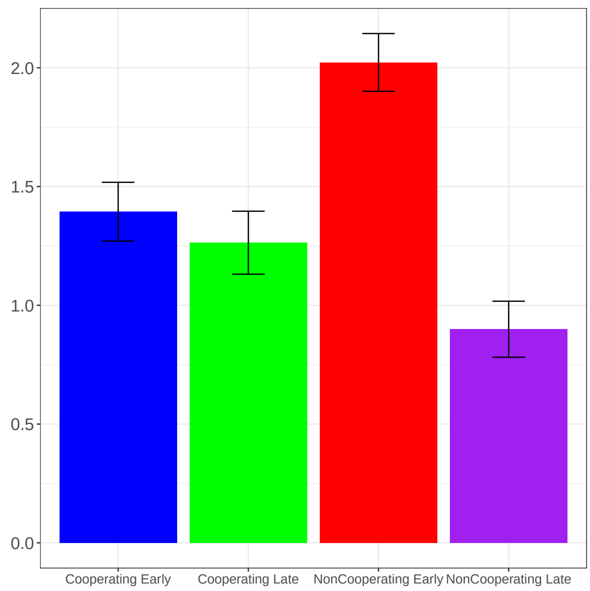


Table 1: Sample quartet from experiment. A tone in parentheses indicates that the tone was not present in all items.

Closure	Cooperating prosody	Non Cooperating prosody
Early	(When Roger leaves) % (the house is dark) * H* L-L% H* (!H*) L-L%	(When Roger leaves the house) % (is dark) * H* L-L% H* L-L%
Late	(When Roger leaves the house) % (it's dark) * H* L-L% H* L-L%	(When Roger leaves) % (the house it's dark) * H* L-L% H* (L-) H* L-L%

The interaction can perhaps be seen most clearly by comparing the mean change in pupil size in the same 3000ms time period, shown in Figure 2. Though the trends are not statistically significant in more traditional LMER analyses, Non Cooperating prosody differentially increased pupil dilation in Early closure structures.

Figure 2: Mean pupil size with standard errors



3. DISCUSSION

The overall pattern is highly consistent with Kjelgaard and Speer's [9] initial offline findings, as well as Engelhardt et al.'s [5] earlier pupillometry study. In general, our findings show that change in pupil dilation reflects mismatches between prosodic grouping and syntactic structure. Furthermore, the cost is greater for mismatches that initially support independently preferred syntactic analyses, i.e., the Non Cooperating Early Closure condition. We entertain two possible interpretations of this interaction.

The first interpretation is a syntactic one. The interaction might indicate that the language processor uses prosodic boundary information in tandem with structure building principles to generate sentence structures in real time. The boundary location in the Non Cooperating Early closure condition would lend prosodic support to the preferred structure, thereby cementing the interpretation of *the house* as an object to the embedded verb *leaves*. The Non Cooperating Late closure condition, however, would present prosodic evidence for the dispreferred structure to the parser, in which *the house* functions as the subject of a separate clause. That is, even though the parser would compute, but later reject, an Early closure structure, the generally preferred Late closure syntax might be more easily recovered.

An alternate interpretation is phonetic. The majority of our items traded on the acoustic similarity between *is* and *it's*. Listeners may have simply misperceived the contraction *it's* as the copula *is* in the Non Cooperating Late closure condition, thereby reducing the penalty for failing to align the prosody and the syntax. Currently, we do not have evidence to properly arbitrate between the two possibilities.

To conclude, we have shown that increases in pupil size track misleading prosody, and that there are greater effects on changes in pupil dilation when the misleading prosody supports a preferred structure. We take this result as a proof of a general methodological concept: pupillometry complements existing online measures (such as visual world eye tracking and EEG), by offering a highly versatile, yet simple to administer method for exploring the online computation of linguistic prosody during sentence processing in naturalistic environments. When paired with a growth curve analysis, it provides a powerful tool for exploring the dynamics of how prosodic and syntactic constituencies are formed and aligned in real time sentence comprehension.

REFERENCES

[1] Baayen, R. H., Davidson, D. J., Bates, D. M. 2008. Mixed-effects modeling with crossed

random effects for subjects and items. *Jrnl of Mem and Lang* 59, 390–412.

[2] Beckman, M. E., Hirschberg, J., Shattuck-Hufnagel, S. 2005. The original ToBI system and the evolution of the ToBI framework. In: Jun, S.-A., (ed), *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press 9–54.

[3] Beckman, M. E., Pierrehumbert, J. B. 1986. Intonational structure in Japanese and English. *Phon Yearbook* 3, 255–309.

[4] Demberg, V., Sayeed, A. 2016. The frequency of rapid pupil dilations as a measure of linguistic processing difficulty. *PloS one* e0146194.

[5] Engelhardt, P. E., Ferreira, F., Patsenko, E. G. 2010. Pupillometry reveals processing load during spoken language comprehension. *Quart Jrnl of Exp Psych* 63, 639–645.

[6] Frazier, L. 1987. Sentence processing: A tutorial review. In: Coltheart, M., (ed), *Attention and Performance XII*. Lawrence Erlbaum Associates.

[7] Jun, S.-A. 1998. The accentual phrase in the Korean prosodic hierarchy. *Phonology* 15, 189–226.

[8] Just, M., Carpenter, P. A. 1993. The intensity dimension of thought: pupillometric indices of sentence processing. *Canadian Jrnl of Exp Psych* 47, 310–339.

[9] Kjelgaard, M. M., Speer, S. R. 1999. Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Jrnl of Mem and Lang* 40, 153–194.

[10] Kuchinke, L., Võ, M. L.-H., Hofmann, M., Jacobs, A. M. 2007. Pupillary responses during lexical decisions vary with word frequency but not emotional valence. *Intl Jrnl of Psychophysiology* 65, 132–140.

[11] Kuchinsky, S. E., Ahlstrom, J. B., Vaden Jr, K. I., Cute, S. L., Humes, L. E., Dubno, J. R., Eckert, M. A. 2013. Pupil size varies with word listening and response selection difficulty in older adults with hearing loss. *Psychophysiology* 50, 23–34.

[12] Ladd, D. R. 1996/2008. *Intonational phonology*. Cambridge: Cambridge University Press.

[13] Laeng, B., Sirois, S., Gredebäck, G. 2012. Pupillometry: A window to the preconscious? *Persp on Psych Sci* 7, 18–27.

[14] Mirman, D. 2016. *Growth curve analysis and visualization using R*. CRC Press.

[15] Price, P. J., Ostendorf, M., Shattuck-Hufnagel, S., Fong, C. 1991. The use of prosody in syntactic disambiguation. *JASA* 90, 2956–2970.

[16] Salverda, A. P., Brown, M., Tanenhaus, M. K. 2011. A goal-based perspective on eye movements in visual world studies. *Acta psychologica* 137, 172–180.

[17] Schafer, A. J. 1997. *Prosodic parsing: The role of prosody in sentence comprehension*. PhD thesis UMass Amherst.

[18] Scheepers, C., Mohr, S., Fischer, M. H., Roberts, A. M. 2013. Listening to limericks: a pupillometry investigation of perceivers' expectancy. *PLoS One* 8, e74986.

[19] Schmidtke, J. 2018. Pupillometry in linguistic research: An introduction and review for second language researchers. *Studies in Second Language Acquisition* 40, 529–549.

[20] Selkirk, E. 2011. The syntax-phonology interface. In: Goldsmith, J., Riggle, J., Yu, A., (eds), *The Handbook of Phonological Theory*. Oxford, UK: Blackwell 435–483 2nd edition.

[21] Shattuck-Hufnagel, S., Turk, A. E. 1996. A prosody tutorial for investigators of auditory sentence processing. *Jrnl of Psych Res* 25, 193–247.

[22] Snedeker, J., Trueswell, J. 2003. Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Jrnl of Mem and Lang* 48, 103–130.

[23] Steinhauer, K., Alter, K., Friederici, A. D. 1999. Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nat Neurosci* 2, 191–196.

[24] Wilhelm, B., Wilhelm, H., Lüdtke, H. 1999. Pupillography: Principles and applications in basic and clinical research. *Pupillography: Principles, methods and applications* 1–11.

[25] Zellin, M., Pannekamp, A., Toepel, U., van der Meer, E. 2011. In the eye of the listener: Pupil dilation elucidates discourse processing. *Intl Jrnl of Psychophysiology* 81, 133–141.