The Effect of Poor Source Code Lexicon and Readability on Developers' Cognitive Load




Sarah Fakhoury Yuzhan Ma Venera Arnaoudova Olusola Adesope

Washington State University, USA

WASHINGTON STATE UNIVERSITY

50%
Program
Comprehension

50%
Other software
Maintenance
Activities

Studying Program Comprehension

Studying Program Comprehension

Think aloud

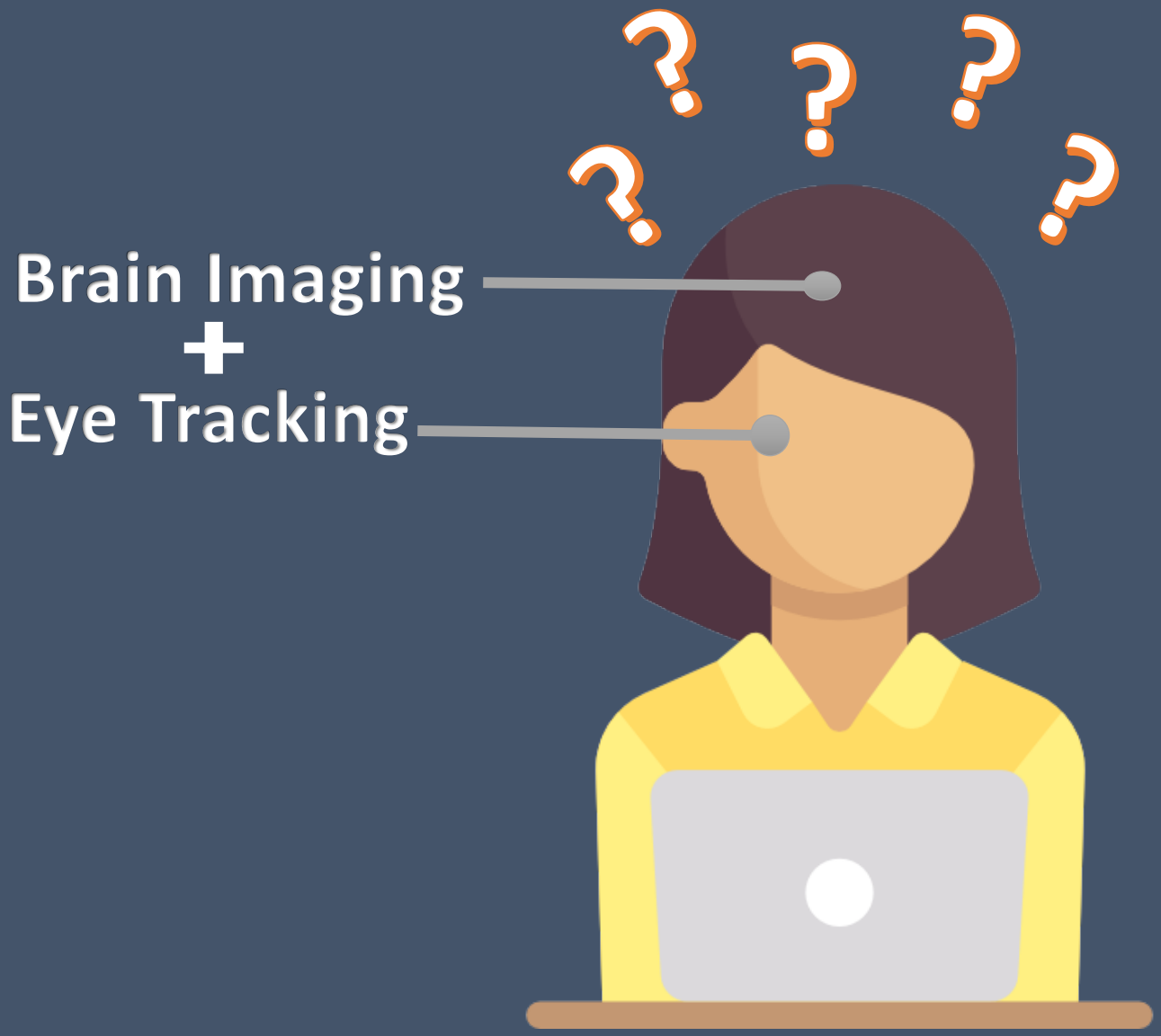
Surveys

Interviews

Comprehension Summaries

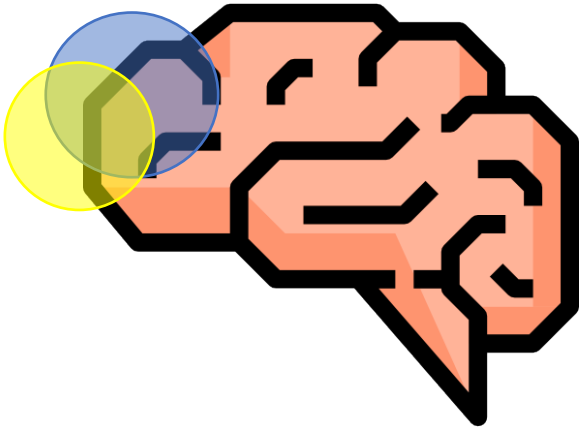
Studying Program Comprehension

Studying Program Comprehension



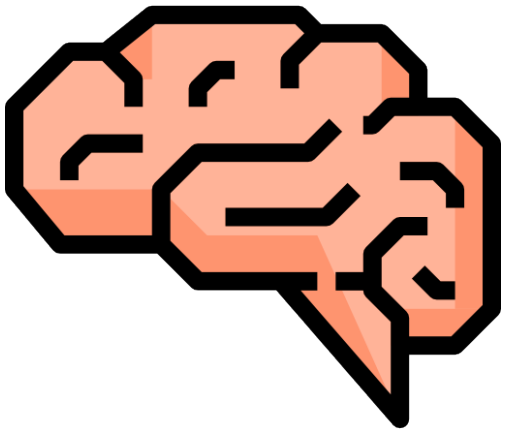
Brain Imaging:
functional Near InfraRed Spectroscopy

Measures oxygenated and deoxygenated hemoglobin



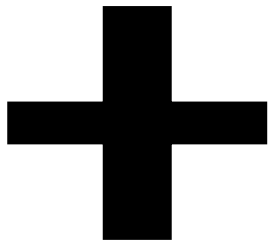
Increased oxygenated hemoglobin indicates an increase in cognitive load

+ Used in a wide variety of working memory research
+ Minimally Invasive
+ Allows replication of real working environments



Brain Imaging

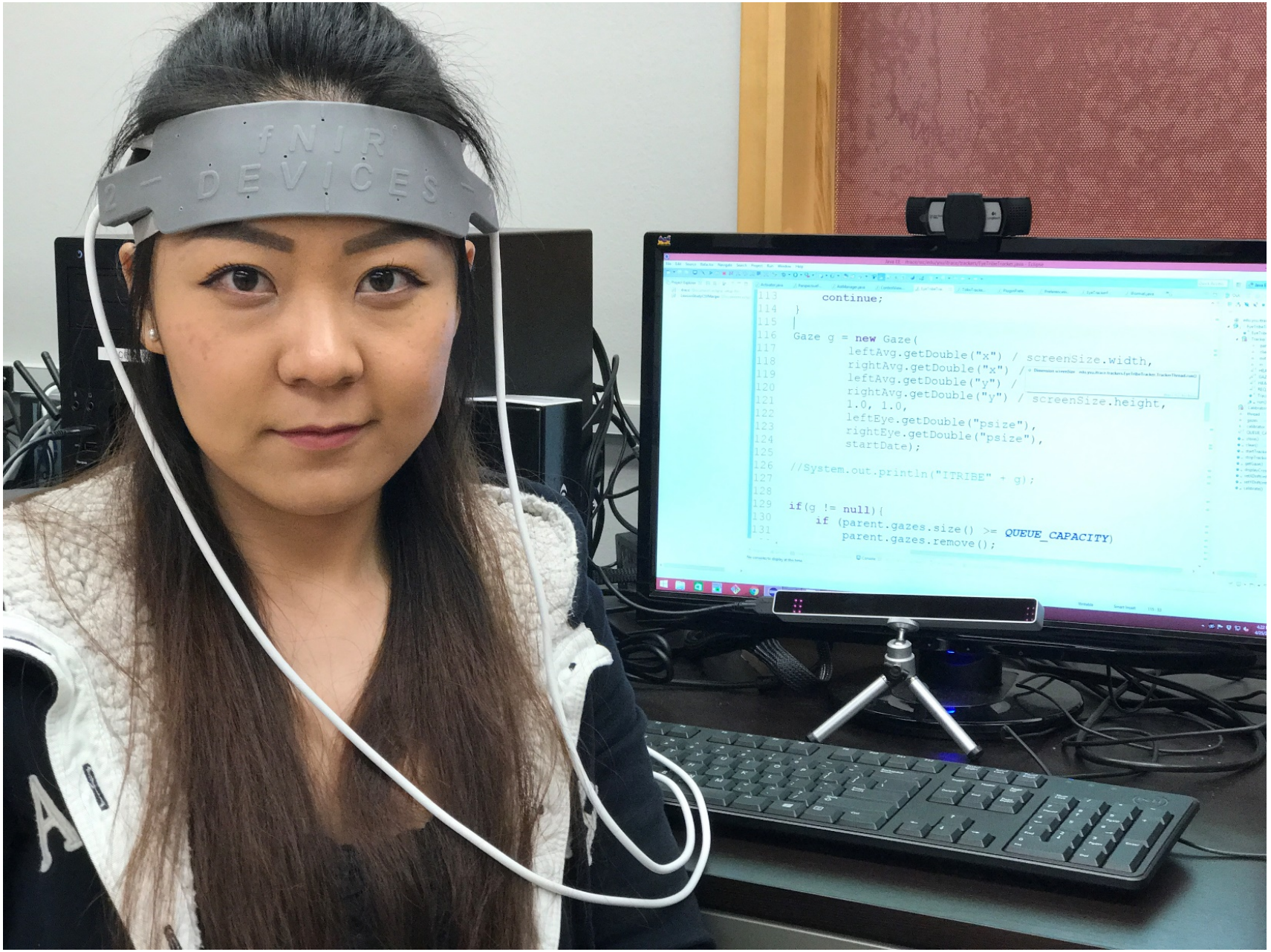
Cognitive Load



Eye Tracking

Specific identifiers in the Source code

```
    if (wort.length() <= (pos + Länge)) {
      return wort.substring(pos);
    }
    return wort.substring(pos, pos + Länge);
  }
}
```


fNIR
DEVICES
continue;
}
Gaze g = new Gaze(
leftAvg.getDouble("x") / screenSize.width,
rightAvg.getDouble("x") /
leftAvg.getDouble("y") /
rightAvg.getDouble("y") / screenSize.height,
1.0, 1.0,
leftEye.getDouble("psize"),
rightEye.getDouble("psize"),
startDate);
//System.out.println("ITRIBE" + g);
if(g != null){
if (parent.gazes.size() >= QUEUE_CAPACITY)
parent.gazes.remove();

Can we accurately associate cognitive load to identifiers using fNIRS and eye tracking devices?

Comprehension Task

RQ1: Can developers' cognitive load be accurately associated with identifiers' terms using fNIRS and eye tracking devices?

Participants asked to read source code in foreign language or prose snippet

Carefully highlight areas of the code that were difficult or took a lot of time to understand

Results

Treatment	SMC	Treatment	SMC
German Code	0.87	Prose	0.81
	0.76		0.81
	0.82		0.70
	0.79		0.73
	0.81		0.65
	0.82		0.75
Average	0.81	Average	0.74
Total Average			**0.78**

Cognitive load can be captured with a similarity of 78% compared to self-reported results

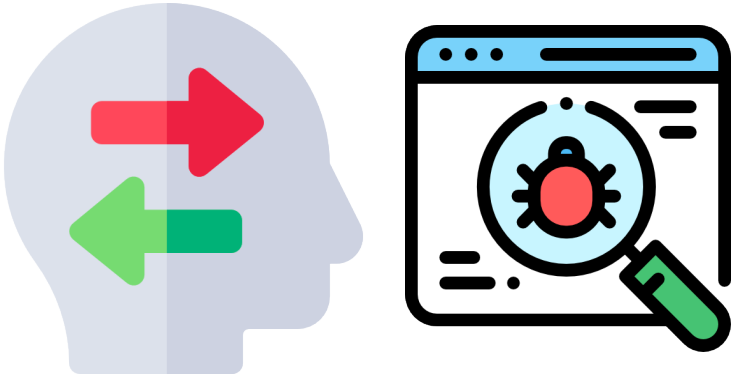
Determine the

Structural vs. Lexical

Implications on Cognitive Load

Bug Localization Task

RQ2: Do the different structural and lexical inconsistencies in the source code cause a measurable increase in developers' cognitive load during program comprehension?



Psychological Complexity of Source Code

Structural Features

Cyclomatic Complexity

LOC

Formatting

Lexical Features

Comments

Identifiers

Documentation

Structural Features

Feature	Corr.	Feature	Corr.
Cyclomatic Complexity	– – –	Halstead vocabulary	–
Number of Arguments	– – –	Halstead length	–
Number of operands	– – –	Number of casts	–
Class References	– – –	**Number of loops**	–
Local Method References	– – –	**Number of expressions**	–
Lines of Code	– – –	**Number of statements**	–
Halstead effort	– –	**Variable Declarations**	–
Halstead bugs	– –	**Number of Comments**	+ +
Max depth of nesting	– –	**Number of Comment Lines**	+ +
External Method References	– –	**Number of Spaces**	+ +
Halstead volume	–	Number of operators	+
Halstead difficulty	–		

[1] Raymond P.L. Buse and Westley R. Weimer. 2010. Learning a metric for code readability. IEEE Transactions on Software Engineering (TSE) 36, 4 (2010), 546–558.
[2] Maurice H Halstead. 1977. Elements of software science. (1977).
[3] Thomas J. McCabe. 1976. A Complexity Measure. IEEE Transactions on Software Engineering (TSE) SE-2, 4 (1976), 308–320.
[4] Daryl Posnett, Abram Hindle, and Premkumar Devanbu. 2011. A Simpler Model of Software Readability. In Proceedings of the Working Conference on Mining Software Repositories (MSR). 73–82.

Lexical Features

- Introduce Linguistic Antipatterns to source code snippets

"Poor recurring practices that create inconsistencies between naming, documentation, and implementation of the software." [1]

[1] Arnaoudova, Venera; Di Penta, Massimiliano; Antoniol, Giuliano Linguistic Antipatterns: What They are and How Developers Perceive Them Empirical Software Engineering (EMSE), 21 (1), pp. 104–158, 2015

Linguistic Antipatterns

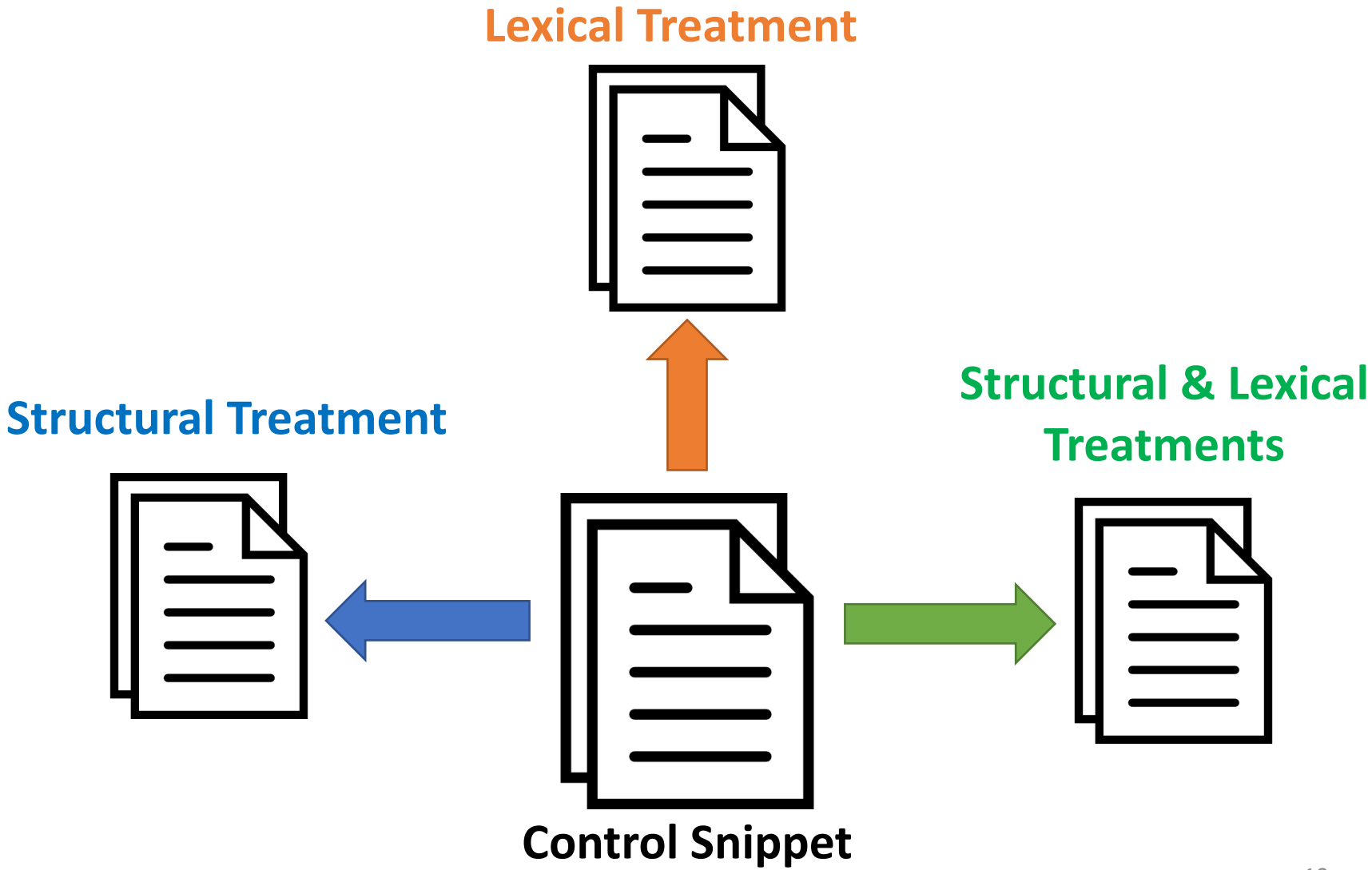
- **A get method that does not return**

```
public void getIdNumber(String userName){
    //...
}
```

- **A method signature and comment are opposite**

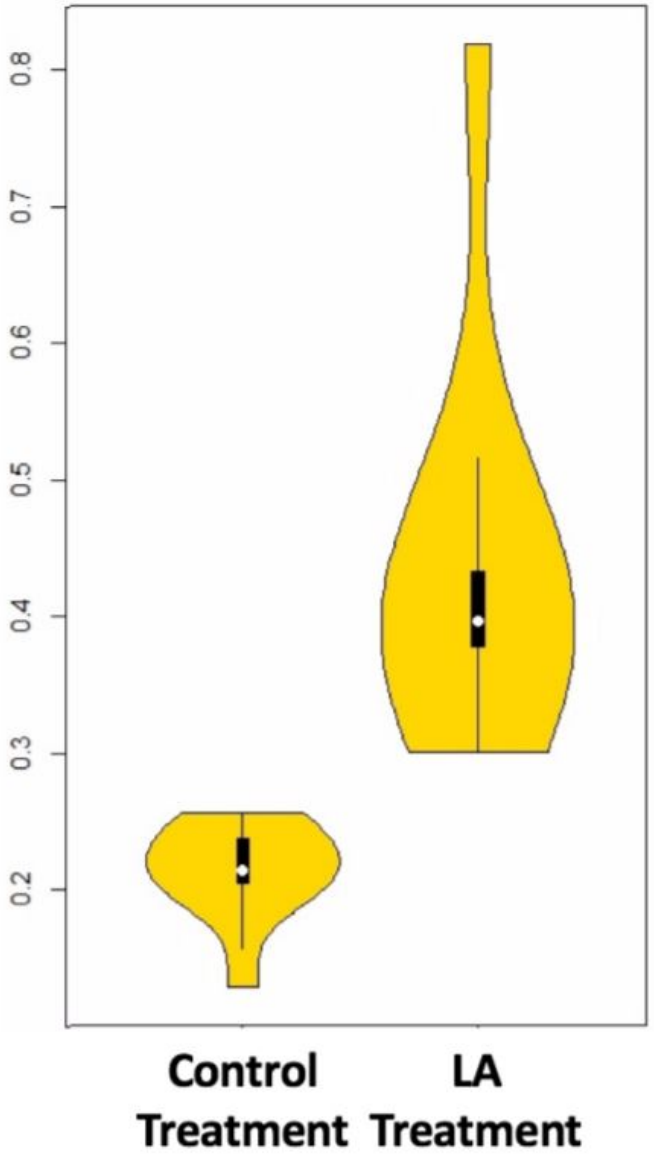
```
/**This method finds all singular words in a sentence**/
public String[] getAllPluralWords(String text) {
    // ...
```

Bug Localization Task



Taken from open source projects Java

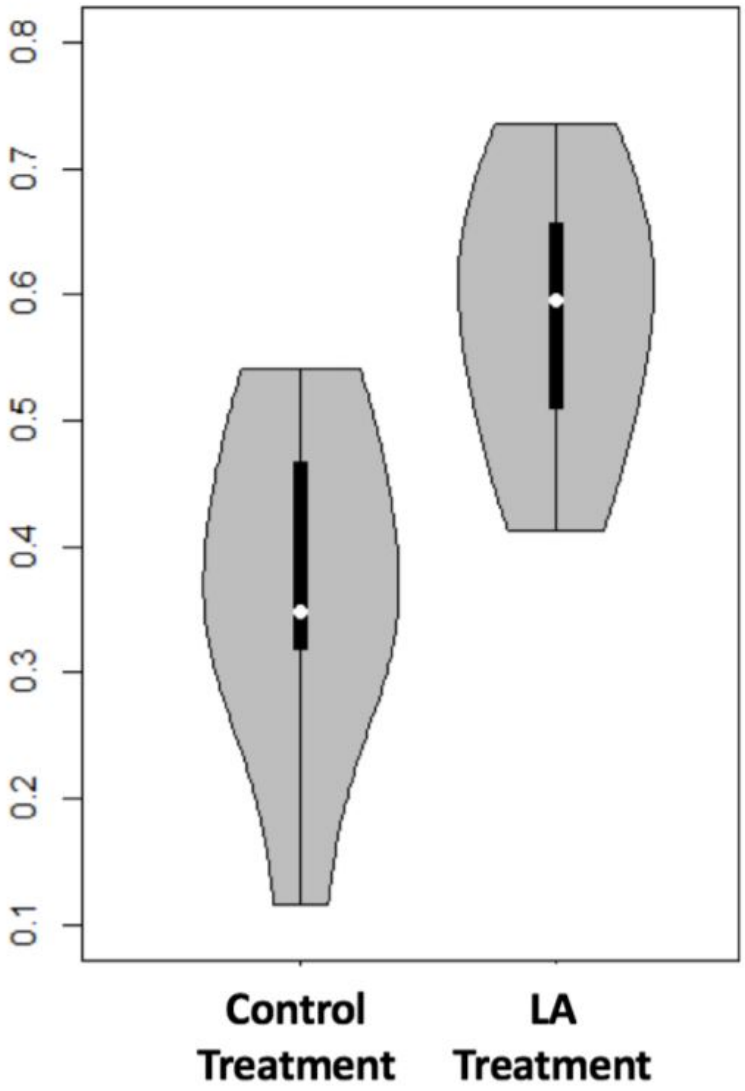
Lexical Treatment



Data points that contain high cognitive load in control snippets vs LA treatment.

***Significant* p-value (0.0009) with large effect size**

Lexical Treatment



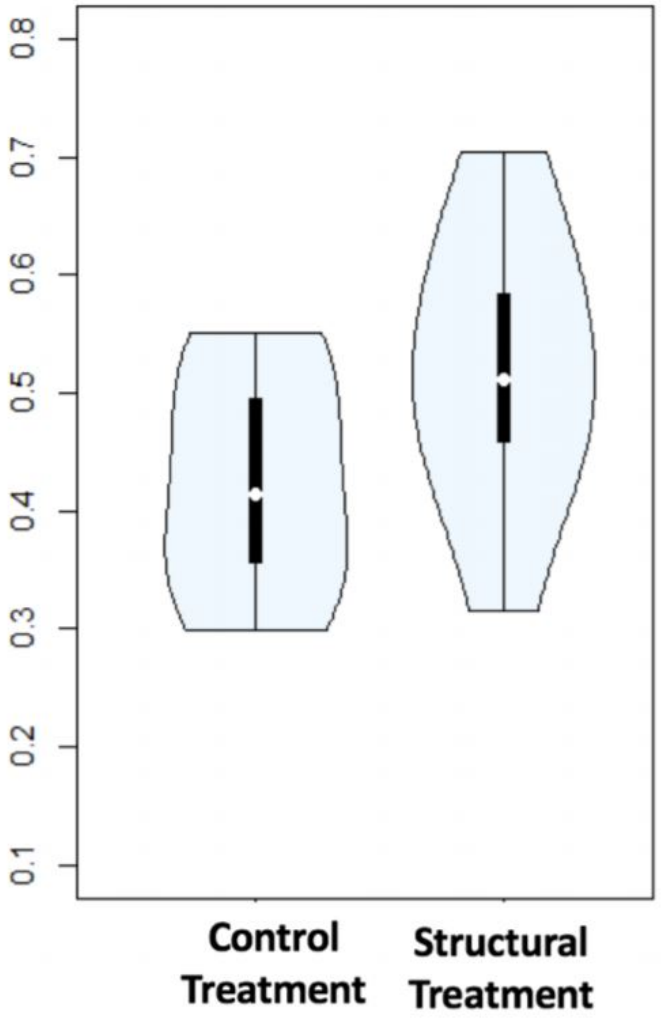
Average normalized oxygenation per participant

***Significant* p-value 0.003 with large effect size**

Lexical Treatment: Conclusions

Participants who encountered linguistic antipatterns had a higher percentage of fixations that indicated high cognitive load and their overall average oxygenation was higher.

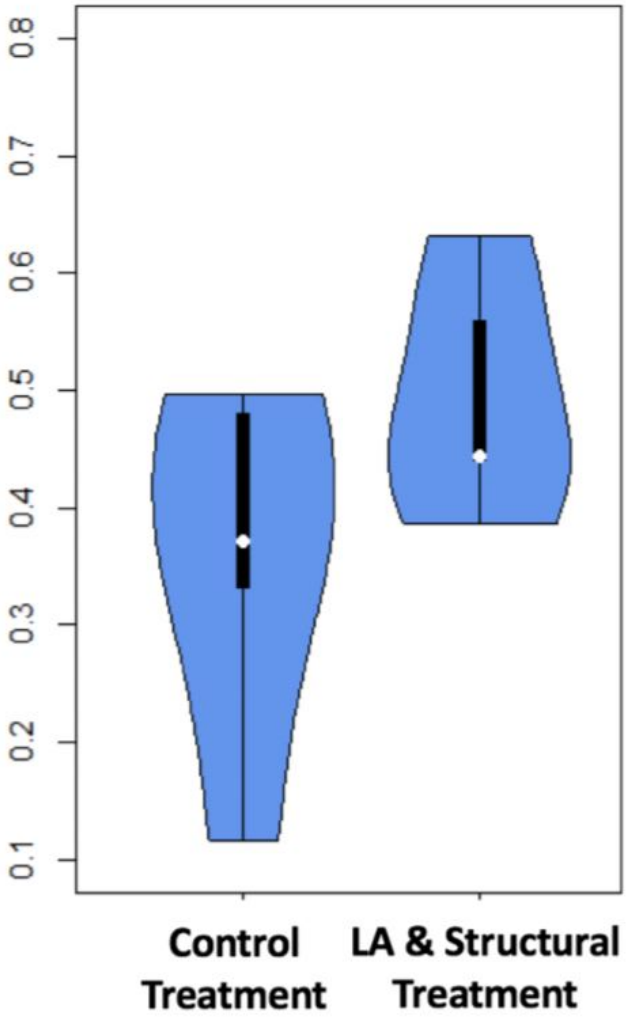
Structural Treatment



Average normalized oxygenation per participant

p-value 0.14 with medium effect size

Lexical & Structural Treatment



Average normalized oxygenation per participant

p-value 0.48 with small effect size

Structural & Lexical Treatment: Conclusions

Participants found structural snippets frustrating and difficult to read but this **did not** cause a significant increase in average oxy.

The treatment containing both structural and LAs mislead more than 60% of participants and they **could not** successfully complete the task.

Overall Conclusions

1. Using fNIRS and eye tracking we can measure cognitive load at a fine grained level.

2. Linguistic antipatterns significantly increase cognitive load of developers during comprehension tasks

3. Structural inconsistencies may cause frustration to the participants but there is no significant increase in their cognitive load as compared to control treatments.