# Understanding Collaborative Program Comprehension: Interlacing Gaze and Dialogues

**Abstract:** We study the interaction of the participants in a pair program comprehension task across different time scales in a dual eye-tracking setup. We identify four layers of interaction episodes at different time scales. Each layer spans across the whole interaction. The present study concerns the relationship between different layers at different time scales. The first and third layers are based on the utterances of the participants while the second and fourth layers are based on participants' gaze.

## Introduction

In CSCL, one main open challenge is to use technology to measure the dynamics of interaction. We report recent developments in eye-tracking which show how gaze can be used to reflect cognitive and collaborative processes at various time scales. Thereby we scale up the social unit of analysis from individual to pair and scale down the temporal unit of analysis from the whole interaction to shorter interaction episodes.

With regards to the social unit of analysis, gaze has traditionally been used to assess individual cognition (e.g. eye-tracking studies about reading, program comprehension, etc.). However, in the context of CSCL, a methodology is needed to describe collaborative gaze. Various measures of "gaze togetherness" have been used to indicate the quality of collaboration in dyadic interaction. In general, good collaboration features convergent gaze. Gaze togetherness increases significantly especially during verbal and deictic references. These measures of togetherness are however related to a global time scale and don't consider the evolution of gaze focus during interaction.

With regards to the temporal granularity of analyses, studies have emphasized overall measures of individual attention. For example, studies (Romero et.al, 2002; Bednarik & Tukiainen, 2006; Bednarik et. al., 2006; Sharif & Maletic, 2010; Hejmady & Narayanan, 2012; Pietinen et. Al., 2008; Pietinen et. Al., 2010; Bednarik & Shipolov, 2011) report the proportion of time that subjects spend fixating different parts of the interface. These measures indicate overall gaze behavior (and may be correlated with expertise) but they cannot serve as real-time indicators of collaboration that could be used to provide immediate feedback. In the context of CSCL, the dynamics of interaction and dialogue are important indicators for collaborative knowledge building (e.g. Stahl, 2000). New gaze indicators are needed to reflect the knowledge building at the micro level.

Time scales have been used to describe behavior at various levels. Eye-trackers allow us to capture attention at a time scale that has more information content than the other measures like interface event logs, dialogues or gestures. In a controlled experiment (Lord & Levy, 2008) the duration of eye-fixations have duration of the order of 100 milliseconds, which gives them a place at the lower end of cognitive behavioral band (Newell, 1994). Cognitive behavioral bands have complex actions (e.g., reading or gestures) at the higher end. Anderson (2002) identifies cognitive modeling as bridging across the behavioral bands by taking the lower level bands into account. We will reuse the levels by Anderson \cite{anderson_spanning_2002} to refer to the Task (where we usually measure understanding), Unit task (where we usually code dialogues) and Operations (where we usually collect raw data).

In this contribution, we address both the social and temporal mismatch of current gaze methodology with the study of collaborative interaction. We propose a method to detect interaction episodes based on both gaze togetherness and stability and show that these measures are related to the level of understanding that a pair achieves at the end of the task. To support our proposal, we present a dual eye-tracking study in a remote pair program comprehension scenario.

The rest of the paper is organized as follows: the second section gives the related work for the present study. The third section describes the main features of the study and the research questions. The fourth section describes the experiment and the various variables. The fifth section presents the analysis results. Finally the sixth section discusses the results and concludes the paper.

## Related Work

### Adaptive Support for CSCL

Adaptive CSCL has been around for 10 years. Jermann et. al. (2001) proposed a feedback model for collaborative interaction regulation. The regulation is based on the collection of collaborative indicators that are assessed by the system or by the human learners and teachers. More recently, Magnisalis (2011) propose that web 2.0 and artificial intelligence are increasingly used to design reactive systems and that the learners benefit from the adaptation of the systems.

## Scaling Up The Social Unit

There are different gaze-based measures of collaboration given by Richardson & Dale (2005), Cherubini et. al. (2008) and Pietinen et. al. (2010). Richardson & Dale (2005) used "gaze togetherness" as a notion of gaze cross recurrence (how much the participants are looking at the same object at the same time). Cherubini et. al. (2008) used eye tracking in a remote collaborative problem solving setup to detect the misunderstanding (distance between the referrer's and the partner's gaze points) between the collaborating (through chat) partners. Pietinen et. al. (2010) gave a new metric, to measure joint visual attention in a co-located pair programming setup, using the number of overlapping fixations and use the fixation duration of overlapping fixation for assessing the quality of collaboration. The problem of these measures is that they characterize togetherness on a global temporal level or on an arbitrarily defined timespan (one could partition the interaction into "n" parts but these would not reflect the underlying interactive dynamics).

## Linking Gaze and Speech

At the level of operations, there are studies about gaze and speech coupling (Mayer et. al., 1998; Griffin & Bock, 2000; Zelsinky & Murphy, 2000). There are different notions eye-voice span given in different studies, but all the notions point towards a strong coupling between speaker's gaze and speech. Allopenna et. al. (1998) showed that the mean delay between hearing a verbal reference and looking at the object of reference (the listeners' voice-eye span) is between 500 and 1000 milliseconds. The combination of eye-voice and voice-eye coupling is that the gaze of speakers and listeners are coupled with a lag of about 2000 milliseconds. This short term coupling between speaker and listener is at the operation level only and does not inform about the relationship of gaze and dialogue in longer episodes. This is problematic when one is interested in knowledge building episodes that usually last for several utterances.

## Linking Dialogue and Understanding

Concerning the relationship between dialogues and understanding, there is a long-standing tradition of research in CSCL. For example, the elaborated explanations (Cohen, 1994; Webb, 1989) were shown to be beneficial for learning. In the field of tutoring, research has shown that dialogue moves of tutors depend on their assessment of the tutee (Eugenio et. al., 2009; Chi et. al., 2008; Chi & Roy, 2010) and that they are predictive of better understanding by the tutee (D'Mello et. al., 2010). What is missing is a gaze indicator at the same temporal level as dialogues.

## The Domain: Program Comprehension and Eye Tracking

There have been studies in the past concerning eye-tracking and programming. Romero et. al. (2002) compared the use of different program representation modalities (propositional and diagrammatic) in a expert novice debugging study where experts had a balanced shift of focus among the different modalities. Sharif et. al. (2012) emphasized the importance on code scan time in a debugging task and conclude that experts perform better and have shorter code scan time. Bednarik & Tukiainen (2006) examined coordination of different program representations in a program understanding task where experts concentrated more on the source code rather than looking at the other representations. Hejmady & Narayanan (2012) compared the gaze shift between different AOIs in a debugging IDE and concluded that good debuggers were switching between code and the expression evaluation and the variable window rather than code and control structure and the data structure window.

# Present Study and Questions

The present dual eye-tracking study examines the relationship between gaze, speech and performance in spatially distributed (remote) pair programming. We chose remote pair programming so that we can have two synchronized streams of eye-tracking data, which is difficult in the co-located pair programming (both programmers looking at the same screen). Baheti & Williams (2002) have shown that pair programming can be conducted remotely without negative effects on performance. We use two synchronized eye-trackers to study the gaze of two persons who have to read, understand, and explain functionality of a JAVA program.

## Methodological Question

The present study identifies different time scales to characterize interaction. Our working hypothesis is that it is necessary to define a gaze measure at each level to reflect corresponding cognitive processes. Indeed, measuring gaze at a global task level does not inform about dynamics of interaction and measuring gaze at the operations level reflects perception more than collaboration, elaboration and dialogue. Hence, our methodological question is what gaze measure reflects the dynamics of dialogue?

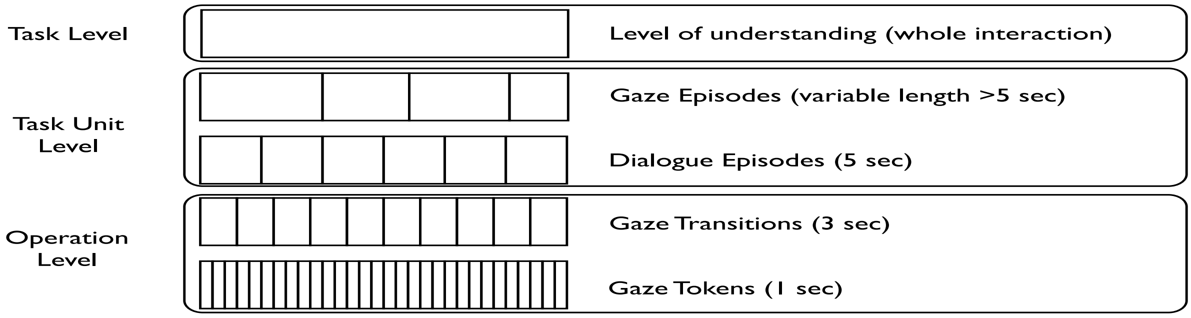


Figure 1. Interaction of the pair divided into different levels of time granularities.

We define gaze measures on two levels.

- On the task unit level, gaze episodes correspond to moments characterized by a stable gaze togetherness and gaze focus. For example, in a focused/together episode, programmers look together at a limited set of objects. These episodes typically last from 5 seconds up to 100 seconds.
- On the operations level we use gaze transitions among different set of objects. The transitions are based on a segmentation of gaze into 1-second slots and last for 3 seconds.

We define cognitive measures on two levels:

- On the task level we rated the level of understanding based on the explanations that were provided by the participants.
- On the task unit level we categorized the dialogues of participants depending on whether they were task related (describing the program) or whether they were about managing the task.

### Research Questions

The answer to the methodological challenge allows for new research questions to be asked about the relationships between two consecutive levels of time granularity:
**Question 1: task level and task unit level:** How does the level of understanding relate to the prevalence of different gaze episodes?
**Question 2: task unit level:** How do the types of gaze episodes relate to the types of dialogue episodes?
**Question 3: task unit level and operation level:** How do different dialogue episodes relate to the different gaze transitions?

## Experiment

In the experiment, pairs of subjects had to solve two types of pair programming tasks. The task consisted of describing the rules of a game implemented as a Java program. The experimental data used for this paper is the same as used in Nüssli (2011) and Jermann & Nüssli (2012), however the questions and analysis presented hereafter are completely different. Eighty-two students participated in the study. The participants were typical bachelor and master students aged from 18 to 29 years old with a median of 23 years old. The participants were paired into forty pairs without further consideration of their level of expertise, gender, age or familiarity. The subjects did a pretest that consisted of individually answering thirteen short programming multiple choice questions and then collaboratively solved the ten program understanding tasks which overall lasted approximately 45 minutes. Gaze was recorded with two synchronized Tobii 1750 eye-trackers that record the position of gaze at 50Hz in screen coordinates (see Figure 2). The interested reader can find technical details about the setting in Nüssli (2011).

### Gaze Tokens

The JAVA program is composed of tokens (see Figure 2, bottom-left). For example, a line of code "location = array [ c ] ;" contains 13 tokens ('location', 'c, '=', 'array', ';', 2 brackets and 6 spaces). Fixations on the individual tokens are detected using a probabilistic model (for details see Nüssli (2011)). We categorized the program tokens into 3 gaze tokens: **Expression**, **Structural** and **Identifier**. Each second of the interaction is categorized as one of the gaze tokens (based on the maximum probability).

### Gaze Transitions

We aggregated three consecutive gaze tokens into following three categories (see Sharma (2012)):
**Expression**: if all the three gaze tokens are expressions.
**Data Flow**: if there is a permutation of expressions and identifiers.
**Read**: if there is a permutation of all the three gaze token categories.

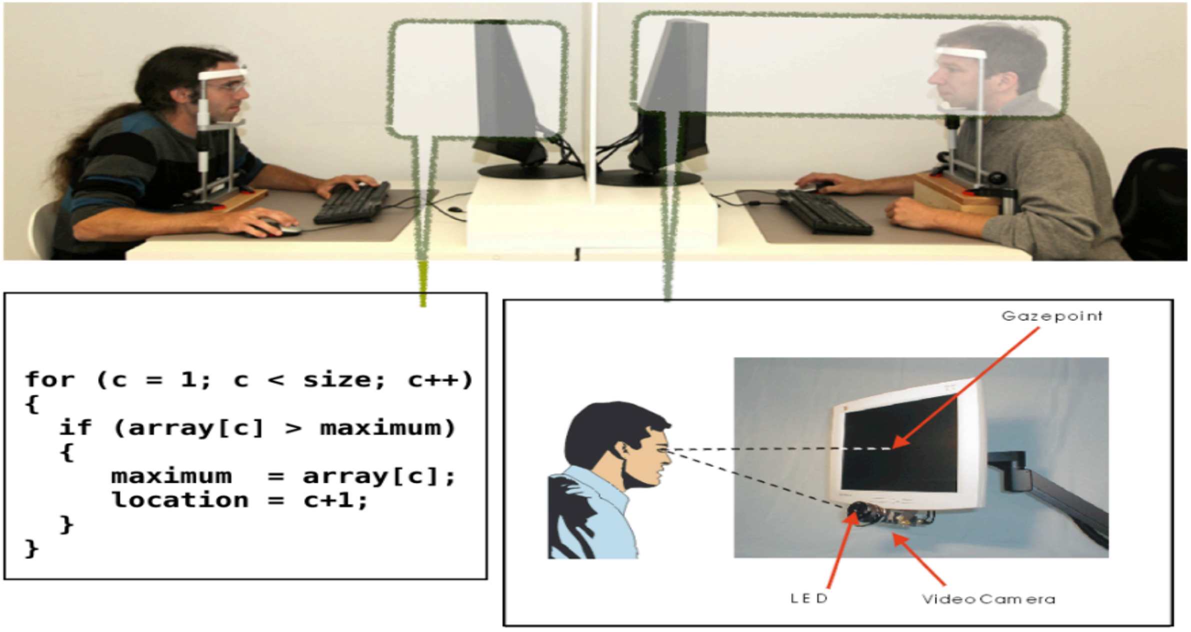


Figure 2. Apparatus used for the experiment. Upper half shows the laboratory setup. The left bubble depicts the stimulus and the right bubble depicts the eye tracking setup.

### Dialogue Episodes

We divided the dialogues into 2 major categories according to the content of dialogues. The first category comprises the dialogues containing the description of program functionality; and the second category contains task management utterances, for example, when participants talk about how to proceed, as well as about the controls of interface or where they should look next. Accordingly, we named the two categories as **"description"** and **"management"** respectively.

### Gaze Episodes

The gaze episodes are identified based on two parameters: the visual focus of gaze of the participants and the similarity of their gaze. In order to characterize the visual focus of one subject, we compute the object density vector over a given time window. This density vector contains the probability of looking at the different objects of the stimulus. In order to compute this vector, we aggregate gaze data over a 1-second time window and we compute for each object the amount of gaze time that was accumulated inside the object.

We then define the visual focus size as the numbers of objects that are looked at during a 1-second time frame. The rationale is to distinguish between moments where subjects look essentially at few objects versus moments where they look more or less uniformly at several objects. In order to get a quantitative indicator of this focus size, we compute the entropy of the density vector. Entropy measures the level of uncertainty of a random variable, which is, in our case, the objects looked at by the subjects. Hence, high entropy indicates that the subjects looked at many objects (*not focused gaze*), while low entropy indicates that they mostly looked at few objects (*focused gaze*).

For each 1-second timeframe, we define the visual focus coupling as the similarity between the objects looked by one subject and the objects looked by the second subject. We quantify this coupling by computing the cosine between the gaze density vector of one subject and the gaze density vector of the other subject.

Episodes are obtained by combining focus size and similarity. An episode lasts as long as the focus size and similarity stay constant. Technically, a run length encoding procedure applied on the 1-second indicators for visual focus and similarity obtains this. When both subjects are focused and similar we define **"focused together"** gaze episodes. Similarly, we define three other types of gaze episodes that are: **"not focused together"**, **"focused not together"**, **"not focused not together"**. Since we are mostly interested in *"what happens during moments of high collaboration?"* we report only what happens in "together" episodes (i.e., **"focused together"** and **"not focused together"**). Typically, a **"focused together"** episode translates in terms of behavior as putting joint efforts to understand code while a **"not focused together"** episode translates as an effort to search some piece of code.

### Level of Understanding

We distinguish between two levels of understanding based on how well they performed the description task. Pairs with **high level of understanding** are able to describe the rules of the game along with initial situation, valid moves and winning conditions. Pairs with **low level of understanding** only describe partial aspects of the game structure, and often give algorithmic descriptions of the program and try to guess the detailed rules from the method names; but they failed to get the winning condition.

## Results

### Question 1: Understanding and Gaze Episodes

The first question concerns the relation between the level of understanding attained by the pair and proportion of time spent by the pair in different gaze episodes. Table 1 shows the ANOVA results for gaze episodes "focused together" and "not focused together" across the two levels of understanding. Pairs with high level of understanding spend more time in gaze episode "focused together" than the pairs with low level of understanding ($F_{[1,16]}=8.70, p=0.01$). Figure 3 shows the difference interval for the two types of gaze episodes across the levels of understanding.

Table 1: ANOVA results for different gaze episodes across two levels of understanding.

| Episode Type | Df1 | Df2 | Sum Sq. | F-value | p-value |
|---|---|---|---|---|---|
| Focused Together | 1 | 15 | 0.09 | 8.70 | 0.01 |
| Not Focused Together | 1 | 15 | 0.06 | 10.60 | 0.005 |

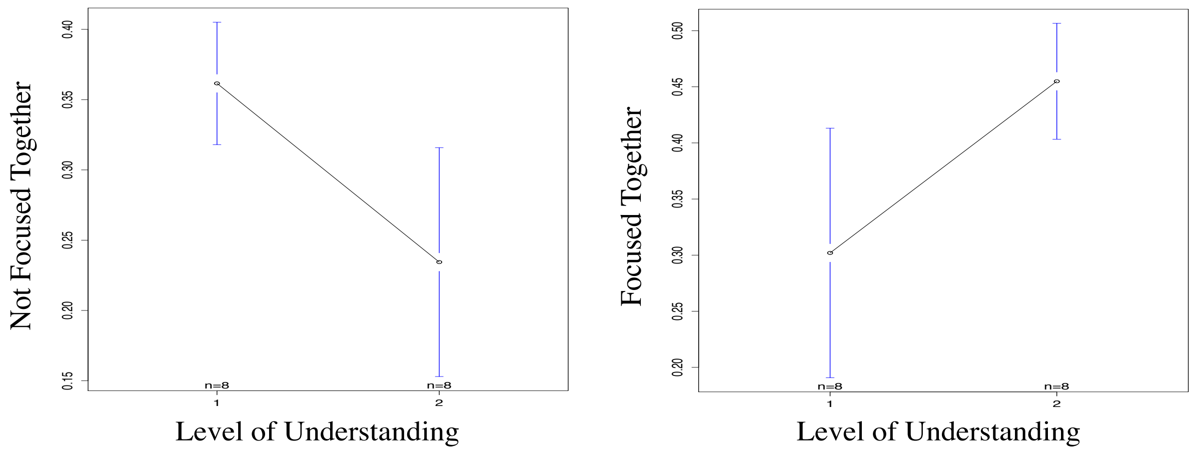


Figure 3. Difference Margin for focused together and not focused together gaze episodes for different levels of understanding.

### Question 2: Gaze Episodes and Dialogue Episodes

The second question addresses the relationship between the gaze episodes and the dialogue episodes. Table 4 shows the mixed effect model for the two types of dialogue episodes with the factors level of understanding and gaze episodes. There is no significant difference between the proportion of total time spent in dialogue episodes and the gaze episodes, but, there is a significant interaction effect of level of understanding and gaze episodes on the proportion of total time spent on the different dialogue episodes ($F_{[1,61]}=7.60$, $p=0.01$, Figure 4).

Table 2: Mixed effect model for dialogue episodes with factors level of understanding (UND) and gaze episodes (EPGAZE) (NS= Not Significant).

| | Dialogue Episodes | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Description Episodes | | | | Management Episodes | | | |
| Model | Df | Sum Sq. | F-value | p-value | Df | Sum Sq. | F-value | p-value |
| UND | 1 | 0.05 | 2.46 | NS | 1 | 0.01 | 1.56 | NS |
| EPGAZE | 1 | 0.04 | 1.71 | NS | 1 | 0.01 | 0.52 | NS |
| UND * EPGAZE | 1 | 0.17 | 7.80 | 0.009 | 1 | 0.07 | 7.60 | 0.01 |

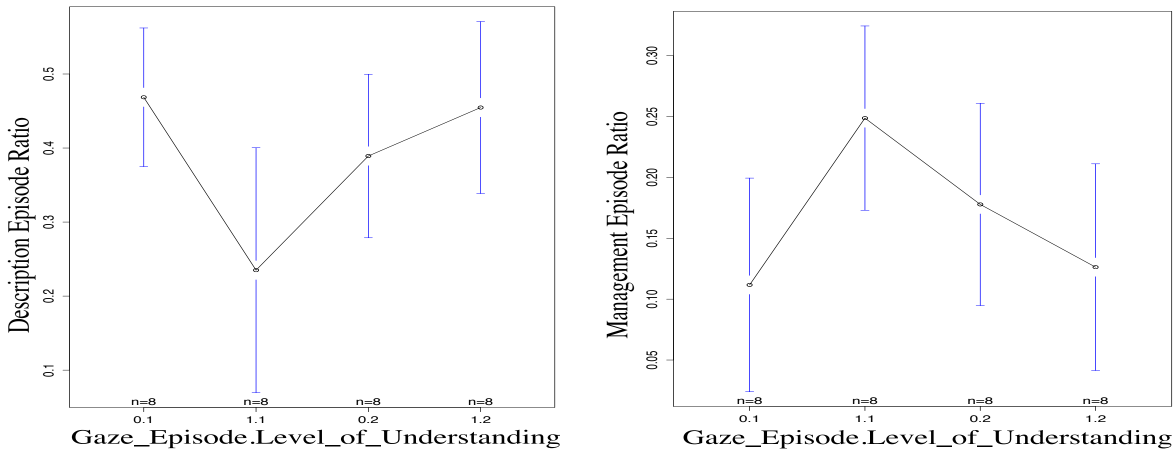


Figure 4. Interaction effect on DESC and MGMT dialogues in focused together and not focused together gaze episodes for different levels of understanding

The pairs with high level of understanding spend more time in "description" dialogue episodes when they are in a "focused together" gaze episode. On the other hand, pairs with low level of understanding spend more time on "management" dialogue episodes when they are in a "focused together" gaze episode. Figure 5 shows the dialogue snippets for pairs with different levels of understanding during different gaze episodes.

## Question 3: Dialogue Episodes and Gaze Transitions

The third question considers the relation between the dialogue episodes and the gaze transitions. Table 3 shows the ANOVA results for different gaze transitions across different dialogue episodes. "Description" dialogue episodes have more gaze transitions as "expressions" than the "management" dialogue episodes. Moreover, "management" dialogue episodes have more gaze transitions as "read" than the "description" dialogue episodes. The differences are irrespective of the level of understanding or the type of gaze episodes. Figure 6 shows the difference intervals for the two gaze transition categories across the dialogue episodes.

Table 3: ANOVA (repeated measures) results for different gaze transitions against dialogue episodes.

| Transition Type | Df1 | Df2 | Sum Sq. | F-value | p-value |
|---|---|---|---|---|---|
| Expressions | 1 | 63 | 0.51 | 8.79 | 0.004 |
| Read | 1 | 63 | 0.45 | 8.31 | 0.005 |

S2: I am looking for checkForWinner... the checkForWinner calls the checkForSum function for all i1, i2, i3.

(a)

S1: look here at choice...
S2: but we don't know where getPlayerMove is...
S1: where is getPlayerMove?
S2: look here choice is getPlayerMove.

(b)

S1: we said before, in order to be a valid action the player should choose a number which is valid, so from 1 to 9... if initial state or he should choose the number from the available list.

(c)

S1: we should look at the current situation
S2: currentGameState...
S1: no, no... let's check the checkForWinner function

(d)

Figure 5. Dialogue snippets for pairs having different levels of understanding during different gaze episodes to show the differences between verbal communications among the participants in the pairs. (a) Low level of understanding, focused together. (b) Low level of understanding, not focused together. (c) High level of understanding, focused together. (d) High level of understanding not focused together.

## Discussion and Conclusion

We conducted the present study with a two-fold motivation. First, identifying gaze and dialogue indicators at different time scales in a pair program comprehension task. Second, bridging different levels of time scales to demonstrate the relationship between gaze and group cognition.

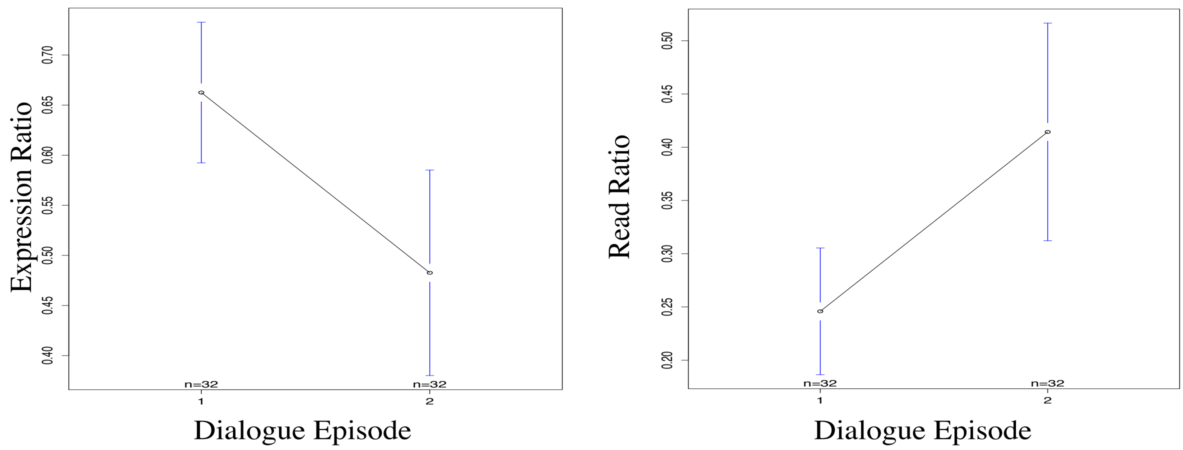


Figure 6. Difference Margin for expression and read gaze transitions for different dialogue episodes.

Concerning the methodological challenge, we have proposed gaze episodes as a description of the gaze of a pair on a task unit level. This measure is task independent and can be applied in a wide range of situations. For example, it could be used to describe the focus and similarity of gaze in a concept-mapping task, or in any text reading task. The level of detail for focus and similarity can be varied depending on the accuracy of the eye-tracker and depending on the task. With low-end eye-trackers, one could measure paragraph level, whereas with high-end machines, similarity can be measured at the word base.

Concerning the bridge between two consecutive time scales, we analyze each pair of time scales (see section "Present Study and Questions" and "Results"). We observed that the pairs with high level of understanding spend more time being "focused together" (see subsection "Understanding and Gaze Episodes") and while they are "focused together" the participants in the pair explain the functionality of the program to each other (Figure 5 (c)). When the pairs with high level of understanding are "not focused together" they talk about their next steps in the task (e.g., they talk about where to look next, Figure 5 (d)). On the other hand, pairs with low level of understanding exhibit the opposite behavior as they spend more time being "not focused together" (see subsection "Understanding and Gaze Episodes"). Moreover, while the pairs with low level of understanding are ``focused together'' they talk about managing their focus and when they are "not focused together" the participants explain to each other a small part of the functionality of program to maintain a shared focus. Based on our observations, we think that this reflects different ways to understand the program. The "focused" way consists of explaining in depth the functionality of the program, whereas the "unfocused" way consists of describing the code to the partner and to "travers" the code together.

One important observation is the interaction effect of level of understanding and gaze episodes on the type of dialogues (see subsection "Gaze Episodes and Dialogue Episodes"). There is no direct relation between the gaze episodes and dialogue episodes. This observation is similar to one that we found in another contribution (*not cited here for anonymity*). However, we see a direct relation between gaze indicators at the level of operations and dialogues. Irrespective of the level of understanding, the pairs have a higher proportion of "expressions" gaze transitions within "description" episodes. Moreover, the pairs have a higher proportion of "read" gaze transitions within "management" episodes. A possible explanation to this observation is that within a "description" episode the participants are more concerned with "what the program does?" This piece of information is contained in expressions within the programming constructs and hence the participants spend their time on understanding the expressions. On the other hand, within a "management" episode participants are talking about where to go next of they are searching a particular piece of code hence the gaze of participants is as if they are scanning the code like English text.

In a nutshell, we showed that there is a relationship between gaze and dialogue indicators at different time scales. These relations help us understand the cognition that underlies program comprehension as well as the collaboration that underlies pair programming. The results are interesting enough to pursue further research in the same direction to find the causality between processes at different time scales.

## References

Allopenna, P. D., Magnuson, & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models* 1,* 2,* 3,* 4,* 5. Journal of memory and language.

Anderson, J.R. (2002). Spanning seven orders of magnitude: A challenge for cognitive modeling. Cognitive Science. 26(1), 85–112.

Baheti, P., & Williams, L., (2002). Exploring pair programming in distributed object-oriented team projects. In

Proceedings of XP/Agile Universe.
Bednarik, R., Myller, R., Sutinen, E., & Tukiainen, M., (2006). Program visualization: Comparing eye-tracking patterns with comprehension summaries and performance. 18th Psychology of Programming Workshop.
Bednarik, R., & Shipilov, A., (2011). Gaze cursor during distant collaborative programming: A preliminary analysis. Proceedings of the DUET.
Bednarik, R., & Tukiainen, M., (2006). An eye-tracking methodology for characterizing program comprehension processes. In Proceedings of ETRA'06.
Cherubini, M., Nüssli, M.-A., & Dillenbourg, P., (2008). Deixis and gaze in collaborative work at a distance (over a shared map): a computational model to detect misunderstandings. In Proceedings of ETRA'08.
Chi, M., & Roy, M., (2010). How adaptive is an expert human tutor? In Intelligent Tutoring Systems, 2010.
Chi, M., Roy, M., & Hausmann, R., (2008). Observing tutorial dialogues collaboratively: In- sights about human tutoring effectiveness from vicarious learning. Cognitive Science, 32(2).
Cohen, E. G., (1994). Restructuring the classroom: Conditions for productive small groups. Review of educational research, 64(1), 1–35.
D'Mello, S., Olney, A., & Person. N., (2010). Mining collaborative patterns in tutorial dialogues. Journal of Educational Data Mining, 2(1), 1–37.
Eugenio B. D., Fossati, D., Ohlsson, S., & Cosejo, D., (2009). Towards explaining effective tutorial dialogues. In Annual Meeting of the Cognitive Science Society.
Griffin, Z.M., & Bock, K., (2000). What the eyes say about speaking. Psychological science, 11(4).
Hejmady, P., & Narayanan, N. H., (2012). Visual attention patterns during program debugging with an IDE. In Proceedings of ETRA'12.
Jermann, P., Soller, A., & Muehlenbrock, M., (2001). From mirroring to guiding: A review of the state of art technology for supporting collaborative learning. In Proceedings of EuroCSCL-2001, 324-331.
Jermann, P., & Nüssli, M.-A., (2012). Effects of sharing text selections on gaze cross-recurrence and interaction quality in a pair programming task. In Proceedings of CSCW '12, 1125–1134.
Lord, R. G., & Levy, P. E., (2008). Moving from cognition to action: A control theory perspective. Applied Psychology, 43(3), 335–367.
Magnisalis, I., Demetriadis, S., & Karakostas, A., (2011). Adaptive and intelligent systems for collaborative learning support: A review of the field. IEEE Transactions on Learning Technologies, 4(1), 5–20.
Meyer, A. S., Sleiderink, A. M., & Levelt, W. J. M., (1998). Viewing and naming objects: Eye movements during noun phrase production. Cognition, 66(2).
Newell, A., (1994). Unified theories of cognition, volume 187. Harvard University Press.
Nüssli, M.-A., (2011). Dual-Eye Tracking Methods for the Study of Remote Collaborative Problem Solving. PhD thesis, Ecole Polytechnique Federale de Lausanne.
Pietinen, S., Bednarik, R., Glotova, T., Tenhunen, V., & Tukiainen, M., (2008). A method to study visual attention aspects of collaboration: eye-tracking pair programmers simultaneously. In Proceedings of ETRA'08.
Pietinen, S., Bednarik, R., & Tukiainen, M., (2010). Shared visual attention in collaborative programming: a descriptive analysis. In Proceedings of the ICSE Workshop on Cooperative and Human Aspects of Software Engineering.
Richardson, D. C., & Dale, R., (2005). Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. Cognitive Science, 29(6).
Richardson, D. C., Dale, R., & Kirkham, N. Z., (2005). The art of conversation is coordination. Psychological Science, 18(5).
Romero, P., Lutz, R., Cox, R., & Boulay, B., (2002). Co-ordination of multiple external representations during java program debugging. In Human Centric Computing Languages and Environments, 2002. Proceedings. IEEE 2002 Symposia on, 2002.
Sharif, B., & Maletic, J.I., (2012). An eye tracking study on camel case and under score identifier styles. In Proceedings of 18th International Conference on Program Comprehension.
Sharif, B., Michael Falcone, & Maletic, J.I., (2012). An eye-tracking study on the role of scan time in finding source code defects. In Proceedings of ETRA'12.
Sharma, K., Jermann, P., Nüssli, M.-A., & Dillenbourg, P., (2012). Gaze Evidence for different activities in program understanding. 24th Psychology of Programming Workshop.
Stahl, G., (2000). A model of collaborative knowledge-building. In Proceedings of 4th International conference of the learning sciences, 70–77, 2000.
Webb, N. M., (1989). Peer interaction and learning in small groups. International journal of Educational research, 13(1), 21–39.
Zelinsky, G. L., & Murphy, G. L., (2000). Synchronizing visual and language processing: An effect of object name length on eye movements. Psychological Science, 11(2).