

Using dual eye-tracking to unveil coordination and expertise in collaborative Tetris

Patrick Jermann
CRAFT

Patrick.Jermann@epfl.ch

Marc-Antoine Nüssli
CRAFT

marcantoine.nuessli@epfl.ch

Weifeng Li
CRAFT

Weifeng.Li@epfl.ch

Ecole polytechnique fédérale de Lausanne (EPFL), Switzerland

The use of dual eye-tracking is investigated in a collaborative game setting. The automatic collection of information about partner's gaze will eventually serve to build adaptive interfaces. Following this agenda, and in order to identify stable gaze patterns, we investigate the impact of social and task related context upon individual gaze and action during a collaborative Tetris game. Results show that experts as well as novices adapt their playing style when interacting in mixed ability pairs. We also present machine learning results about the prediction of player's social context.

Eye-tracking, awareness tools, machine learning, coordination, expertise

1. INTRODUCTION

1.1 Dual eye tracking and HCI

Dual eye-tracking (DUET) is a novel methodology that consists of synchronously recording the gaze of two collaborators working on a shared application. This approach leads to new insights about the collaboration processes occurring in computer supported collaborative work and during other synchronous interactions, for instance team gaming. This knowledge about interaction is of importance for the HCI community as it provides clues on how to design efficient interfaces to sustain fruitful collaboration.

Moreover, as eye-trackers are becoming more reliable and cheaper, there is great chance that within one decade, all personal computers will be able to record and analyse user's gaze. This gives software interaction designers the opportunity to use gaze based information to build more efficient and more intuitive user interfaces. This is of course also true for collaborative applications which could take advantage of dual-gaze patterns to offer efficient communication facilities or to provide meaningful feedback about the part-ner(s) and/or about the collaboration in general.

The main approach in the field of gaze sensitive applications consists of building mouse-replacement or mouse-enhancement systems [20] in which the gaze is used as a pointing device, either for disabled people or to offer more efficient pointing mechanisms. While this is surely a valid approach, especially for disabled people, there are some criticisms concerning the use of the gaze for direct control. First, the eyes are a perceptive device and not a motor device in the sense that their primary purpose is to acquire

information [20]. A second problem is the inaccuracy of gaze data, mainly because of tracking errors¹ but also because of intrinsic properties of eye-movements due to physiological constraints.

We take a somehow different approach regarding the development of gaze-sensitive applications. The central idea is to take advantage of existing patterns in eye-movements to provide users with task specific feedback. This requires the development of algorithms able to interpret eye-movement patterns and to make predictions about the user or about the processes in which the user is engaged in. These predictions could then be used to give higher-level psychologically meaningful feedback, or to adapt the interface given the current activity of the user. This approach is currently pioneered in the field of adaptive user interfaces (e.g. the Text 2.0 framework [1]).

More specifically, we are interested in improving remote collaboration and thus, in building tools that give feedback about the state of the collaboration, such as so-called aware

¹This is especially true for remote, camera-based eye-trackers and such devices are the most likely to be used in end-user gaze sensitive applications.

ness tools([6]) or group mirrors([8]). There are some tentative works about gaze-awareness tools to improve collaboration. For example, RealTourist [12] allows a remote tourist consultant to see tourist's gaze on a shared map while helping remotely to plan a trip. Also, Brennan [3] did a study where two people had to solve a visual search task on remote computers (e.g. subjects have to search for a letter O hidden in a cloud of Q letters). They showed that they were much faster in solving the problem if they could see the gaze of their partner, which suggests that gaze is a good clue to coordinate

parallel activities. Finally, various systems, such as Clearboard [7] or GAZE groupware ([18]) have been designed to give access to collaborators to the gaze directions of others, similarly to a face-to-face situation. All these systems rely on a simple technique that consists of displaying the last few seconds of gaze of one collaborator to the others. We have piloted this approach in an experiment on collaborative Tetris and came to the conclusion that even if it is not necessarily detrimental, feedback can be quite disturbing because it interferes with the main task representation and because it is not always precise enough to support deictic references. Finally, users are not trained to interpret minimally aggregated eye-movements patterns.

The provision of higher level feedback about collaboration

(e.g. about the quality of the communication) relies on the constitution of a computational model that can be used to diagnose the current state of interaction. Cherubini [4] has made a step in this direction by developing an algorithm able to predict, up to certain point, the occurrences of misreferencing between two people annotating a shared map. Another tentative is the REGARD system [14], which monitors dual-gaze and speech in real-time in order to learn to associate words to objects on the screen.

The first step in the development of intelligent gaze-aware tools is the discovery of stable meaningful patterns in the gaze of two people collaborating remotely. Our approach is to use machine learning techniques to discover such patterns and to use them for predictions. Our work is at the frontier of engineering -make reliable predictions from temporal signals-and cognitive science -understanding the relationship between gazes and collaboration. The present work has both a psychological aspect about the relationship between gaze and collaboration processes, and a technology-oriented aspect about how gaze data may be used to feed machine learning algorithms in order to predict aspects of interaction.

1.2 Research Questions

We conducted the current study about Tetris, a well known arcade game where players have to place falling tetrominoes (also called zoids) on a stack. Tetris is particularly well suited for eye-tracking because it is a sensori-motor control task that does not require off-screen deliberations. It is therefore very likely that there is a strong link between eye-movements and the cognitive processes involved in solving the task (see [9] for a detailed analysis). In the collaborative version of the Tetris game that we developed, two players share the game field and each controls their own zoid. One player's zoids appear on the left side and the other's appear on the right side of the game field.

The first question concerns the effect of collaborating with another player. More specifically we are

interested in the social context, namely to investigate how the play with a more or less expert player influences behavior. Previous research has shown that experts and novices do not look the same way at the game [16]. Experts produce more lateral gaze sequences than novices while there are no difference between novices and experts with regards to vertical scan path, or fixation durations. Do experts adapt their behavior when interacting with novices? Do novices copy experts' performance and adopt a more expert way of looking at the game? Beyond Tetris, the interaction of collaborators with different levels of expertise is of a particular interest as it is known to play an important role when learning through social interactions ([19]). Learning is indeed facilitated by the presence of a more expert peer or parent who is able to guide the discovery of new concepts.

The second question concerns the coordination of collaborative activity. The benefits of collaboration (distributing cognitive load, combining areas of expertise, etc.) come at the cost of coordination. In Tetris, playing together gives rise to situations where the players have to coordinate their moves, either because players want to place their zoid on the other players' side, or because they are in competition for the best position available on the stack. How are players handling coordination? Is it possible to distinguish coordination from task completion activities? How can we use the response to these questions to design gaze aware tools that facilitate coordination?

The third question asks whether the differences found in gaze and actions data between various expertise roles may be used to actually predict the roles of the players. In this respect, we want to compare the predictive power of different machine learning techniques and of using different features, such as only actions, only gazes or a combination of both.

2. METHOD

2.1 Task Description

We implemented a multiplayer, collaborative, version of the classical Tetris game. In this version, two pieces fall at the same time in the same game area, one on each side of the area. Each player can control one of the two pieces (either the left piece or the right piece) and they can execute the usual Tetris movements: left or right translations, rotations, down accelerations and drop. Of course, the game prevents collisions between the two pieces, thus producing novel situations compared to the original game. For example, if a player drops a zoid while the second player has its zoid right below, the drop of the first player's zoid will be stopped just above the

blocking zoid. Note that, like in the given example, the two pieces do not necessarily fall synchronously as some players may play their zoids faster than their partner. The game is collaborative and thus, both players share the same score independently of who plays what.

2.2 Procedure

The subjects played on two computers placed side by side and separated by a shelf in order to prevent them to look at each other. However, they could freely talk to each other.

We recorded eye-movements of both players using two Tobii 1750 eye-trackers. All actions (translations, rotations,

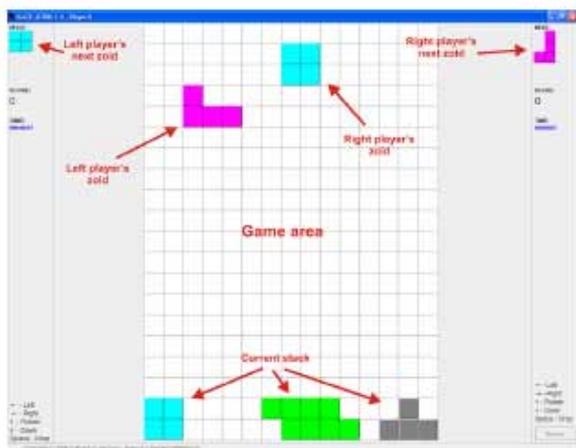


Figure 1: Screenshot of the collaborative version of Tetris.

etc..) and game state changes (game stack variations, scores, pieces position, etc..) were also logged by the Tetris program. All these data (eye-movements and Tetris actions and states) were recorded on a separate computer in order to synchronize the timestamps of the two eye-trackers along with the timestamps of the Tetris application. Subjects first played normal Tetris game individually in a training phase for 5 minutes. The score from this training phase was used to determine the level of expertise of the player. Then, they played in pairs for 10 minutes in a collaborative training phase and finally, they played again together for 15 minutes for collaborative testing phase.

2.3 Social context

We split the players into novices and experts by a quartile cut of the subjects' score during the individual training phase. The 25% lowest scoring players were assigned to the novice (N) group and the 25% highest scoring to the expert (E) group. In other terms, we have one group composed of the 25% worst players and one group with the 25% best players.

The social context is the combination of the players' level along with their partner's level. The four possible combinations are NN for a novice playing

with another novice, NE for a novice playing with an expert, EN for an expert playing with a novice, and EE for an expert playing with another expert. Out of the 2829 zoids that we analysed, 744 (26%) were played in EE pairs, 788 (28% EN) plus 655 (23% NE) were played in mixed ability pairs, and 642 (23%) were played in NN pairs.

2.4 Game context

The game context represents the state of the game during the playing of one specific zoid. Two variables are particularly interesting with regard to our second question as they identify situations that require active coordination by the players.

• Conflict

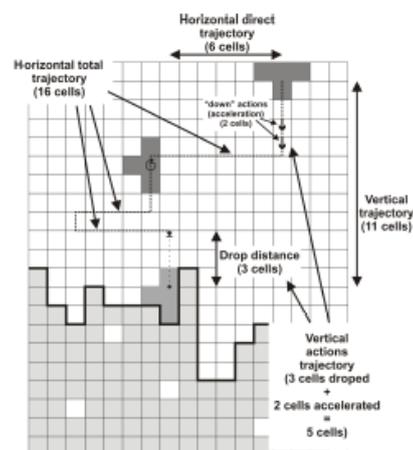


Figure 2: Schema depicting Tetris complex action features. This is an example of a possible trajectory for a zoid.

This is a score between 0 and 1 assessing how much the two players have to place their zoid at the same place. Conflict is computed in two steps. First, we computed the number of best fit positions for each player, a best fit position being defined as a position in which the zoid can be placed without leaving a hole below it. Secondly, we computed the number of common best fit positions by comparing the best fit positions of the two players. Finally, we applied the following formula to compute the conflict level of one player:

$$\text{conflict} = (1 - \text{self}/15) + (\text{common}/\text{self}) \quad (1)$$

where self is the number of best fit positions for the player and common is the number of positions which are best fits for both players. Note that the constant value 15 corresponds the horizontal number of cells in the grid. Thus, conflict value will be high if the player has few best fit positions and/or if her best fits are the same than the ones

of her partner. It is also important to note that this score is not symmetric as it depends on the number of best fits of the concerned player which can be potentially different than for her partner.

- *Crossing*

This is a simple boolean value indicating if a crossing between the two players (for example, the left player going on the right side of her partner) occurred. Thus, it allows to identify episodes which required a certain amount of coordination between the two players.

2.5 Features

2.5.1 Actions

In addition to the simple players actions counts, we defined several more advanced features which represent the gaming style more accurately. These features are defined below and the most complex ones are summarized on Figure 2.

- *Rotations, translations, drops, downs*

These are simple counts of the five Tetris actions that players can perform. Left and right translations have been grouped together under the name translations as there were no reason to distinguish between them.

- *Horizontal direct trajectory and horizontal total trajectory*

These features describe the trajectory of the zoid on the horizontal axis. The horizontal direct trajectory is simply the horizontal distance between the zoid initial position and its destination, while the horizontal total trajectory represents the horizontal distance actually travelled by the zoid (see figure 2, direct trajectory is 6 cells and total trajectory is 16 cells)

- *Horizontal directness*

This feature represents how direct the horizontal positioning of the zoid is. In other terms, it measures how many superfluous horizontal translations the players did. It is computed by dividing the horizontal direct trajectory by the horizontal total trajectory. Thus, it is the proportion of the total (actual) trajectory which was really useful in order to place the zoid in its final position. In the example of the figure 2, the horizontal directness is $6/16 = 0.375$

- *Vertical trajectory and vertical actions trajectory*

These are the vertical counterparts of the horizontal trajectories defined above. However, things are slightly different because vertical moves occur

also automatically (zoids fall down if the player does nothing). Moreover, it is not possible to go upward thus preventing any superfluous vertical moves. We have defined the vertical trajectory in the same way than the horizontal direct trajectory : it is the vertical distance between the zoid initial and final positions. The vertical actions trajectory represents the vertical distance travelled by doing down actions (accelerations) or by dropping the zoid (see figure 2, trajectory is 11 cells and actions trajectory is 5 cells)

- *Vertical acceleration*

This is somewhat similar to the horizontal directness but for the vertical movements. Here, we have computed the ratio of the vertical actions trajectory over the vertical trajectory. Thus, it is the proportion of the vertical distance travelled by doing accelerations (downs or drop), the complement of this value being the ratio of the direct vertical distance travelled by automatic vertical movements. In figure 2, the vertical acceleration is $5/11 = 0.454545$

2.5.2 Gaze

Gaze variables aim to describe mainly where the players were looking at and more generally, some characteristics of their eye-movements.

- *Hit on self, other and contour*

These three gaze variables are the ratio of the total fixation time spent looking at the three main objects of interest which are the player self's zoid, the other player's zoid and the contour of the game stack (See figure 3).

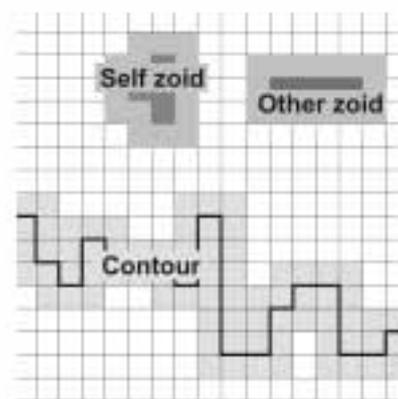


Figure 3: Schema depicting Tetris areas of interest. The light-gray areas shows the tolerance zone to consider a fixation as a hit on the corresponding area.

- Average fixation duration, average saccade length, ratio of horizontal and vertical saccades

Average fixation duration and average saccade

length are simple usual characteristics of the eye-movements. The ratios of horizontal and vertical saccades are the proportion of saccades being horizontal or vertical with tolerance of 20 degrees.

3. RESULTS

3.1 Descriptive statistics

The dataset contains a total of 2829 zoids played by 28 individuals. In this section we report descriptive statistics as well as statistical tests that illustrate the effect of different contextual variables on gaze and action features. Action features include rotations, vertical acceleration and horizontal directness. Gaze features include the ratio of time spent by the players fixating on the contour, on their own zoid or on their partner's zoid.

3.1.1 Social context

We fitted a series of linear mixed effect models to assess whether the social context influences action and gaze features (see table 1). Only two variables differ significantly among the four categories:

Vertical acceleration is key in Tetris. Rather than waiting until zoids fall down, expert players accelerate the game by downs and drops. Figure 4 shows that vertical acceleration is featured by EE and EN players to a similar extent. It is especially interesting to see the NE players accelerate the game more than NN players under the influence of their more expert partner. This is either due to direct tutoring ('e.g. 'drop it now') or simply to imitation by the novice.

Gaze on self is strongest for NN players. From figure 5 we see that novice players change the way they look at the game when interacting with experts. Indeed, NE players look at own zoid less than NN players (and spend the time looking more at the contour). We also see that, while the overall time ratio spent on other's zoids is very small, EN experts spend more time than any other players looking at their novice partner's zoid.

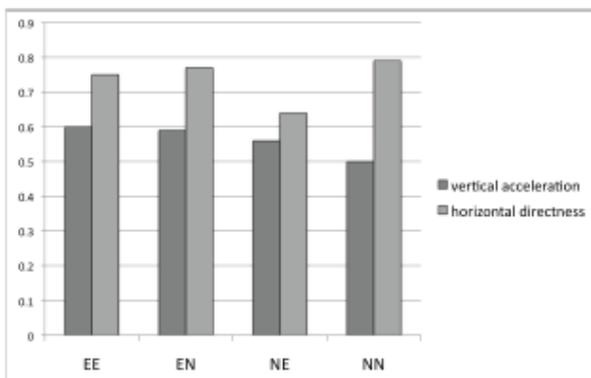


Figure 4: Action features.

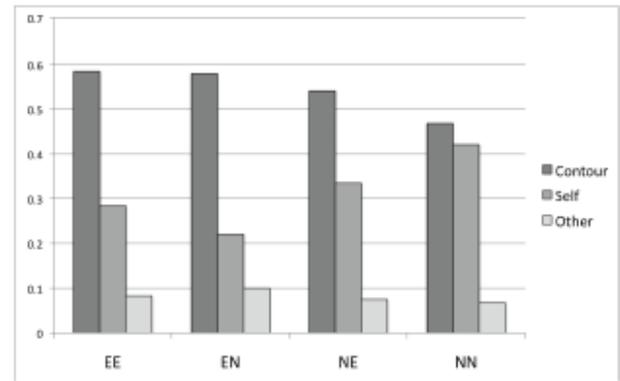


Figure 5: Time ratio spent fixating on the contour, self and other.

Table 1: Effect of social context on action and gaze features. All F-tests performed with 1 and 11 degrees of freedom. Numbers represent means.

Var	Pair Composition				ANOVA	
	EE	EN	NE	NN	F	p
Rotations	1.6	1.7	2.0	1.7	0.9	NS
Vert.	0.6	0.59	0.56	0.5	4.1	<.04
Horiz.	0.75	0.77	0.64	0.79	2.1	NS
Contour	0.58	0.57	0.54	0.45	1.9	NS
Self	0.29	0.23	0.34	0.43	4.8	<.02
Other	0.084	0.10	0.076	0.072	0.87	NS

3.1.2 Game context

Many different situations arise during the game. We identified two variables which allow us to contrast behavioral measures: conflict and crossings. Both variables have a social aspect: conflict results in players having the same goal (potentially leading to a competitive situation) and crossing results in players having to coordinate their actions to reach different goals.

A binary conflict level was defined by a median cut, with values of conflict above 0.23 corresponding to high conflict and values below the median corresponding to low conflict. We fitted a series of linear mixed effect models to assess the influence of conflict on the action and gaze features (see table 2). It appears that in case of conflict subjects play slower and do more translations moves. They also focus relatively more on the contour, less on their own zoid or their partner's zoid.

Table 2: Effect of conflict on action and gaze features. All F-tests performed with 1 and 2871 degrees of freedom. Numbers represent means.

Var	Conflict		ANOVA	
	High	Low	F	p
Rotations	1.7	1.8	3.5	NS
Vert. accel.	0.54	0.59	27.7	<.0001
Horiz. direct.	0.71	0.77	23.7	.0001
Contour	0.56	0.51	36.6	<.0001
Self	0.3	0.33	10.9	<.001
Other	0.076	0.094	13.21	.0003

Crossing happened during the play of 430 zoids (15%). This indicates that this phenomenon is rather

rare and that players mostly stay on their side of the playing field. The proportion of zoids played with crossing is slightly higher (17%) in mixed ability pairs (NE and EN) than in homogeneous pairs (12% in EE and NN). We conducted a series of linear mixed effect analyses of variance to investigate the effect of crossing on the action and gaze features (see table 3). While crossing, players rotate their zoid more often, vertically accelerate less, and make more horizontal moves. The coordination that is required when crossing zoids also impacts the fixation targets, as players fixate slightly less on the contour and slightly more on their partner's zoid.

Table 3: Effect of crossing on action and gaze features. All F-tests performed with 1 and 2871 degrees of freedom. Numbers represent means.

Var	Crossing		ANOVA	
	No	Yes	F	p
Rotations	1.7	2	5.4	<.05
Vert. accel.	0.58	0.5	23.1	p<.0001
Horiz. direct.	0.75	0.67	5.1	<.02
Contour	0.54	0.51	23.9	<.0001
Self	0.32	0.30	0.49	NS
Other	0.075	0.14	93.5	<.0001

3.2 Principal Component Analysis

We conducted a Principal Component Analysis (PCA) to further investigate the structure of behavioral variables. The analysis included five gaze related variables as well as three action related features. We retain two dimensions with eigenvalues higher than 1 for the presentation of the results. The first dimension accounts for 23% of the total variance, the second dimension accounts

for 20%. The correlations between the initial variables and the orthogonal dimensions of the PCA are represented in figure 6. The correlational structure uncovers two rather independent aspects of the behavior. We use the mapping of contextual variables onto the two components to guide the interpretation of the dimensions (see figure 7). Along with contextual variables, we represent the seven different zoid types (L, J, S, Z, I, O and T).

3.2.1 Dimension 1: zoid control versus solution search Two gaze features are strongly correlated with the first dimension. It represents the opposition between fixations on the players' own zoid (self) and fixations on the contour. The players fixate their own zoid when it first appears as well as when they "drive" it through rotations and translations down towards its final position. They fixate the contour to search for a "good" spot to place their zoid, as is suggested by the correlation of the ratio of horizontal saccades with the dimension. In figure 7, this dimension opposes novices (NN) and experts (EN and EE). Interestingly, NE players are situated to the left of NN players, indicating that when novices play along with an expert, they fixate their own zoid less than when they play with another novice. The contrast between high and low conflict is also represented on this dimension. Conflictual situations are solved by a more intense search for solutions on the contour. Finally, the zoids which are most fixated are L, J and T, the ones with the most possible orientations and hence the most rotations performed.

3.2.2 Dimension 2: speed versus coordination Vertical acceleration has the strongest correlation with the second dimension. Horizontal directness also correlates positively with the dimension, indicating that few superfluous translations are done. Rotations and fixations on the part-

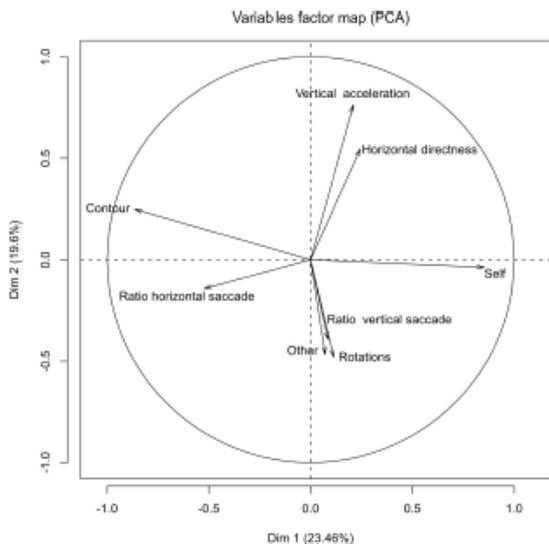


Figure 6: Correlation circle for the two first dimensions. The arrows represent the correlation between original variables and the dimensions of the PCA.

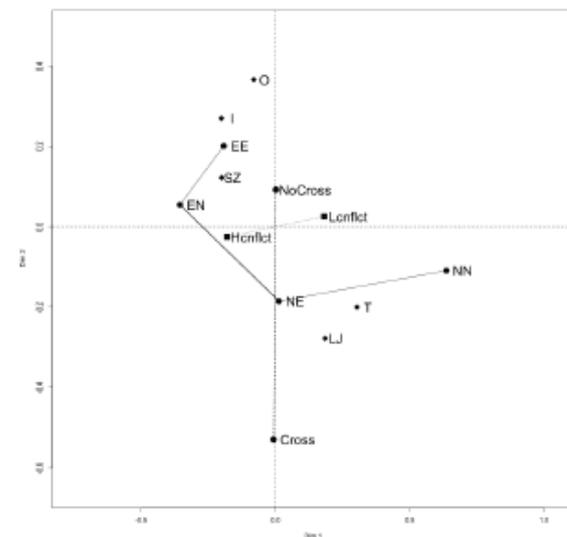


Figure 7: Centroids of the supplementary categorical variables.

ner's zoid correlate negatively with the dimension, indicating a slower pace of play. In figure 7 experts (EE and EN) are positioned higher on this dimension than novices (NN and NE), although it appears that EN experts slow down because they play with novices. The opposition between zoids with crossing and without crossing is most striking on this dimension and lends it the coordination flavour. We see that when players are crossing their zoids, the speed is lowered, and the fixations are higher on the partner's zoid. Finally, we see that I and O zoids are played the fastest whereas the L, J and T zoids appear lower on the dimension, mostly because of the time it takes to rotate them.

3.3 Automatic recognition of pair composition

We experimented with the prediction of the roles of the player and their contexts using both generative Gaussian Mixture Model (GMM) [2] and discriminative Support Vector Machines (SVM) [5]. GMM is a parametric probability density estimation technique which has been successfully used to deal with the speaker verification [13] and speaker recognition [11]. For a feature vector x , the mixture density is represented by the form

$$p(x|\Theta) = \sum_{k=1}^K w_k \mathcal{N}(x; \mu_k, \Sigma_k), \quad (2)$$

where $\mathcal{N}(x; \mu_k, \Sigma_k)$ is the k th unimodal Gaussian density with mean vectors μ_k and covariance matrices Σ_k , and w_k is the corresponding mixture weight. The GMM parameter set $\Theta = \{w_k, \mu_k, \Sigma_k\}_{k=1}^K$ is estimated over the all the feature vectors of the training examples. Here K denotes the number of the Gaussian mixtures.

In our experiments, each social context category (i.e. NN, NE, EN, and EE) is represented by such a GMM and is referred to as model Θ_c , $c = 1, \dots, C$. Here C denotes the number of social context categories. For a test vector x_i , the c -th GMM likelihood which can be written in the log domain can be calculated as

$$L(x_i|\Theta_c) = \log p(x_i|\Theta_c). \quad (3)$$

This calculation is performed across the C GMMs, and the one providing the largest log likelihood is found as the recognized social context class.

SVM belongs to the class of maximum margin based discriminative classifiers. They perform pattern recognition between two classes by finding a decision surface that has maximum distance to the closest points in the training set which are termed as support vectors. In our case, given a training set of examples x_j , $j = 1, 2, \dots, N$ with the labels y_j and a new test vector x_i , the decision value of a SVM has the form:

$$f(x_i) = \sum_{j=1}^N \alpha_j y_j K(x_i, x_j) + b, \quad (4)$$

where the coefficients α_j and b are the solutions of a quadratic programming problem [17] using the training data. Those training examples with $\alpha_j = 0$ are identified as support vectors. $K(x_i, x_j)$ is the kernel function that can provide SVM with non-linear abilities. In our experiments, the radial basic function kernel are used:

$$K(x_i, x_j) = \exp(-\gamma |x_i - x_j|^2), \quad \gamma > 0. \quad (5)$$

are used. SVM classifier utilizes the kernel trick [15] to find a linear separating hyperplane in a transformed, arbitrarily high-dimensional space, which is capable of capturing nonlinearities in the evidence space. For support vector machines are originally binary classifiers, Max-win strategy is used in our work for multi-class recognition. In our studies the penalty parameter C and γ in Eq. (5) is set experimentally.

	Action-based	Gaze-based	both
GMM			
CRR [%]	51.34	56.62	70.84
Kappa	0.34	0.42	0.63
SVM			
CRR [%]	54.45	60.12	75.28
Kappa	0.39	0.49	0.68

Table 4: Recognition results using different features in terms of correct recognition rate (CRR) and Kappa coefficient (the higher the better).

In our experiments, action-based features, gaze-based features, and their combination are used for the recognition of social context. The whole data for each social context group is divided by five folds, and five-fold cross validation are adopted for our evaluation.

Table 4 summarizes the recognition performance using different classifiers and different feature sets, in terms of correct recognition rate (CRR) and Kappa coefficient. It is shown that gaze-based feature set performs better than action-based feature set. This demonstrates that gaze-based features are more important for classifying different social contexts. Combining the two feature set significantly improves the recognition performance, which shows that the information from the two feature sets are complementary. It can be also found that SVM performs better than GMM. This might be explained by the fact that SVM is a more discriminative model than GMM which is based on the estimation of the probability density function (pdf) statistically.

4. DISCUSSION

Results show that players adapt their behavior to the social context of interaction. A possible explanation is that experts tutor novices by telling them where and how to place their zoid on the stack, thereby showing the novices the tricks of the trade. As a consequence, they play slower and monitor their partner's zoid

more than if they interacted with another expert. In parallel, novices who interact with experts look less at their own zoid and monitor the contour more than if they interacted with novices. Dialogue between players contain many coordination episodes (e.g. “move away!”), explicit evaluations of actions (e.g. “well done”) as well as strategic decisions on occasion (“We should make four lines at the time”). Explicit tutoring often takes the form of advice (e.g. “put your piece here”). However, not all pairs correspond to this somewhat idealized image of a tutor helping out a novice. In several cases, players remained silent during the whole game. In addition, tutoring is not limited to mixed ability pairs, so that we can’t simply equate tutoring and expert-novice interaction. A closer analysis of audio transcripts may shed more light on the inner workings of interaction and allow us to refine the analysis of what characterizes successful tutoring episodes in all social contexts. Our study presents several limitations. High and low ability players were considered as novices and experts. This definition of expertise is somewhat arbitrary and depends on the ability of the student population the subjects were drawn from. We did not collect indicators about learning in this study. It takes of course longer than half an hour to train someone to be an expert Tetris player. Hence, we observed tutoring interactions but cannot measure whether they led to sustainable and positive learning effects. Finally, our results and overall approach were obtained with a rather simple perceptual-motor game. We are eventually aiming at more deliberative activities, for example pair programming. Beyond Tetris, our results suggest that it is feasible to link behavioral features with attributes of social interaction. This is particularly interesting for the design of systems in the field of computer supported collaborative learning (CSCL). In the model of computer support for interaction regulation that we proposed [8], the provision of feedback to the collaborators relies on the diagnosis (by a computer or a human coach) of a computational model of interaction. The use of gaze features appears to be a promising option to enrich the automatic collection of indicators about collaborative learning. The results of the Principal Component Analysis that we conducted suggest that playing collaborative Tetris consists of two concurrent activities: playing the game and collaborating. The first dimension represents gaze features related to the resolution of the basic Tetris task (analysing
2In order to achieve accurate pdf estimation, generally large amount of training data is needed. the contour and controlling the zoid) while the second dimension represents the cost of coordinating zoid movements that could result in collisions (slowing down and observing the partner’s zoid). We have seen that the responses of the players to crossings and conflicts are different. Players handle crossings

by monitoring their partner’s zoid and avoiding each other whereas conflict is rather solved by exploring the contour for alternative positions. The game field was large enough for players to play side-by-side most of the time and only occasionally have to coordinate actions and plans. The imbalance between individual instrumental actions and socially oriented actions is common to many multiplayer games where coordination is only occasionally required [10] between players.

Beyond Tetris, our results suggest that it could be possible to detect coordination episodes based on gaze features. The signature for explicit coordination seems to be increased visual attention to the partner’s objects. The visual structure of collaborative Tetris is particularly well suited to measure behavioral cues that reflect either individual or coordination actions. In other situations, for example a shared text editor or a shared concept mapping tool, it might be more difficult to directly link fixations on areas of interest with the underlying dimension of ownership. We have observed that in situations like collaborative concept mapping collaborators develop a sense of ownership for the concepts they created, and rarely delete or modify concepts created by their partners. More work is required to deepen our understanding of the links between gaze, coordination and division of labor in a group.

An obvious application of the dual eye-tracking methodology is to build awareness tools that display the partners’ gaze upon request to support referential communication. Players simply press a key to display their gaze while verbally signalling a deictic reference (e.g. saying “here”). As a response, the system could display whether their partner looks at the place they are referring to. The possibility to computationally predict the social context (and quite probably the task related context as well) of the players allows to envisage the design of adaptive gaze awareness tools. For example, the system could display the position referred to by the expert as a zoid superimposed on the contour in the optimal orientation, thereby lightening the mental load of the novice. With an expert partner, the same reference would be signaled more simply by a simple visual marker.

5. REFERENCES

- [1] R. Biedert, G. Buscher, S. Schwarz, M. Moller, A. Dengel, and T. Lottermann. The text 2.0 framework. writing web-based gaze-controlled realtime applications quickly and easily. IUI 2010 Workshop on Eye Gaze in Intelligent Human Machine Interaction, 2010.
- [2] C. M. Bishop. Neural Networks for Pattern Recognition. Oxford University Press, 1995.
- [3] S. E. Brennan, X. Chen, C. A. Dickinson, M. B.

- Neider, and G. J. Zelinsky. Coordinating cognition: The costs and benefits of shared gaze during collaborative search. *Cognition*,106(3):1465 – 1477, 2008.
- [4] M. Cherubini, M.-A. Nüssli, and P. Dillenbourg. Deixis and gaze in collaborative work at a distance: a computational model to detect misunderstandings. In ETRA '08: Proceedings of the 2008 symposium on Eye tracking research and applications, pages 173–180, New York, NY, USA, 2008. ACM.
- [5] N. Cristianini and J. Shawe-Taylor. *An Introduction to Support Vector Machines*. Cambridge University Press, Cambridge, 1999.
- [6] C. Gutwin and S. Greenberg. The effects of workspace awareness support on the usability of real-time distributed groupware. *ACM Trans. Comput.-Hum. Interact.*,6(3):243–281,1999.
- [7] H. Ishii and M. Kobayashi. Clearboard: A seamless medium for shared drawing and conversation with eye contact. In Proceedings of the SIGCHI conference on Human factors in computing systems, pages 525–532, Monterey, CA, USA, May 3-7 1992. ACM Press.
- [8] P. Jermann and P. Dillenbourg. Group mirrors to support interaction regulation in collaborative problem solving. *Computers and Education*,51:279–296,2008.
- [9] D. Kirsh and P. Maglio. On distinguishing epistemic from pragmatic action. *Cognitive Science*,18(4):513 – 549, 1994.
- [10] T. Manninen. Interaction forms and communicative actions in multiplayer games. *Game Studies*,3, 12:[online journal], 2003.
- [11] S. Nakagawa, W. Zhang, and M. Takahashi. Text-independent speaker recognition by combining speaker-specific gmm with speaker adapted syllable-based hmm. In Proc. the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), pages 81–84, 2004.
- [12] P. Qvarfordt, D. Beymer, and S. Zhai. Realtourist -a study of augmenting human-human and human-computer dialogue with eye-gaze overlay. In INTERACT 2005, pages 767–780, 2005.
- [13] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn. Speaker verification using adapted gaussian mixture models. *Digital signal processing*,10:19–41,2000.
- [14] M. Sangin. *Peer Knowledge Modeling in Computer Supported Collaborative Learning*. PhD thesis, Ecole Polytechnique Fédérale de Lausanne, 2009.
- [15] B. Schölkopf, C. J. Burges, and A. J. Smola. *Advances in Kernel Methods -Support Vector Learning*. MIT Press, Cambridge, 1999.
- [16] J. Underwood. Cognitive Processes in Eye Guidance, chapter Novice and expert performance with a dynamic control task: Scanpaths during a computer game, pages 303–324. Oxford University Press, 2005.
- [17] V. Vapnik. *Statistical Learning Theory*. John Wiley and Sons, New Year, 1998.
- [18] R. Vertegaal. The gaze groupware system: mediating joint attention in multiparty communication and collaboration. In CHI '99: Proceedings of the SIGCHI conference on Human factors in computing systems, pages 294–301, New York, NY, USA, 1999. ACM.
- [19] L. S. Vygotsky. *Mind in Society: Development of Higher Psychological Processes*. Harvard University Press, 14th edition, March 1978.
- [20] S. Zhai. What's in the eyes for attentive input. *Communications of the ACM*,46(3):34–39,2003.