Article

Eye movements of online Chinese learners

Ursula Stickler and Lijing Shi

Abstract

Although online tutorials are becoming commonplace for language teaching, very few studies to date have provided insights into learners' behaviours in synchronous online interactions from their own perspective.

This study employs eyetracking technology to investigate ten learners' attention during synchronous online language learning in a multimodal environment. The participants were learners of Chinese as a Foreign Language at beginner's or lower-intermediate level. While learners took part in two different online activities, one focusing on reading, the other on interaction with others, their gaze focus was tracked, and in subsequent stimulated recall interviews the learners reflected on their engagement with the screen and their intentions while reading or speaking online.

Our findings show that during reading tasks, when Pinyin transcriptions as well as Chinese characters were presented, all beginner and lower intermediate participants focused to some degree on the Pinyin. In the interactive task learners' gaze was drawn to elements of the screen that were not immediately necessary for technical or linguistic reasons but that could be interpreted as containing social presence information, e.g. names listed and emoticons employed by other users.

KEYWORDS: CHINESE AS A FOREIGN LANGUAGE; COMPUTER ASSISTED LANGUAGE LEARNING; EYETRACKING; STIMULATED RECALL; SYNCHRONOUS COMPUTER MEDIATED COMMUNICATION (SCMC)

Affiliation

The Open University, UK.
email: ursula.stickler@open.ac.uk (corresponding author)

equinoxonline

Introduction

Despite the fact that online language learning is becoming 'normalised' (Bax, 2003, 2011) research has not yet produced a clear picture of learners' attention focus during synchronous online language learning sessions. As suggested by various researchers (Chun, 2013; Fischer, 2007), finding out what learners actually do when they learn online is an essential step in advancing our knowledge about CALL (Computer Assisted Language Learning) or CMC (Computer Mediated Communication) for online language learning.

One suggested method that has gained prominence over the past years is eyetracking (O'Rourke, 2012; Smith, 2010), a research technique which enables the detailed study of a person's gaze movement. By recording reflections from the user's pupils, the exact position and duration of fixation at a given point is captured. Eyetracking can be a valuable method for revealing learners' attention focus during online activities. Our study employs eyetracking to investigate the learning of Mandarin Chinese in an online multimodal environment.

Online language tutorials at beginner's and lower intermediate level, particularly for a non-alphabetic language like Chinese, typically combine different activities including speaking, listening, reading, pronunciation practice, vocabulary learning, and silently practising along with peer learners (Stickler and Shi, 2013). Specific online tasks can also include dragging and dropping items on a screen area, using emoticons or checking the meaning of unfamiliar vocabulary items in an online dictionary. In a well-designed online tutorial, interaction with materials and with other speakers is integrated to maximize scaffolding and support, crucial for a successful learning experience.

In other words, online language learning tutorials are multimodal (Hampel and Stickler, 2012; Jewitt *et al.*, 2001). In this article we are referring to the modes most prominently apparent in a synchronous audio-graphic environment, i.e. speaking, written teaching text, synchronous writing (textchat), images, and potentially a video broadcast. Multimodality plays a particular role in learning Chinese as a Foreign Language (CFL) because of the use of characters. Chinese characters are logograms: each of them is like a picture or a symbol, and is pronounced as one syllable. The fact that the form of characters bears little relation to the pronunciation has implications for beginner learners (Lee and Kalyuga, 2011) and makes reading Chinese difficult at lower proficiency levels (Wang, 2014). Therefore Chinese textbooks for both first and second language learners of Chinese use a phonetic transcription system called Pinyin. Pinyin uses the Roman alphabet and spelling to approximate the pronunciation of Chinese characters. For example, the character 海 is written as hǎi in Pinyin and pronounced [haɪ]; meaning 'sea'. At more advanced stages, Pinyin is gradually taken away.

Little research has been done in the online learning of Chinese, particularly the processes of online reading and of interactive speaking. Understanding what online learners find difficult to cope with and the underlying reasons is meaningful as it would help us in supporting our students. To fill this gap we followed ten learners of Chinese at an English speaking university during online tutorials, using eyetracking and stimulated recall to investigate their attention during online reading and interactive tasks.

This article will focus on two research questions:

1. During online Chinese tutorials, what is learners' attention focus in reading tasks? And why?
2. During online Chinese tutorials, what is learners' attention focus in interactive tasks? And why?

We will first locate the study within the current trends of online language teaching and learning, and the use of eyetracking in language education and SCMC (Synchronous Computer Mediated Communication) research.

Literature Review

Developments in online language teaching

For the past decade, research in online language teaching and learning has moved from textbased and often asynchronous interaction towards multimodal and synchronous interaction (Blake, 2011; Liu *et al.*, 2003; Stockwell, 2007). It is a trend that interaction in a second language (L2) for the purpose of learning has become more commonplace and distributed (Godwin-Jones, 2012). With developments in software applications and improvement in connectivity, multimodal online environments are now easier to use for online language teaching (Wang *et al.*, 2010).

In line with technological advances, our understanding of the online language learning process as well as online pedagogy have developed (Sun, 2011). Hampel and de los Arcos (2013), for example, described the technological development at the Open University in the UK, placing it in the context of changing pedagogical and theoretical frameworks. Teachers and researchers have become aware of the centrality of learners' contributions (Blake, 2011; Lai and Morrison, 2013), of their active role in creating their own 'learner-context interface', as White (2009) calls it. This means that the intention of the teacher in designing, creating and conducting a task cannot necessarily predict the learning behaviour of a student (Montoro Sanjosé, 2012). Hence, learners' actual behaviour, such as their attention focus, needs to be investigated.

Online teaching of Chinese as a Foreign Language

Since Mandarin Chinese has been identified in the UK as a language important for strategic or political reasons (Dearing and King, 2007; Tinsley and Board, 2014), more people are learning Chinese. Increasingly it is delivered online, for example using internet communication systems or internet telephony such as Skype or WeChat conversations, or via learning management systems such as Blackboard or Moodle. Once the technical hurdles (e.g. inputting Chinese characters) were overcome, computer-assisted teaching of CFL advanced rapidly (Robin, 2013). Indeed, in parallel with other languages, the communication modes in online Chinese teaching have expanded from mainly text to a combination of text, audio, and emoticons. For example, Wang and her colleagues have been pioneering the use of SCMC tools (e.g. video conferencing) for CFL teaching since 2004 (Wang, 2004; Wang and Chen, 2007, 2009, 2010). Based on their studies on theories of SLA (Second Language Acquisition) and distance learning, they argue that synchronous interaction is a crucial component for online language learning. As their studies were based on retrospective user reflections, an investigation of the exact behaviours such as attention focus of online learners in real time is still outstanding.

A previous study (Stickler and Shi, 2013) sought to identify how Chinese teachers' intentions match with students' perceptions or expectations during online multimodal tutorials. Employing multimodal analysis of synchronous online speaking interactions and stimulated recall, this study revealed mismatches between teacher's intentions and students' perceptions during online tutorials. Such mismatches can lead to communication failure, anxiety and even the total abandoning of language learning online. To understand why learners are disheartened and how they could be encouraged to continue online language learning, it is useful to find out exactly what students' attention is focused on during online tutorials.

The shift in our own research focus from teaching to the learners' perspective echoes the debates about future research directions in the field of CALL more generally. The call for identifying 'what learners do' originated in Fischer's paper 'How do we know what students are actually doing? Monitoring students' behavior in CALL" (Fischer, 2007), and has been taken up recently in a CALICO *Festschrift* devoted to his work (Hubbard *et al.*, 2013).

Among the various methods suggested for the investigation of students' behaviour during online learning tasks, Chun (2013) identified eyetracking as one promising method. Eyetracking can provide 'a dynamic trace of where a person's *attention* is being directed in relation to a visual display' (Poole and Ball, 2006: 213).

Eyetracking as a research tool for SCMC

A small number of researchers have been pioneering the use of eyetracking in SCMC for language learning by adapting the techniques of HCI (human-computer interaction) studies to explore individual SCMC processes (O'Rourke, 2008, 2012; Smith, 2010, 2012). In O'Rourke's study Irish university students of French or German interacted with native speakers via a text-based SCMC environment. O'Rourke (2012) employed three ways of analysing eyetracking data collected during these online tandem sessions: first, he examined reading patterns in native speaker and non-native speaker interaction; he held that gaze replay could yield insight into individual linguistic-cognitive strategies in L2 SCMC. Second, eyetracking data of two students were combined with event data (keypresses, mouse-clicks) to find out patterns of self-monitoring of their writing. Third, O'Rourke triangulated one student's eyetracking data with a log extract and screen video, attempting to show that 'SCMC discourse text is not just an ordinary conversation jumbled up; it is a form of conversation that is multilinear even with just two participants, and its coherence must be actively inferred and tracked by the participants, and by the analyst' (O'Rourke, 2012: 31).

While O'Rourke's study is based on quasi-naturalistic sessions, Smith (2012) carried out his eyetracking study in an experimental research set-up. The focus of Smith's investigation was noticing of recasts by 18 learners of English at university level. Learners engaged in a short text chat with a native speaker who provided intensive and explicit corrective recasts. Smith compared learners' gaze focus during the experiment and a stimulated recall session. Noticing events were compiled from these two techniques. A pre-test, immediate and delayed post-tests of English proficiency were also used. Both eyetracking and stimulated recall data suggested that learners were able to notice semantic and syntactic targets more easily than morphological targets. Smith argued that 'the use of eye gaze data seems to be potentially valuable in helping to determine which features of the input are likely to be noticed and which are not since we can see precisely what learners view and arguably attend to' (Smith, 2012: 72).

Both O'Rourke and Smith mention the importance of investigating eyetracking as a promising method in SCMC research, however, they both limited their own studies to textbased SCMC and one-to-one interactions. Their studies inspired us to adapt eyetracking technique for our investigation of synchronous language tutorials where multiple levels of interactions take place (teacher—learners, learners—learners, learner—computer) through multiple modes (spoken, written, graphic) and modalities (e.g. text, audio, emoticons).

Project Description and Research Methodology

The sections below introduce our participants, the online conferencing system, learning activities, methodology, and data collection instruments.

Participants

Our participants were ten adult learners of Chinese at early stages, beginners to lower intermediate. Nine were Western learners and one was a heritage speaker of Cantonese. Eight participants had previously completed Dì yī bù (第一步), a ten-month distance beginners' course at the Open University (OU) in the UK (Kan and McCormick, 2012; Stickler and Shi, 2013) that leads to an equivalent of A2 level on the Common European Framework of Reference for Languages (CEFR). One learner had taken part in equivalent courses at Adult Continuing Education institutions, covering a similar length of study and level of achievement. The last learner was less advanced, having only covered an equivalent of three months of study, placing her approximately at A1 of the CEFR level. All learners were computer literate adults in full-time or part-time employment, and had taken Chinese as an optional course.

For this study, the learners took part in one reading and one interactive online activity, both of which were recorded in the OU eyetracking lab. The project lasted 13 weeks between July and October 2012. All participants filled in the pre-study questionnaire before the start of the first activity[1] (see Appendix A for selected responses).

To anonymize our participants, they were given Chinese names, which is a common and well-received practice in CFL classrooms. The project followed BERA (British Educational Research Association) ethical guidelines and was given full institutional approval by the Open University's Ethics committee.

Elluminate: Online conferencing system

The online conferencing system used in our project was *Elluminate* (see Figure 1). This software was also the platform used for tutorials at the OU and was therefore familiar to the participants. Depending on the intensity and manner of their prior study, they were more or less skilled in handling the different features of *Elluminate* for interaction with the computer (activity 1, reading) and for interaction with other learners and the tutor (activity 2, interactive).

In addition to speaking by activating the microphone button, *Elluminate* allows for written exchanges using textchat to manipulate whiteboard content, e.g. dragging and dropping elements (see Heiser *et al.*, 2013). An important element of the software is the participants' window, showing a list of participants' names in the online session, with small icons indicating the activities that they are engaged with, e.g. writing in the textchat, manipulating elements

on the whiteboard. Participants can also indicate their emotional state using a set of emoticons, and vote with a 'yes/no' button. By clicking on a raised-hand icon, learners can signal their intention to speak.
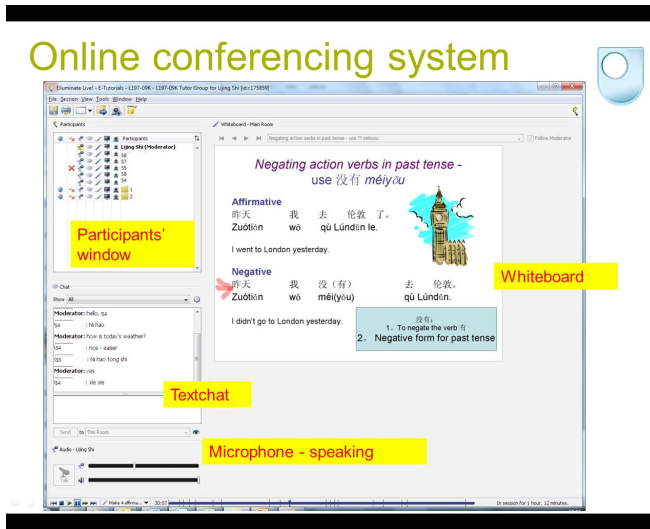

Figure 1. *Elluminate* screenprint with main areas labelled.

Reading and Interactive activities

In the first activity, reading content was presented in the whiteboard area and participants worked their way independently through a series of whiteboard screens presenting instructions and tasks. After a set of brief warming up tasks, the main reading task was presented as a short text in characters with Pinyin transcription below, followed by three comprehension questions in English (for details of the tasks see Figure 2; for the full text see Appendix B).

The second activity, which was interactive, centred around the theme of 'transportation' and was led by an experienced online tutor, one of the authors of this paper. It involved synchronous online spoken interaction with learners who took part remotely while the participant was being recorded in the eyetracking laboratory. Each participant had to take part in a separate tutorial as only one eyetracker was available to us in the labs. The instructions were given verbally by the online tutor and the tasks involved the manipulation of whiteboard elements, as well as speaking interactively with the tutor and with other online participants. The four interactive tasks, lasting approximately 15 minutes, were designed to help learners recall vocabulary items, practise their pronunciation, and employ revised words and structures in short simple dialogues (see Appendix C for details).

Methodology

Eyetracking for the investigation of synchronous online language learning is relatively new. However, this method has been used for more than 100 years in reading research (Just and Carpenter, 1976; Rayner, 2009), and it has been gaining popularity in HCI and usability research. To answer our research question, we have studied how eyetracking has been used in those areas. The following summarizes the concepts which shaped our research design.

Eyetracking can be defined as a technique 'whereby an individual's eye movements are measured so that the researcher knows where a person is looking at any given time and the sequence in which their eyes are shifting from one location to another' (Poole and Ball, 2006: 211). In reading research, the two most widely used measures of eye movements that have been developed are eye fixations and saccades (Duchowski, 2003). Fixations are 'those moments when the eyes are relatively stationary and reflect when information is being encoded' (Smith, 2012: 55), while saccades refer to 'the eye's rapid movements from one fixation to the next' (Nielsen and Pernice, 2010: 7). Reading researchers have come to a general agreement that average fixation duration lasts approximately 200–250 milliseconds, and readers normally make about 3 to 4 saccadic movements per second.

In the context of interface design and usability evaluation the following three metrics are mainly used: fixation-derived metrics (e.g. fixation duration, number of fixations overall), saccade-derived metrics (e.g. number, amplitude), and scanpath-derived metrics (Poole and Ball, 2006).

Usability researchers have discovered that eyetracking data could be influenced by participants' physical features such as the size of their pupils, the kinds of spectacles they wear, and the design of the task. To increase the validity and reliability of eyetracking data, Nielsen and Pernice (2010) suggested combining eyetracking with other research methods such as stimulated recall, questionnaires, interviews, and observation. Hence, we combined eyetracking with questionnaires and stimulated recall in this study.

Data collection

In brief, the data collection instruments used were:

- the pre-study questionnaire
- the eyetracking recordings of two separate activities (reading and interactive)
- stimulated recall interviews following each eyetracking activity.

Questionnaire

Before taking part in the eyetracking activities, all 10 participants filled in a questionnaire detailing their personal and learning background, their

self-evaluated level of Chinese and of general ICT skills, the latter partly based on Spitzberg's (2006) CMC competence questionnaire.

Eyetracking

The eyetracking sessions were carried out in a laboratory room at the OU that is equipped with ceiling mounted cameras, to record procedures and follow-up interviews. The equipment used was a table mounted Tobii 60 eyetracker. In this set-up, eye movements were recorded by reflecting an infrared light beam off the pupil of the eye. A video of the screen was recorded with gaze focus points overlaid visually. Eye focus was recorded numerically at a rate of 16 ms (or 62.5 hertz).

Individual participants engaged with the tasks while seated at the computer with eyetracking equipment. Before starting the actual task, the equipment had to be calibrated for every individual user until satisfactory accuracy was achieved.

Eyetracking data analysis methods

Tobii Studio 3.2.1 software was used to capture and analyse eyetracking data; it created instant visualizations. One type of visualization, gazeplot, shows focus points as numbered dots and movement in-between the focus points as lines (see Figure 2 for an example). The gazeplot can also be presented in dynamic form as a video showing movement of eye focus from one area to the next. This can be replayed to the participants, e.g. for stimulated recall interviews.
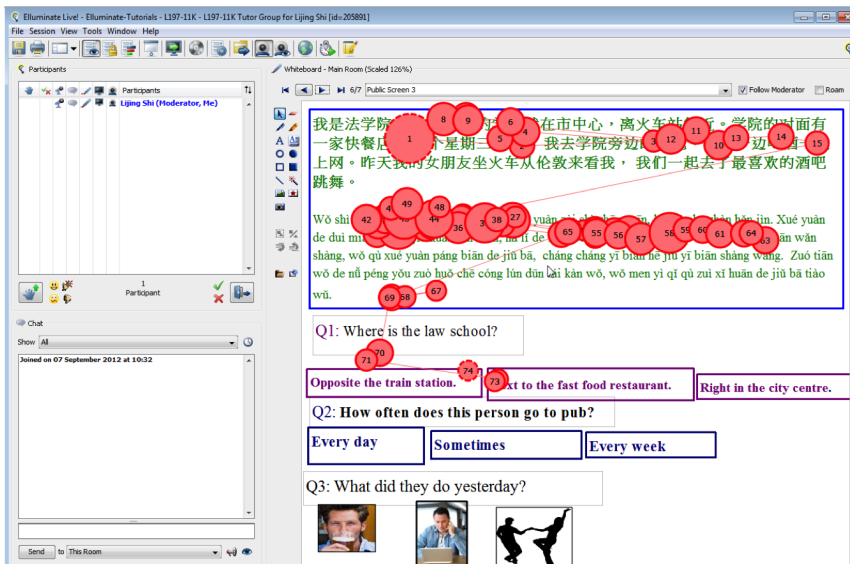

Figure 2. Gazeplot image of activity 1, reading task

A different type of visualization is called heatmap, a static image of an accumulation of focus points where the longer the focus remains on a certain area of the screen, the 'hotter' the colour becomes, changing from light green to deep red (see Figure 3 for an example). A reversed heat map, leaving the focus areas visible and the areas that attracted little attention blackened out, is called gaze opacity map (see Figure 7 for an example).

To analyse the data recorded during eyetracking, two measurements were taken: fixation duration and fixation count. The length of fixation in total (total fixation duration) shows where the main attention is focused during online work, whereas the number of fixations at a certain point (fixation count) can identify areas of increased difficulty. To study specific areas, Areas of Interest (AoIs) can be defined manually by an outline on the screen; this allows detailed analysis and comparison. For example, a screen of reading material prepared for a Chinese tutorial can be divided into areas of Chinese characters (named AoI 1) and areas with Pinyin transcription (AoI 2).

Stimulated recall

After each activity, the researchers in this study played back the gazeplot video of the eyetracking to the participants and asked them to recall what they were doing following a stimulated recall method (Gass and Mackey, 2000). Using the stimulus of the gazeplot video allowed participants to reflect and to explore their own learning experience, be it the interaction with the computer or the more complex interaction with other participants online. The participants used the opportunity to elaborate on the reasons for their gaze focus, offering explanations related to learning strategies and sometimes speculating on possible alternative explanations.

Data analysis and findings

Findings of our study are presented in two sections according to our two research questions.
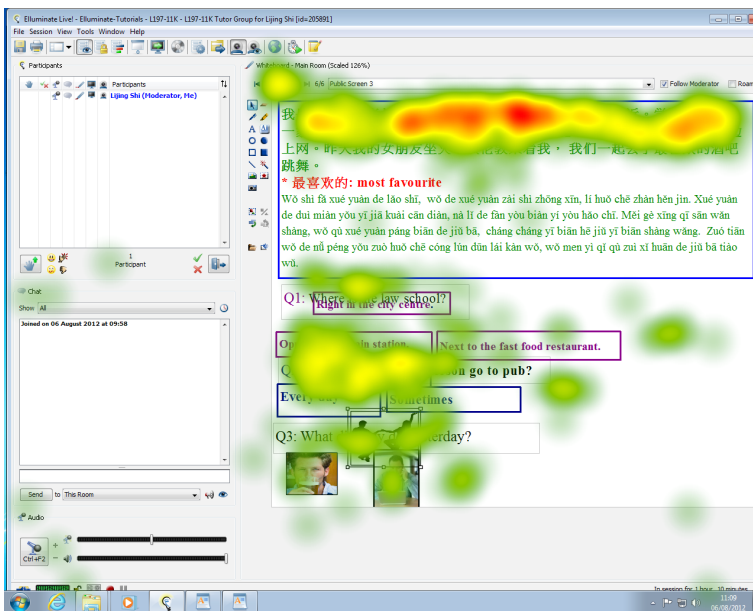
Activity 1: Reading

To answer our first research question, we studied participants' eye movements during an online reading task and their reflections revealed in stimulated recall.

First we collected parameters derived from the reading activity. Table 1 shows how long participants took to answer three reading comprehension questions, how many of the questions they answered correctly, and evaluations of their Chinese language skills. The self-rating is taken from students' pre-questionnaire. However, this proved inconclusive as learners tended to underestimate their levels. For this reason the teacher's rating was added.

Table 1. Participants' time on task and results of reading task

Pseudonym	Mai Kemu	Ai Mi	Cha Li	Deng Kan	Ma Li	Jie Ning	Li Sha	Wu Xi	Lin Da	Su Shan
Overall time to answer	1'27"	1'28"	2'02"	2'50"	3'01"	3'05"	4'00"	4'13"	4'46"	5'58"
Reading task accuracy	66%	100%	100%	33%	66%	66%	100%	100%	100%	66%
Student overall self-rating	Average	Average	Average	Very poor	Good	Poor	Poor	Poor	Average	Average
Teacher-rated overall Chinese skills	Excellent	Excellent	Very good	Average	Good	Poor	Very good	Good	Very good	Good

Heatmaps were generated from Tobii to ascertain the main areas of the gaze focus for each learner. Then AoIs were created manually to distinguish between Chinese text and the Pinyin transcription. The heatmaps of all the participants share similar features: they focus on Pinyin or characters, respectively; their secondary focus is on the comprehension questions; and they ignore the left-hand side of the screen, keeping their gaze predominantly on the whiteboard area. Variations in length of fixation can be broadly categorized according to the language level of individual learners (excellent, very good, good and poor). Figures 3a–3d show heatmaps in these four categories. Deng Kan's reading skills were evaluated as 'average', but he did not participate in the interactive task, therefore, his heatmap is not included.

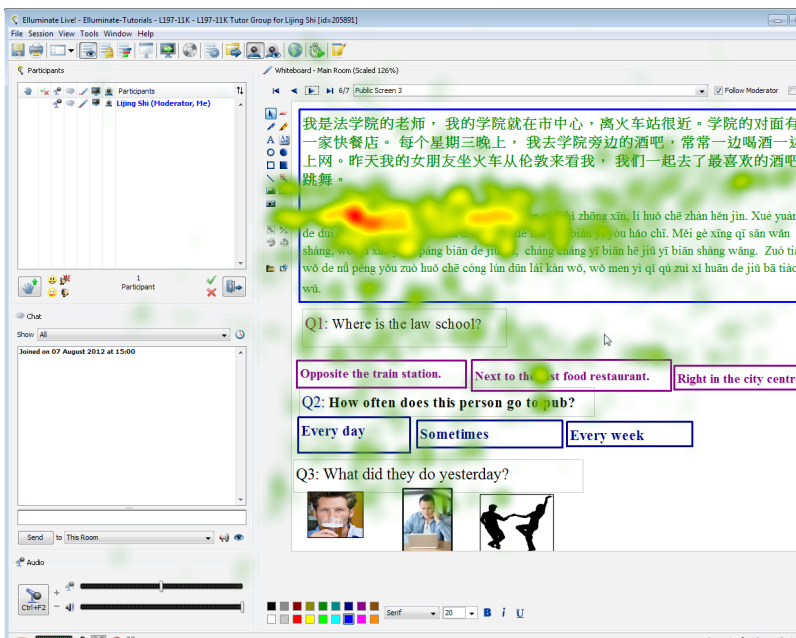**Figure 3a.** Activity 1 (reading task), participant 爱米 (Ai Mi) – excellent

Figure 3b. Activity 1 (reading task), participant 林妲 **(Lin Da)** – very good
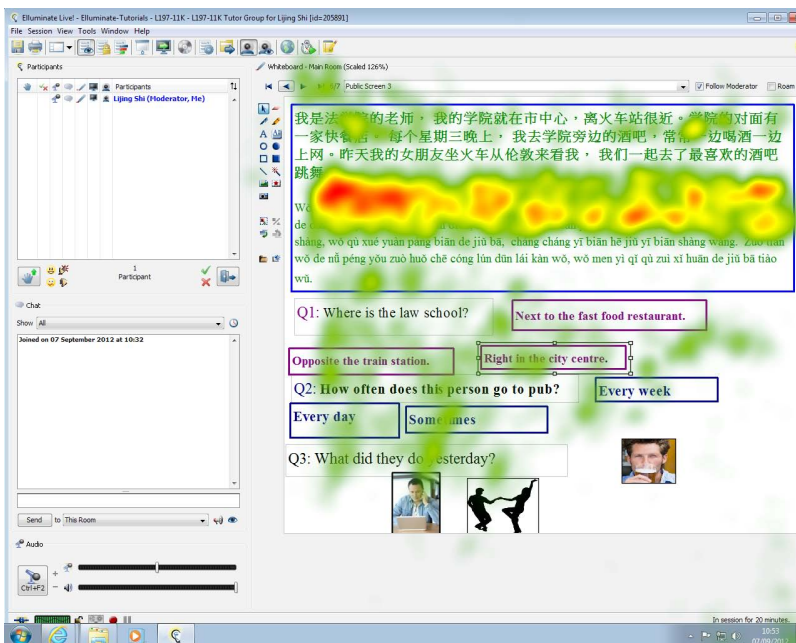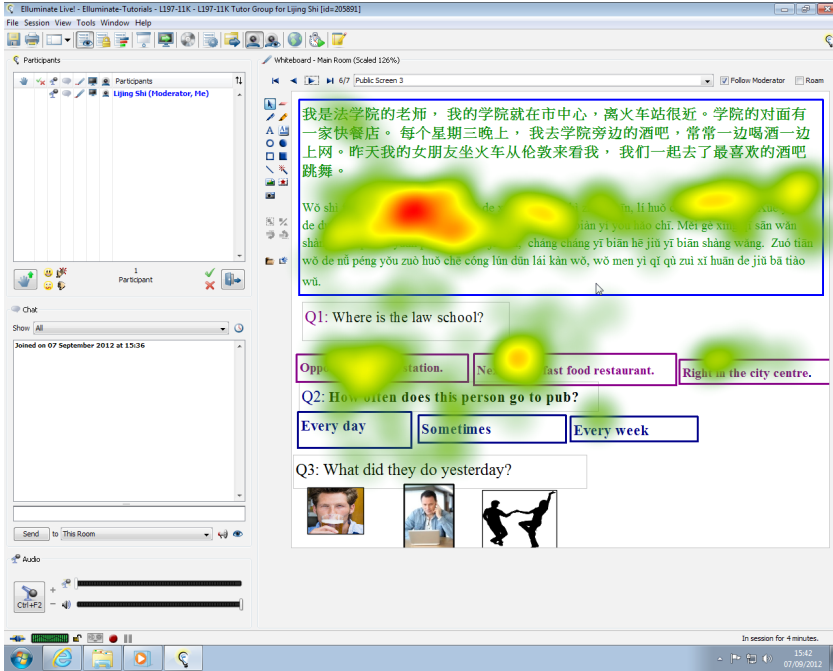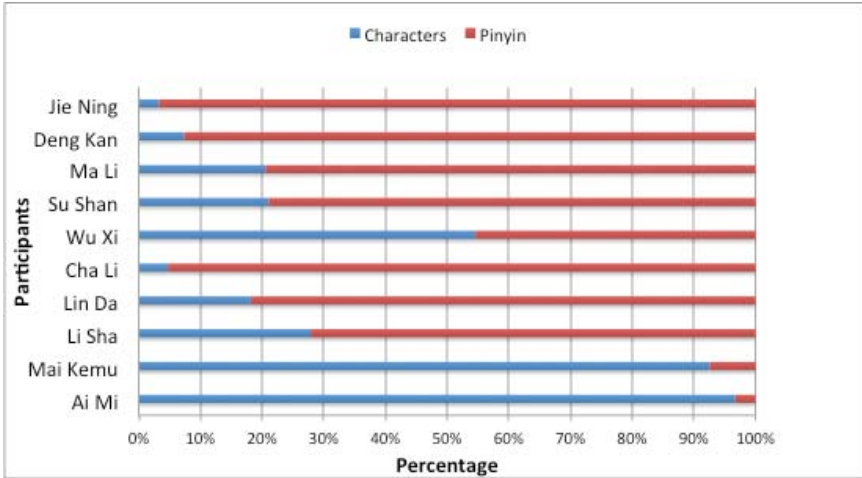

Figure 3c. Activity 1 (reading task), participant 苏山 (Su Shan) – good

**Figure 3d.** Activity 1 (reading task), participant 杰妮 (Jie Ning) – poor

The heatmaps in Figures 3a to 3d illustrate the increasing use of Pinyin with diminishing Chinese reading skills. Whereas the gaze focus (the hot spots) in the heatmap of the learner at excellent level is almost entirely on the character part of the reading text, for the learner at very good level the gaze focus is already distributed between characters and Pinyin transcription, placed directly underneath. The learner at good level still shows some minor attention (green spots) on the characters, whereas the gaze focus of the learner at poor level seems entirely concentrated on the Pinyin part of the reading text. There is no attention focus on the social areas and only very occasional glances at the technical areas.

Further analysis of fixation duration in the selected AoIs confirms these differences. Everyone used Pinyin, but to a different degree, with fixation duration on Pinyin ranging from 3% (Ai Mi, a learner at excellent level) to 97% (Jie Ning, a learner at poor level) of the time spent on the task (see Figure 4). To Ai Mi, a heritage speaker of Cantonese who is familiar with traditional Chinese characters, characters convey more meaning than Pinyin.

Figure 4. Fixation duration of activity 1 (reading task), Characters vs. Pinyin

To visualize the link between participants' fixation duration (on characters and Pinyin) and their language ability Figure 5 is derived from Table 1 and Figure 4.
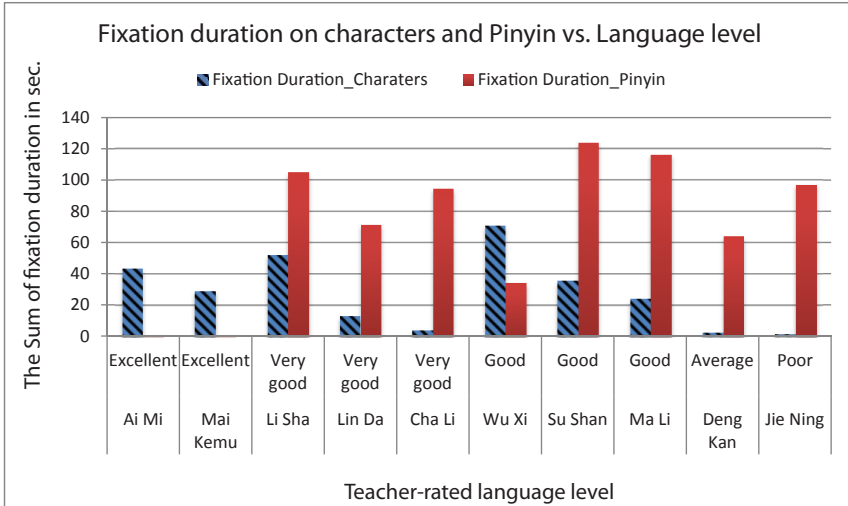
Figure 5. Fixation duration on characters vs. Pinyin and Language level

The eyetracking software (Tobii) has, thus far, provided us with graphical and numerical data in terms of how students pay attention to Chinese characters and Pinyin in online reading. The stimulated recall interviews

supplemented the eyetracking data by providing participants' reasons which can be classified as: (a) simply because Pinyin was available (convenience); (b) using Pinyin for confirmation; and (c) relying on Pinyin for comprehension.

A good example for 'convenience' usage is Lin Da, a very competent learner at beginner to intermediate stage. Lin Da mentioned in her stimulated recall interview that she made a conscious effort to focus on characters but resorted to Pinyin as a matter of convenience and speed of interpretation.

> *Lin Da: If the Pinyin are not there I will read the characters. And I try to make myself look at the characters because I remember, you know, when I first started I spent most of the time looking at Pinyin because I was in a hurry and I can't get out of the habit. I think it is very hard but I think if the characters are the only things there then I will look at them.*
>
> Participant Lin Da: 20'02" – 20'20"

As an analysis of the fixation duration of the two distinct AoIs (characters and Pinyin) shows this strategy of convenient Pinyin use resulted in Lin Da using characters for less than 20% of the time compared to Pinyin, which she used over 80% (see Figure 4).

Other participants differed in their approach. A good example of working with Pinyin and characters for confirmation is Su Shan, who strategically decided to answer the most questions correctly in the minimum amount of time. She used different clues available, and worked from questions back to the text rather than starting by reading the text.

> *US: …, you just want to answer the questions, I suppose.*
>
> *Su Shan: and I think that's it, I am getting it done in five minutes.*
>
> *US: Okay.*
>
> *Su Shan: I think it goes all the way. And I don't focus on any question, I just read them all and then when I find something that could be part or could be answering one of the other questions and then I follow it up a bit and then if I can't come to a conclusion I go to the next question.*
>
> Participant Su Shan: 27'29" – 28'15"

She used characters for 21% of the time for strategic confirmation. As she explained: '*I am going back to the characters to see if I can find more vital clues within the characters*' (Participant Su Shan 26'18" – 26'27").

In contrast, Jie Ning exemplifies the third reason, i.e. relying on Pinyin for comprehension. She declared that she only knew a few characters and did not even attempt to use them to support her comprehension task. Her looking at characters accounted for less than 5%.

Activity 2: Interactive

To answer the second research question, focusing on interaction with others online, we examined eyetracking and stimulated recall data from a second activity conducted with eight of the participants who originally took part in activity 1. One recording with a calibration below 55% accuracy was discarded and not used for analysis.

All the heatmaps for this interactive activity showed similar features, therefore they were combined (see Figure 6), illustrating that participants' attention was concentrated on the Pinyin sections of the whiteboard, the names of fellow participants, the area indicating their state, and the microphone button.
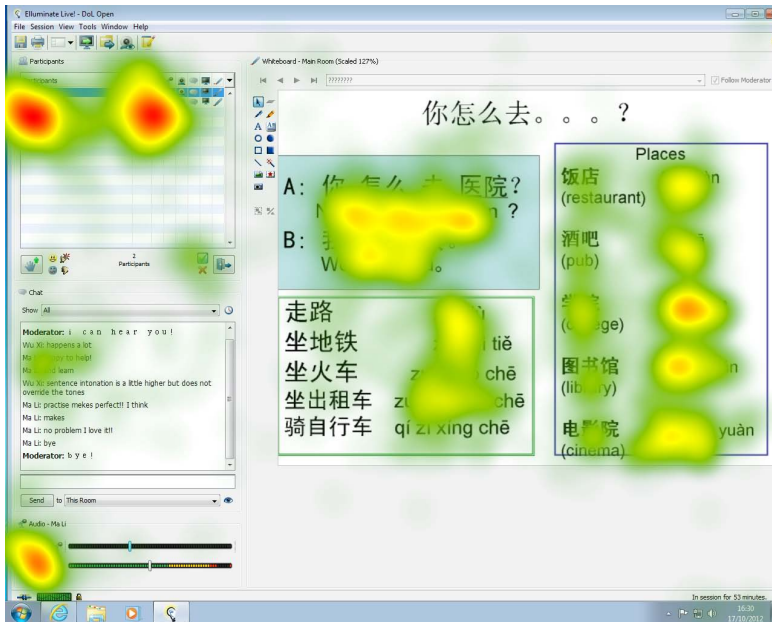

Figure 6. Activity 2 (interactive), Task 4, combined heatmap of all participants

On the combined heatmap we manually marked concentrations of attention as AoIs, and subsequently categorized these AoIs into three types. The following three types were identified:

- 'Content' is where the learning material is displayed on the whiteboard.
- 'Social' shows participants' presence and their interaction mode (e.g. typing, speaking) represented in icons.
- 'Technical' is where participants can activate microphone or textchat to communicate and select small icons (emoticons) or Yes/No voting buttons.

For numerical analysis, fixation duration on same type AoIs was clustered and added together. Content accounts for approximately 70% of the overall fixation duration, social for approximately 20%, and technical for approximately 10%. These types of AoIs with their associated fixation duration expressed as percentage are illustrated in the gaze opacity map (Figure 7), which is the reverse image of the combined heatmap.
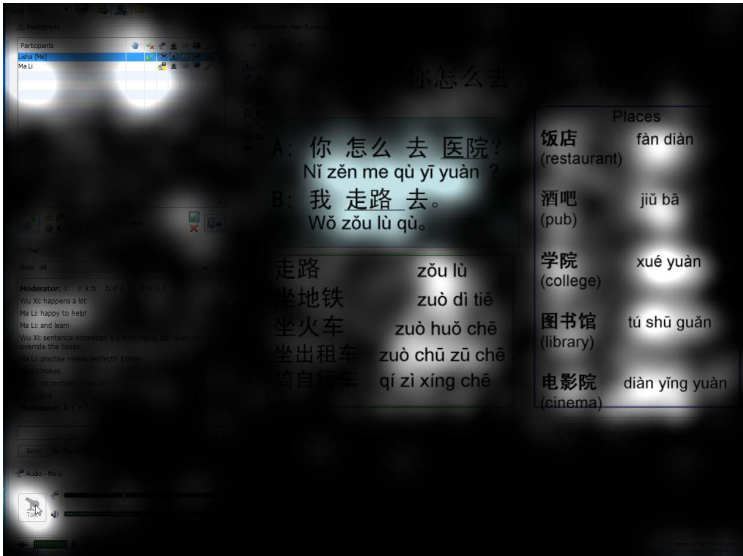
Figure 7. Three types of AoIs and relative fixation duration

As in activity 1 (reading), stimulated recall data were used to explain actions of learners during activity 2 (interactive). The second stimulated recall interview included reflections on speaking interactions with others and interpretation of both reading and interactive activities. For example, participants talked about why their attention was drawn towards the content, social or technical AoIs. Some examples of the recalls are as follows:

US: *And checking on the left hand side there.*

Lin Da: *Uh, Yah. I think I tend to look at that when somebody is speaking, and check out who they are.*

(Lin Da 13'34 - 13' 43)

Su Shan expressed a very similar opinion.

Su Shan: *So, here I'm checking who is there. Interactive task, so waiting for <xx> interactive. Checking the names again, checking what I need to do, what tools I've got.*

(Su Shan 54'09 – 54'18).

And

> Su Shan: So I'm checking who is going to say the next one. Focusing on that … translation.
>
> <div align="right">(Su Shan 57'46 – 57'53)</div>

Ma Li described in detail the social presence of others she noticed during the online session.

> Ma Li: *Yes. I think I kept looking when any activity, /. You're looking here and then, out of the corner of your eye, you could see there's somebody's joined or they put a happy face there or something. And there is quite a lot going on here, more than in other sessions I have been to, uhm, because here, you were typing some stuff. People were, at the beginning I think there is more activity in the text area.*

> US: Mhm.

> Ma Li: *When other people are doing things. I keep looking at the names. Yeah, wondering if I knew Marina, checking, checking the spelling because I am Ma Li there, checking who is speaking over the microphone.*
>
> <div align="right">(Ma Li 1:07'38)</div>

Ma Li also explained why she looked at the technical area.

> Ma Li: *Yeah. And, of course, every time you to speak, you have to look to get your cursor over there, and this is repeating the /…*
>
> <div align="right">(Ma Li 1:08'10)</div>

Analysing eyetracking data helped use to answer the first part of our research questions, i.e. what our learners focus on, and stimulated recall supplied the reasons for their attention focus.

Discussion

We set out to study real-time learners' attention focus during online Chinese tutorials. In our first activity, eyetracking data indicated the ratio of participants' attention on Pinyin ranged from 3% to 97% (Figure 4). Using teacher-rated language levels, we tried to establish a link between learners' language skills and the length of time their attention focused on characters and Pinyin, respectively.

The process for CFL learners to develop their reading skills (Ren and Yang, 2010) from Pinyin to characters varied. From Figure 5, one can see two extremes: learners at excellent level rely predominantly on characters for comprehension tasks, whereas learners at poor and average levels rely on Pinyin. For learners in-between, their fixation duration on characters and Pinyin combined was higher than either of the two extremes so they spent longer

time overall. Additionally, their fixation duration on Pinyin dominated, with one exception (i.e. Wu Xi). This indicates that they actively used both characters and Pinyin to complete the reading task.

Following Nielsen and Pernice's (2010) suggestion, we supplemented eyetracking with stimulated recall interviews in the context of online Chinese reading. From the stimulated recall data, we could deduce that some participants needed to rely on Pinyin for comprehension because of their limited character recognition (e.g. Jie Ning), or their intention to finish the task efficiently by targeting Pinyin primarily, with characters only for confirmation (e.g. Su Shan). In some cases, the glance was drawn to Pinyin simply because it was displayed on the reading task screen and it was familiar to the student (e.g. Lin Da).

There has been an on-going debate on the role of Pinyin in CFL. Some teachers think that Pinyin should be withdrawn as soon as possible, but others regard Pinyin as the 'lifeline' to CFL learners' spoken vocabulary acquisition and reading development (Everson, 2008). Pinyin serves as a crucial scaffold for comprehension and speaking especially at the early stage of CFL learning, while it later becomes a crutch delaying the full development of reading skills in characters (Ye, 2011, 2013; Koda, 1992). From our stimulated recall interviews we identified some factors influencing learners' attention (e.g. comprehension, confirmation, consolidation) on Pinyin and character reading. These factors are summarized in Figure 8 which was partly inspired by Ye's depiction of the connections between Chinese characters, Pinyin, and meaning.
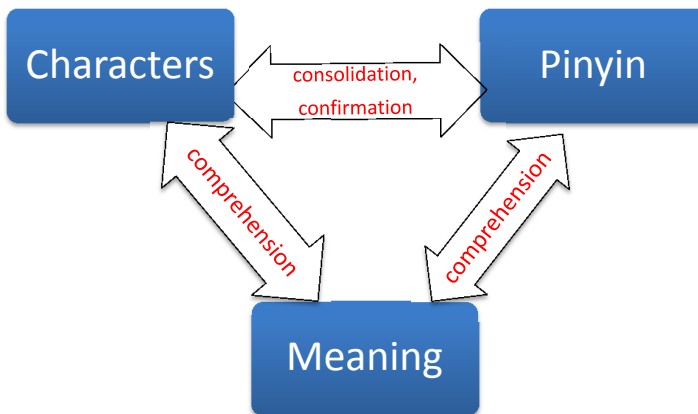
**Figure 8.** How learners make meaning in Chinese

The stimulated recall interviews not only supported the eyetracking data, but also extended our knowledge about the reasons for learners' use

of characters and Pinyin in online reading tasks. For example, Pinyin can help those learners who remember the sound of a Chinese word and associate this sound with the meaning. On the other hand, some ideographic elements of Chinese characters can aid memory and comprehension even without an exact knowledge of the sound of the Chinese word.

Based on our eyetracking and stimulated recall data, we discovered that learners used Pinyin for meaning comprehension, as well as for consolidation and confirmation of characters and Pinyin simultaneously. This suggests that CFL teachers could be more flexible in their approach to teaching characters, and not withdraw Pinyin too early.

The eyetracking data of the interactive activity showed participants still spend more than two-thirds of their attention on content, but a significant amount of attention was devoted to social and technical features. Only the use of some technical features (e.g. microphone) was necessitated by the task (speaking), not the social features which are purely informative. In terms of the technical areas, participants devoted 9.98% of their attention in order to communicate with the others.

The gaze opacity map (Figure 7) indicated that for speaking purposes learners' attention was drawn to Pinyin considerably more than to characters. This is not surprising for beginner level learners (Everson, 2008). During online synchronous interaction language learners face a number of challenges to achieve successful communication. At the most basic level, like any other language learner they have to cope with linguistic challenges. In the case of Chinese, this is exacerbated by the difficulty of relating characters to pronunciation (see Figure 9).
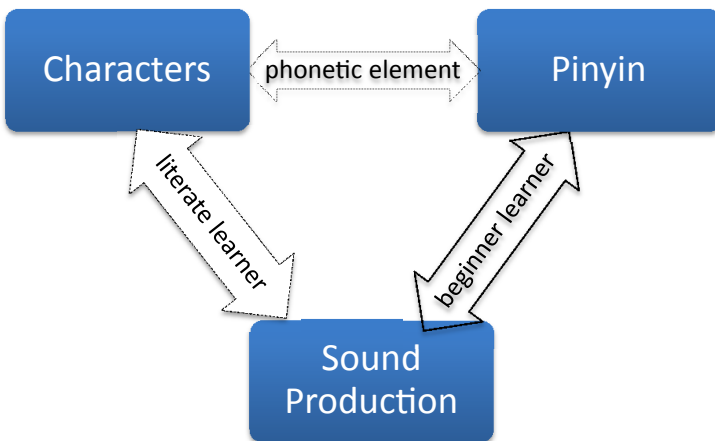
Figure 9. How learners produce sounds in Chinese

The eyetracking data (Figure 6) showed that all the participants engaging in the online interactive task constantly moved their gaze to the social areas, representing one-fifth of the participants' attention. Although the participants did not receive any instructions about using the left-hand side of *Elluminate*, their unguided gaze focus during the interactive tasks showed concentration on names as one of the visual representations of interlocutors in online learning environments. When recalling their interactions, participants mentioned their need to 'see who's speaking', 'check who it was', etc. verbalizing their interpretation of social presence indicators.

In addition to challenges common in face-to-face classrooms, online learners also have to cope with a different approach to social presence: rather than physical presence in a classroom, online learning spaces offer representations of the others that have to be consciously or subconsciously interpreted by the interlocutors. Our findings confirmed that from the learner's point of view, social areas played a significant role, whether they were relevant for the immediate task or not.

Whereas Yamada and Akahori (2007, 2009) claimed that image of the interlocutor had the most effect on learners' perception of social presence, our data show that the simple representation of interlocutor by name in a list was significant. Learners' used whatever means presented (i.e. name list, icons) to interpret and link on to the social presence of others. Learners in our small sample picked up on the clues projected by their fellow students, e.g. the 'happy faces' mentioned by Ma Li, confirming what Satar (2010) has shown: the importance of being able to project as well as understand social presence in synchronous online language learning.

Conclusion

Eyetracking has proven to be a unique tool to scrutinize online language learners' attention during reading in Chinese. Combining it with stimulated recall interviews, we established the reasons for devoting learners' attention to the specific AoIs. Without eyetracking, we would not have been able to establish the contributions of Pinyin for beginner to lower-intermediate CFL learners. Without stimulated recall interviews, our interpretation of learners' reasons would have been speculative at best. Pinyin assists learners' switching between consolidation and confirmation to aid reading comprehension as evidenced in their attention.

Eyetracking is also valuable in revealing learners' attention in an interactive online learning situation. We established the percentage of learners' attention on the content, social and technical AoIs. Within the content areas, heatmaps showed that participants' attention was mainly on Pinyin, confirming the important role of Pinyin for speaking tasks. We found that in

online interactive tutorials a considerable amount of attention was given to the social areas. This demonstrated learners' need for person representations (or 'social presence indicators') during online interactions. Language learning is not just a cognitive activity but also interactive and social. Different online tasks need different support and technical affordances; whereas the online reading task can be done without any attention to the social (or technical) areas of the screen, as we have shown, the interactive task is supported by realizing the presence of others through various means (name lists, emoticons, image, etc.).

Giving a voice to our participants through stimulated recall, and ensuring that they benefit from an enhanced reflectivity on their learning through ongoing discussions and a follow-up questionnaire as part of our research project has changed the research perspective from an outsider view on cognitive processes to an insider-outsider perspective on socio-cultural learning events online. Although eyetracking in SLA research has so far predominantly taken place within interactionist or cognitivist frameworks, combining eyetracking with stimulated recall interviews has proven worthwhile for deepening our understanding of the language learning processes from a socio-cultural perspective[2].

What we set out to do was to find out '[w]hat students are actually doing' (Fischer, 2007) when they work online. Our understanding has advanced to a certain extent by tracking our participants' eye movements and collecting their interpretation of their actions through stimulated recall interviews. Additionally, we have also encouraged learners to reflect on their own learning behaviours, and we have found out in the process that eyetracking data can be a powerful pedagogical tool: watching eyetracking visualizations can enhance learners' awareness of their own learning as well as being useful for teacher training and staff development to advance the teachers' understanding of a learner's point of view.

Limitations of this study and future directions

For research purposes we artificially separated individual reading tasks from interactive speaking tasks to make a comparison of learner behaviour more easily visible. In reality, good online language tutorials usually combine the different elements. Analysing these online language tutorials with their various elements has highlighted the nature of online language learning as a continuous and dynamic intertwining of different modes, tools and tasks. Guided by the materials or the teacher, online learners move between reading, interacting with a screen and with other participants, listening comprehension, and spoken production. As we had only ten participants in our study, the results need to be interpreted with caution.

Eyetracking research opens options for future SCMC research in two directions. On one hand, we need to compare Western learners of Chinese at different levels, and their exact use of Pinyin versus characters. This can be achieved either through a large-scale eyetracking study at a certain point in time, or through a longitudinal study following a group of learners from their very first encounters with Chinese characters to a higher level of reading competence. These studies could confirm threshold levels for the necessity of Pinyin and hence inform pedagogy. Such a method could be extended to other non-Roman languages, such as Arabic.

On the other hand, our research opened a new and innovative avenue for reflective action research. Even this small-scale study has proven the value of combining eyetracking with more qualitative methods in a socio-cultural paradigm of L2 learning. This could lead to engaging the learners more fully in utilising their experience for awareness raising activities, for example, learner strategy training. Action research can also involve online teachers learning about their own attention focus during online tutorials and sharing their expertise with novice online teachers.

Notes

1.   Responses to an additional follow-up questionnaire sent five weeks after the last activity was completed are not included in the data for this article.

2.   The extent of this paper has not allowed sufficient space to provide details of the reasons behind developing the methodology for this study. This will be done in a future publication.

About the Authors

Ursula Stickler is Senior Lecturer in the Department of Languages at the Open University (UK) and convenor of the Open Languages Research Group. Her research interests focus on independent and technology-supported language learning. She is actively involved in several European projects collaborating in the area of autonomous online learning, and teacher training. She has written a number of articles and book chapters on language learning and teaching, and has supervised students on the university's PhD and EdD programmes.

Lijing Shi (Ph.D. in Technology Enhanced Language Learning, M.Ed. in Educational Technology & TESOL), is currently an assistant co-ordinator of Mandarin in the London School of Economics and Political Sciences where she teaches Mandarin Chinese degree and certificate courses. Her main research areas are online language teaching and learning, and intercultural communicative competence.

Acknowledgement

References

Bax, S. (2003). CALL—past, present and future. *System*, 31 (1): 13–28. http://dx.doi.org/10.1016/S0346-251X(02)00071-4

Bax, S. (2011). Normalisation revisited: The effective use of technology in language education. *International Journal of Computer-Assisted Language Learning and Teaching*, 1 (2): 1–15. http://dx.doi.org/10.4018/ijcallt.2011040101

Blake, R. J. (2011). Current trends in online language learning. *Annual Review of Applied Linguistics,* 31: 19-35. http://dx.doi.org/10.1017/S026719051100002X

Chun, D. M. (2013). Contributions of Tracking User Behavior to SLA Research. *CALICO Journal,* 1–2: 256–262.

Dearing, R. and King, L. (2007). *Languages Review*. London: DfES.

Duchowski, A. T. (2003). *Eye Tracking Methodology. Theory and practice*. London: Springer. http://dx.doi.org/10.1007/978-1-4471-3750-4

Everson, M. E. (2008). Issues in Chinese literacy learning and implications for teacher development. In P. Duff and E. Lester (Eds), *Issues in Chinese Language Education and Teacher Development*, 70–78. University of British Columbia: UBC Centre for Research in Chinese Language and Literacy Education.

Fischer, R. (2007). How do we Know what Students are Actually Doing? Monitoring Students' Behavior in CALL. *Computer Assisted Language Learning, 20*(5), 409-442. http://dx.doi.org/10.1080/09588220701746013

Gass, S. M. and Mackey, A. (2000). *Stimulated Recall in Second Language Research*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Godwin-Jones, R. (2012). Emerging technologies. Challenging hegemonies in online learning. *Language Learning & Technology (LLT),* 16 (2): 4–13.

Hampel, R. and de los Arcos, B. (2013). Interacting at a distance: A critical review of the role of ICT in developing the learner–context interface in a university language programme. *Innovation in Language Learning and Teaching*, 7 (2): 158–178. http://dx.doi.org/10.1080/17501229.2013.776051

Hampel, R. and Stickler, U., 2012. The use of videoconferencing to support multimodal interaction in an online language classroom. *ReCALL* 24 (2): 116–137. http://dx.doi.org/10.1017/S095834401200002X

Heiser, S., Stickler, U. and Furnborough, C. (2013). Ready, steady, speak online: Student training in the use of an online synchronous conferencing tool. *CALICO Journal, 30* (2): 227–251. http://dx.doi.org/10.11139/cj.30.2.226-251

Hubbard, P., Schulze, M. and Smith, B. (Eds). (2013). Learner-computer interaction in language education: A festschrift in honor of Robert Fischer. *CALICO Journal,* 1–2.

Jewitt, C., Kress, G., Ogborn, J. and Tsatsarelis, C. (2001). Exploring learning through visual, actional and linguistic communication: The multimodal environment of a science classroom. *Educational Review*, 53 (1): 5–18. http://dx.doi.org/10.1080/00131910123753

Just, M. A. and Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology,* 8 (4): 441–480. http://dx.doi.org/10.1016/0010-0285(76)90015-3

Kan, Q. and McCormick, R. (2012). Building course cohesion: The use of online forums in distance Chinese language learning. *Computer Assisted Language Learning* 27 (1): 44–69. http://dx.doi.org/10.1080/09588221.2012.695739

Koda, K. (1992). The effects of lower-level processing skills on FL reading performance: Implications for instruction. *The Modern Language Journal*, 76 (4): 502–512. http://dx.doi.org/10.1111/j.1540-4781.1992.tb05400.x

Lai, C. and Morrison, B. (2013). Towards an agenda for learner preparation in technology-enhanced language learning environments. *CALICO Journal,* 30 (2): 154–162. http://dx.doi.org/10.11139/cj.30.2.154-162

Lee, C. H. and Kalyuga, S. (2011). Effectiveness of different Pinyin presentation formats in learning Chinese characters: A cognitive load perspective. *Language Learning,* 61 (4): 1099–1118. http://dx.doi.org/10.1111/j.1467-9922.2011.00666.x

Liu, M., Moore, Z., Graham, L. and Lee, S. (2003). A look at the research on computer-based technology use in second language learning: A review of the literature from 1990–2000. *Journal of Research on Technology in Education,* 34 (3): 250–273. http://dx.doi.org/10.1080/15391523.2002.10782348

Montoro Sanjosé, C. R. (2012). *The language learning activity of individual learners using online tasks.* Unpublished doctoral thesis. Milton Keynes: The Open University.

Nielsen, J. and Pernice, K. (2010). *Eyetracking Web Usability*. Berkeley, CA: New Riders.

O'Rourke, B. (2008). The other C in CMC: What alternative data sources can tell us about text-based synchronous computer mediated communication and language learning. *Computer Assisted Language Learning,* 21 (3): 227–251. http://dx.doi.org/10.1080/09588220802090253

O'Rourke, B. (2012). Using eye-tracking to investigate gaze behaviour in synchronous computer-mediated communication for language learning. In M. Dooley and R. O'Dowd (Eds), *Researching Online Interaction and Exchange in Foreign Language Education: Theories, Methods and Challenges,* 305–342. Frankfurt am Main: Peter Lang.

Poole, A. and Ball, L. J. (2006). Eye tracking in HCI and usability research. In Ghaoui, C. (Ed.). *Encyclopedia of Human Computer Interaction,* 211–219. Hershey, PA, etc.: IGI. http://dx.doi.org/10.4018/978-1-59140-562-7.ch034

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology,* 62 (8): 1457–1506. http://dx.doi.org/10.1080/17470210902816461

Ren, G.-Q. and Yang, Y. (2010). Syntactic boundaries and comma placement during silent reading of Chinese text: Evidence from eye movements. *Journal of Research in Reading,* 33 (2): 168–177. http://dx.doi.org/10.1111/j.1467-9817.2009.01406.x

Robin, R. M. (2013). CALL and less commonly taught languages. In M. Thomas, H. Reinders and M. Warschauer (Eds), *Contemporary Computer-Assisted Language Learning,* 303–321. London: Bloomsbury Academic.

Satar, H. M. (2010). *Social presence in online multimodal communication: A framework to analyse online interactions between language learners.* PhD, The Open University, Milton Keynes.

Smith, B. (2010). Employing eye-tracking technology in researching the effectiveness of recasts in CMC. In F. M. Hult (Ed.), *Directions and Prospects for Educational Linguistics* (Vol. 11), 79-97: Dordrecht: Springer Netherlands. http://dx.doi.org/10.1007/978-90-481-9136-9_6

Smith, B. (2012). Eye tracking as a measure of noticing: a study of explicit recasts in SCMC. *Language Learning & Technology (LLT),* 16 (3): 53–81.

Spitzberg, B. (2006). Preliminary development of a model and measure of computer-mediated communication (CMC) competence. *Journal of Computer-Mediated Communication,* 11 (2), online. http://dx.doi.org/10.1111/j.1083-6101.2006.00030.x

Stickler, U. and Shi, L. (2013). Supporting Chinese speaking skills online. *System,* 41 (1): 50–69. http://dx.doi.org/10.1016/j.system.2012.12.001

Stockwell, G. (2007). A review of technology choice for teaching language skills and areas in the CALL literature. *ReCALL,* 19 (2): 105–120. http://dx.doi.org/10.1017/S0958344007000225

Sun, S. Y. H. (2011). Online language teaching: The Pedagogical challenges. *Knowledge Management & E-Learning: An International Journal,* 3 (3): 428–447.

Tinsley, T. and Board, K. (2014). *Languages for the Future*. London: The British Council.

Wang, J. (2014). Reading strategies and reading comprehension retention with and without a pop-up dictionary by beginning learners of Chinese. *CALICO Journal,* 31 (2): 244–260. http://dx.doi.org/10.11139/cj.31.2.244-260

Wang, Y. (2004). Distance language learning: Interactive fourth-generation Internet-based videoconferencing. *CALICO Journal,* 21 (2): 373–395.

Wang, Y. and Chen, N.-S. (2007). Online Synchronous Language Learning: SLMS over the Internet. *Innovate, 3*(3), 8 pages (online).

Wang, Y. and Chen, N.-S. (2009). Criteria for evaluating synchronous learning management system: arguments from the distance language classroom. *Computer Assisted Language Learning,* 22 (1): 1–18. http://dx.doi.org/10.1080/09588220802613773

Wang, Y. and Chen, N.-S. (2010). The collaborative language learning attributes of cyber face-to-face interaction: the perspectives of the learner. *Interactive Learning Environments*, 20 (4): 311–330. http://dx.doi.org/10.1080/10494821003769081

Wang, Y., Chen, N.-S. and Levy, M. (2010). Teacher training in a synchronous cyber face-to-face classroom: Characterizing and supporting the online teachers' learning process. *Computer Assisted Language Learning,* 23 (4): 277–293. http://dx.doi.org/10.1080/09588221.2010.493523

White, C. (2009). Towards a learner-based theory of distance language learning: The concept of the learner-context interface. In P. Hubbard (Ed.), *Computer Assisted Language Learning: Critical Concepts in Linguistics,* 97–112. London: Routledge.

Yamada, M. and Akahori, K. (2007). Social presence in synchronous CMC-based language learning: How does it affect the productive performance and consciousness of learning objectives? *Computer Assisted Language Learning,* 20 (1): 37–65. http://dx.doi.org/10.1080/09588220601118503

Yamada, M. and Akahori, K. (2009). Awareness and performance through self- and partner's image in videoconferencing. *CALICO Journal,* 27 (1): 1–25. http://dx.doi.org/10.11139/cj.27.1.1-25

Ye, L. (2011). *Teaching and learning Chinese as a foreign language in the United States: To delay or not to delay the character introduction.* Unpublished dissertation. Georgia State University.

Ye, L. (2013). Shall we delay teaching characters in teaching Chinese as a foreign language? *Foreign Language Annals*, 46 (4): 610–627. http://dx.doi.org/10.1111/flan.12049

Appendix A

Demographic information of participants from pre-study questionnaire:

Participants	Ethnicity	First language	Age
Mai Kemu	white British	English	60+
Cha Li	white British	English	50+
Deng Kan	white British	English	50+
Li Sha	Phillipine	English	50+
Ma Li	white British	English	40+
Wu Xi	white European	German	40+
Ai Mi	Cantonese	Cantonese	40+
Jie Ning	white British	English	30+
Lin Da	white British	English	60+
Su Shan	white European	German	40+

Appendix B

Details of the reading text used for activity 1:

我是法学院的老师，我的学院就在市中心，离火车站很近。学院的对面有一家快餐店，我经常去那里吃午饭，因为那里的饭又便宜又好吃。每个星期三晚上，我去学院旁边的酒吧，常常一边喝酒一边上网。昨天我的女朋友坐火车从伦敦来看我，我们一起去了最喜欢的酒吧跳舞。

wǒ shì fǎ xué yuàn de lǎo shī, wǒ de xué yuàn jiù zài shì zhōng xīn, lí huǒ chē zhǎn hěn jìn. xué yuàn de duì miàn yǒu yī jiā kuài cān diàn, wǒ jīng cháng qù nà lǐ chī wǔ fàn, yīn wèi nà lǐ de fàn yòu pián yí yòu hào chī. měi gè xīng qí sān wǎn shàng, wǒ qù xué yuàn pang biān de jiǔ bā, cháng cháng yī biān hē jiǔ yī biān shàng wǎng. Zuó tiān wǒ de nǚ péng yǒu zuò huǒ chē cóng lún dūn lái kàn wǒ, wǒ men yī qǐ qù le zuì xǐ huān de jiǔ bā tiào wǔ.

Appendix C

Screenprints of interactive activity 2:

Elluminate whiteboard Screenprint for activity 2, Task 1: dragging and dropping images or English translations onto Chinese characters and Pinyin phrases

Elluminate whiteboard Screenprint for activity 2, Task 2: pronunciation practice with the tutor

Elluminate whiteboard Screenprint for activity 2, Task 3: practising phrases by substitution in different contexts.

Elluminate whiteboard Screenprint for activity 2, Task 4: A group speaking practice utilising phrases and the structure learned.