

# Learners' motivational state in CALL practice: a new challenge for HLT-oriented CALL research

S. Bodnar, B. Penning de Vries, C. Cucchiarini, R. van Hout, H. Strik

Centre for Language and Speech Technology, Radboud University Nijmegen

## Introduction

A compelling feature of Computer-assisted Language Learning systems that make use of Human Language Technologies (HLT) are the opportunities they provide for interactive output practice. Development of such systems is challenging because of the variety of expertise required: researchers and developers must satisfy technological, pedagogical, and computer interface usability requirements. Evaluating new systems is therefore important to verify that a system is an effective practice environment. In HLT-enabled systems, evaluation typically has three major foci: performance of the technology, effect on learning gains, and subjective evaluation by learners. Subjective evaluation provides an indication of learners' attitudes towards practice with the system: if their experience is not positive, learners can lose motivation and stop practicing. Generally speaking, the subjective component of system evaluations is limited to post-practice questionnaires which elicit learners' attitudes after practice has finished. Recent developments in L2 motivation theory, however, suggest that alternative approaches are likely to benefit evaluations by providing a more detailed account of the practice experience. Most relevant here are recommendations to 1) capture learners' fluctuating motivational state during practice and 2) contextualise motivational state data with objective records of system and learner behaviour.

## Evaluating learners' situated motivation in CALL: an example

In a recent study, we have attempted to follow these recommendations to evaluate the situated motivational impact of our CALL system. The system, named GREET, provides L2 learners of Dutch with a spoken grammar practice exercise. The exercise contains two parts: Learners begin by watching short (and often humorous) video clips depicting Dutch-speaking characters interacting with each other in everyday situations. Following each video clip, the system quizzes the learner on events that occur in the video. To respond to a quiz question, the learner must record a spoken answer. An Automatic Speech Recognition (ASR) component then analyses the recorded answer and responds with corrective feedback (CF), in cases where an error is detected, or a congratulatory message, if the answer is correct. Figure 1 presents a screenshot of the video and quiz exercise.

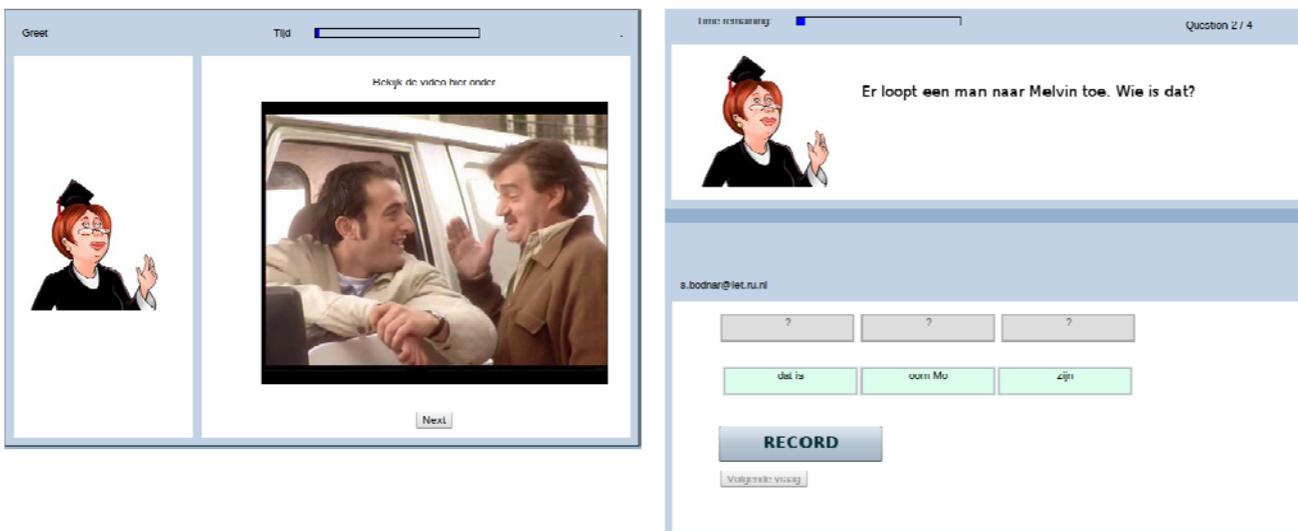


Figure 1. Learners practice speaking Dutch by first watching a video clip (shown left). Next, the system quizzes the learner with questions about events that occur in the clip (shown on the right). Learners first formulate an answer using the word blocks, and then submit their answer to the system by recording themselves speaking it aloud. An ASR component in the system analyses the spoken answer and responds with CF, in the case of errors, or a message of encouragement, if the answer is spoken correctly.

In evaluating GREET, we attempted to monitor how motivation changes during practice and contextualise the subjective data recorded with objective measures of learner-system interaction. We captured subjective data using a periodic 3-item mini-motivation questionnaire designed to survey learners' attitudes towards practice, motivation to practice, and confidence in their ability to do well in the quizzes. To complement this subjective data, we also recorded a variety of system and learner behaviours. We use three metrics to describe learner behaviour in a practice session: number of video clips watched, number of questions attempted, and number of recordings per question.

With the setup described above, we conducted an experiment with 31 adult participants studying Dutch at an A2 level. We organised the participants into two groups who each practiced with a different version of GREET. The CF group used a version of our system that provided immediate feedback and required the learner to produce the correct answer before advancing to the next question. The NOCF group practiced with a version that withheld feedback and the learner was free to advance to the next question whenever they felt ready. Our expectation was that the CF group would have a more positive practice experience and report higher situated motivation levels than the NOCF group due to having their errors corrected. However, there was also the possibility that the confrontational nature of the CF would negatively impact their practice experience.

In analysing the results (trajectories shown in Figure 2), we found that there was no significant motivational difference between the two practice groups. On the one hand, this is surprising, as we expected the CF group to report an overall more positive experience. On the other hand, this finding also confirms that the feedback did not appear to negatively affect the practice experience of the CF group. This is an interesting finding, considering learners in the CF group exerted more effort during practice (on average they made significantly more attempts to complete a question); this extra effort did not negatively impact their motivation. A second finding is that there are indications of improvement in the logs for the CF group. Our analysis indicates that participants in the CF group were able to complete questions with increasingly fewer attempts as they proceeded through practice. Although we did not observe significant differences between the groups on L2 proficiency tests, the logged behaviour suggests that the CF group gradually improves at answering the quizzes.

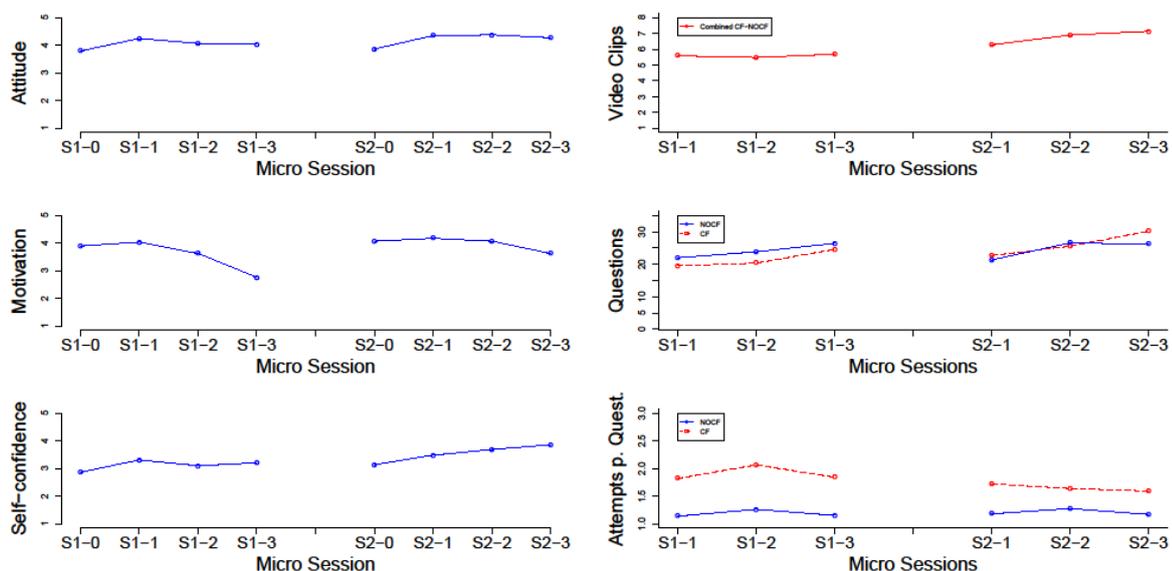


Figure 2. The set of three graphs on the left depict changes in motivational state for the participants. We show the mean for all participants as there was no significant difference observed between the two groups. The top graph is attitude, the middle motivation, and the bottom is self-confidence. The set of three graphs on the right show changes in practice behaviour. The top graph presents mean video clips for all participants (there is no significant difference between the groups) for each 15-minute micro-session. The middle graph shows the mean number of questions completed, while the bottom graph depicts the mean number of attempts per question. In both cases a significant group-time interaction was observed between the groups. In these graphs, the dashed red line corresponds to the CF group, while the solid blue line shows the NOCF group trajectory.

### *Future directions*

The experiment described above is an example of how an analysis of situated motivation can be included in the evaluation of a CALL system. The methods used above, questionnaires and log analyses, were sufficient for studying fluctuations of motivation. However, we believe more can be done to exploit technology for the purpose of evaluating motivational state. Applications of HLT for measuring motivational state are a virtually untouched area in CALL, despite recent advances in affective computing, such as automatic emotion recognition in human speech and sentiment analysis in written modes of communication. Aimed towards HLT specialists working in CALL research, the final part of this presentation aims to draw attention to the problem of diagnosing motivational state using HLT. We point to a number of areas in HLT that could help to develop an automatic, unobtrusive and temporally fine-grained measure of motivational state. The development of such a measure has the potential to contribute to experiments such as the one above, and more generally, educational practices and our understanding of second language acquisition processes.