

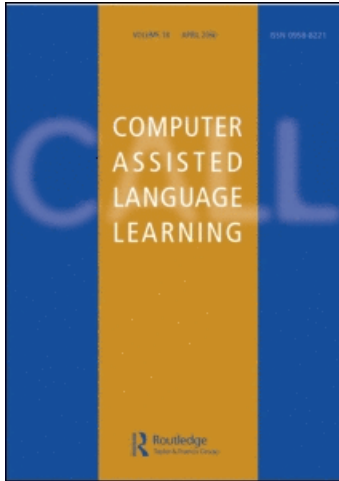
This article was downloaded by: [Radboud University Nijmegen]

On: 15 October 2010

Access details: Access Details: [subscription number 907171856]

Publisher Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Computer Assisted Language Learning

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title~content=t716100697>

Speech interactive computer-assisted language learning: a cross-cultural evaluation

Hazel Morton^a; Mervyn Jack^a

^a CCIR, University of Edinburgh, Edinburgh, UK

Online publication date: 16 September 2010

To cite this Article Morton, Hazel and Jack, Mervyn(2010) 'Speech interactive computer-assisted language learning: a cross-cultural evaluation', *Computer Assisted Language Learning*, 23: 4, 295 – 319

To link to this Article: DOI: 10.1080/09588221.2010.493524

URL: <http://dx.doi.org/10.1080/09588221.2010.493524>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Speech interactive computer-assisted language learning: a cross-cultural evaluation

Hazel Morton* and Mervyn Jack

CCIR, University of Edinburgh, AGB Building, King's Buildings, Mayfield Road, Edinburgh, EH9 3JL, UK

The evaluation in real time of a computer-assisted language learning (CALL) program with end users is an invaluable stage in the design and development of such a program. By using a sample of target users in the evaluation, it is possible to gather information on attitudes towards the program and performance of the program with an intended population. However, such users come to the research with personal attitudes towards the target language and the learning process. This article describes the evaluation of a speech-interactive CALL program that creates an environment in which the learners, using speech recognition technology, can converse with virtual characters in the target language in real time. A cross-cultural empirical evaluation was adopted where two groups of users participated in the research: students of French in Scotland and students of English as a foreign language in mainland China. The evaluations sought to investigate user attitudes towards the speech-interactive CALL program and the performance of the system. In addition, users' motivations towards studying the target language were investigated. Differences were found between the two groups of learners with regards to their motivation towards studying the target language, which may be considered in relation to their differences in attitude towards the speech interactive CALL program.

Keywords: speech recognition technology; speech interactive CALL; motivation; usability; virtual characters; animated pedagogical agents

Introduction

User-centred design and usability studies are common in the field of human-computer interaction (HCI). Some researchers in the field of computer-assisted language learning (CALL) advocate the consideration of usability issues in the design and evaluation of CALL systems (Allum, 2001). However, when using a cohort of 'end users' from a potential target group, individual differences and attitudes to their studies may have an impact on their overall attitudes to the system. Users may come to the research sessions with their own perceptions and feelings towards learning the target language. Such feelings may impact upon the users' attitudes towards using the program. Therefore, it may be appropriate to investigate such attitudes towards learning the target language alongside investigation of attitudes towards the CALL program.

*Corresponding author. Email: hmorton@ccir.ed.ac.uk

The purpose of the study described in this article is to investigate user perceptions of and attitudes towards the speech interactive CALL program and to investigate the users' attitudes towards learning the target language and whether the use of the program had any impact upon those attitudes or whether those attitudes help to understand further the users' attitudes to the program. Data from both system performance and user motivation were collected as part of this study. The data are presented on user attitudes towards using the program and interacting with the characters. User attitudes towards learning the target language (pre and post CALL program usage) are then presented. In addition, user response data on the utterances made while interacting with the characters are presented together with performance data on the speech recognition component of the program.

Background to the study

A previous evaluation showed that the CALL program was found to be enjoyable to use and had high satisfaction scores by learners of Italian and Japanese (Morton, Davidson, & Jack, 2008). This article describes an evaluation of the CALL program for more commonly taught languages. The evaluation was conducted with two groups of language learners: secondary school pupils in Scotland learning French and junior high school students in China learning English as a foreign language. Although it would be impossible to account for the differences between the two groups of learners, these groups were seen to have similar characteristics. Both are foreign language rather than second language contexts. The languages are likely to be taught principally by non-native speakers of that language (this was found to be the case for the sample groups involved in the evaluation). Finally, there are limited opportunities for practising the language outside the classroom. Although these two learning environments have some similar characteristics, the participants involved may have differing attitudes towards the CALL program and in their motivations towards learning the target language.

CALL program

The program described in this article, SPELL, combines virtual worlds and virtual agents with automatic speech recognition technology to create a speech interactive CALL application in which learners can interact in the target language with virtual characters who 'listen' by means of a speech recogniser. The 3-D virtual worlds, created in virtual reality modelling language (VRML), depict the contextualised environment in which the interaction takes place.

The aim in the program is for learners to engage in a dialogue with the virtual characters within a defined context. Research in the use of animated agents in pedagogical applications has found that animated agents can improve the learning experience (Lester et al., 1997; Massaro, 1998). It has also been found (Johnson, Rickel & Lester, 2000) that animated pedagogical agents are able to increase the computer's ability to engage and motivate students. Based on the interaction hypothesis (Long, 1996), the virtual characters are designed to offer modifications of their input in cases where the user appears to be having difficulties. Interaction provides learners with opportunities to receive comprehensible input and feedback (Gass, 1997; Long, 1996; Pica, 1994). Further, interaction allows learners to make changes to their own linguistic output (Swain, 1985, 1995). It has been suggested (Garcia-Carbonella, Rising, Montero, & Watts, 2001) that the use of simulations in

language education is beneficial to learners as they can enhance interactions through negotiated meaning. In the SPELL program, the learners are not told in advance what to say, nor are they given a finite list from which to choose their utterances; the speech recognition grammars are programmed with predicted responses for each individual stage of the dialogue, accounting for grammatical and some ungrammatical responses.

It has been suggested (Wachowicz & Scott, 1999) that implicit feedback is preferable to corrective feedback for speech interactive CALL systems, as implicit feedback is likely to minimise potential problems resulting from imperfect speech recognition. Feedback in the SPELL program is given implicitly in the form of recasts and reformulations. If the system detects that the learner has made an error in their utterance, the animated agent recasts the learner's utterance. If the learner does not respond, the animated agent repeats the question. If the system detects that the learner has given an answer that is not appropriate to the given stage, the system 'rejects' this and the animated agent reformulates the question, possibly offering a hint to the learner. These feedback strategies allow the dialogue with the learner to continue without explicit reference to a problem. This has the advantage of continuing the flow of the dialogue (and where necessary giving the learner another opportunity to respond, or implicitly correcting their response); and, by being implicit in the feedback, this minimises attention to any potential errors made by the speech recognition component.

The SPELL program offers the learner three scenario types within each 'lesson': observational, one-to-one and interactive. Supplementary materials are also available to the learners to access if they require: vocabulary, grammar files, a transcription of the observational dialogue and cultural information. The observational scenario gives the learner an opportunity to observe the animated agents within the scene engage in a contextualised dialogue. The learner can watch the interaction between the animated agents in this scenario. Figure 1 shows the animated agents in the observational scenario.

A series of one-to-one scenarios, entitled in the system 'Talk to John' or 'Parle à Jean' are available. Each one-to-one scenario raises a topic which is relevant to the

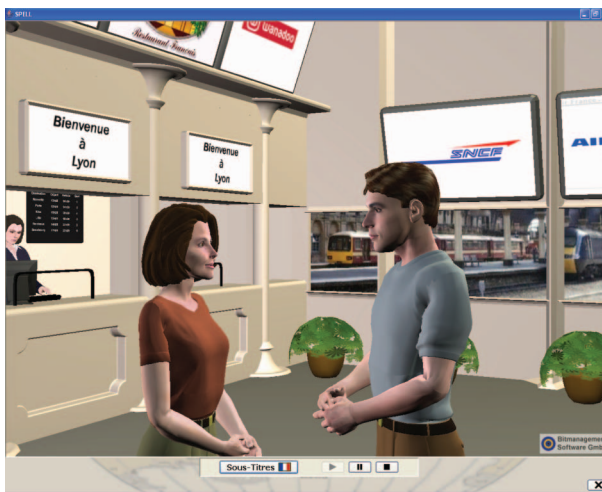


Figure 1. Animated agents in observational scenario.

lesson. The learner, using headphones and a microphone, is asked a number of questions relating to the given scene by one of the animated characters. For example,

A quelle heure part le train pour Nice?

These questions introduced topics and sentence structures which are relevant in the given context and which will be useful for their participation in the interactive scenario (INT). The one-to-one scenarios give the learner the opportunity for extended sentence practice prior to becoming 'immersed' in the INT.

The INT creates an environment in which the learner acts as an active dialogue participant. In this example, the learner 'enters' the virtual railway station and orders tickets at the counter. Figure 2 shows the virtual agents interacting with the learner in the INT.

The INT presented to the learner is the station scene as depicted in the observational scenario. In the 'At the station' scene, the goal is to purchase tickets to your preferred destination in the host country. The ticket agent asks the learner where they would like to go and subsequently takes them through a series of questions in order to sell the train ticket.

In the context of CALL, it has been suggested (Chapelle, 1998) that it may be important for learners to have an audience for their linguistic output so that learners can 'attempt to construct meanings for communication rather than solely for practice' (p. 24). A key element in the design of SPELL is that learner input in the dialogue is necessary for the dialogue to continue. In the case of such transactional dialogues, whatever the learner says has consequence for the rest of the dialogue. Errors from the learner will either result in the agent giving implicit feedback in the form of a recast, or will prompt the agent to reformulate the initial proposition so that the learner can respond again.

A commercially available speaker-independent recogniser that had been trained on native speaker models only was used in the SPELL program. However, the

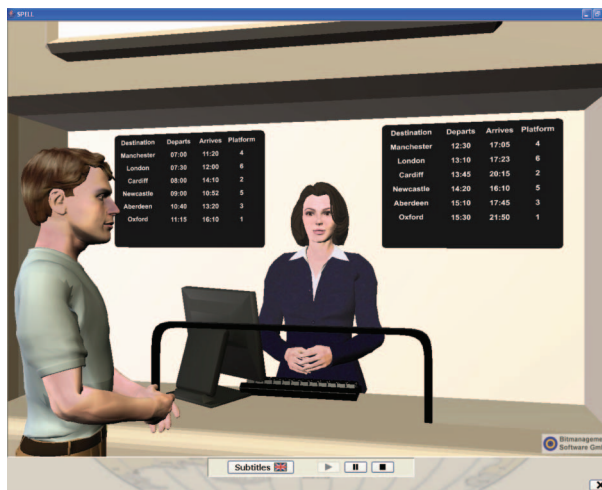


Figure 2. Animated agents in interactive scenario.

recognition grammars in the SPELL program were created specifically for the non-native speaker learners of the program, including both grammatical and ungrammatical utterances constrained to each stage in the interaction. By using individual recognition grammars for each stage, the possible utterances to each given dialogue stage are constrained thus limiting the list from which the recogniser attempts to make a match.

Grammatical errors accounted for in the grammar recognition files for the French lesson included verb inflection errors, article errors, pronoun errors and noun–adjective disagreement. For example, ‘*Jean et Sylvie va à Paris*’ is included in the recognition files, and is flagged as containing a verb inflection error. For the English version, grammatical errors accounted for in the grammar recognition files included omission of third person present ‘-s’, use of present progressive for simple present and article errors. For example, ‘*The train leave at 6*’ is included in the recognition files, but is flagged as containing an omission of third person present ‘-s’ error. The recast command is triggered if the utterance is within the ‘error’ category. This category contains those items that are ungrammatical as well as included responses which are not full sentence utterances.

For a full description of the speech recognition grammars used in SPELL and further details on the design, see Morton and Jack (2005).

Evaluation of SPELL lesson: ‘At the railway station’

The evaluation of the SPELL program took the form of a short, standardised procedure in which target users interacted with a fully functional prototype of the system. Researchers were present throughout in order that any problems or issues arising in the use of the system could be observed and recorded. This approach has the advantage that the researcher may notice aspects of the interaction which the user is unaware of; it also allows the researcher to guide the user through the session in a pre-designed path so that each user who takes part experiences, as much as possible, the same procedure.

This evaluation sought to make an investigation of the ‘At the station’ lesson across two languages, French and English. Investigation was sought on various aspects of the design and interactive elements of the program:

- (1) User attitudes to using the program and interacting with the characters.
- (2) User motivation towards learning the foreign language.
- (3) User response types and the recognition accuracy of the speech recognition component.

Participants

A total of 76 participants took part in the evaluation of the program: 28 students of French and 48 students of English as a foreign language. Participants in the French group were recruited from three local secondary schools in Edinburgh, Scotland. The evaluations took part in the host schools, in a dedicated room. The evaluation of the English as a foreign language (EFL) version was conducted on location in a junior high school in Beijing, China. Table 1 details the participants in this evaluation. Participants were aged between 14 and 16 years at the time of the evaluation. The students had been studying the respective languages for varying

lengths of time. Overall, the French group had been studying French for an average of 4.7 years; the EFL group had been studying English for an average of 6.8 years.

Experimental procedure

The participants were first asked to complete a self-administered motivation questionnaire (M1) which consisted of 16 statements on a seven-point Likert (1932) style scale. They were then given a short tutorial on using the program (using the navigation and functionality controls, accessing the supplementary materials). After this, the participants were asked to attempt various aspects of the 'At the railway station' lesson. The participants were asked to watch the observational scenario, then try two of the one-to-one scenarios (here referenced as O-O 1 and O-O 2) and then try the INT. The participants were informed that they could access other features in the program, for example subtitles or vocabulary, as they wished. The researcher remained present during the program use. After each scenario, the participants were asked to complete an attitude questionnaire. Upon completion of all the scenarios, the participants were then asked to complete the motivation questionnaire (M2).

User attitude questionnaires

User attitude questionnaires were used for each of the different scenarios that the students experienced in the lesson: observational, one-to-one and INTs. The usability questionnaire was created in order to gather attitude to each of the scenarios that the participants experienced. The questionnaire consists of a series of short, simple statements, each with a set of tick-boxes on a seven-point scale labelled from 'strongly agree' through 'neutral' to 'strongly disagree', see Figure 3.

The polarity of the statements is balanced to avoid the response acquiescence effect, where respondents may have a natural tendency to agree with proposals. The usability attributes covered affective, engagement and interaction issues. A set of 14 statements was used in the questionnaire for the one-to-one and INTs. The observational questionnaire omitted statements specific to the interaction issues (an eight-statement questionnaire was used for the observational scenario). The questionnaire was translated into Mandarin for the EFL group.

Table 1. Experimental participants.

	Males	Females	Total
French	14	14	28
EFL	22	26	48
Total	36	40	76

	Strongly Agree	Agree	Slightly Agree	Neither agree nor disagree	Slightly Disagree	Disagree	Strongly Disagree
Q1 I felt stressed talking to the character.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Figure 3. Example user attitude questionnaire statement.

When analysing the results, responses to the questionnaire are first given a numerical value from 1 to 7; these values are then normalised for the polarity of the statements such that a ‘strongly agree’ response to a positive statement is given a value of 7, whereas a ‘strongly agree’ response to a negative statement is given a value of 1. After normalisation of the data, the overall attitude for each participant can be calculated as a mean of all of the scores on the items in the questionnaire. These values can then be used to calculate the overall attitude for all items in the questionnaire across all participants in the study. Additionally, mean scores for individual items in the questionnaire can be obtained for all participants.

Motivation questionnaires

Motivation in language learning is believed to be a predictor of success. In an analysis of 75 motivation studies, Masgoret and Gardner (2003, p. 205) concluded that ‘motivation is more highly related to second language achievement’ than any other factor. There are various kinds of motivation, however, with some believed to indicate a higher potential for success than others. Research which makes comparisons between people whose motivation is intrinsic and those whose motivation is extrinsic has found that those with intrinsic motivation have more interest and confidence in the activity, resulting in enhanced performance or persistence (Deci & Ryan, 1991) and enhanced self-esteem (Deci & Ryan, 1985).

In investigating such learner orientations to second language learning, some researchers (e.g. Noels, 2001) advocate the use of Deci and Ryan’s (1985) self-determination theory in conceptualising a framework for understanding second language learners’ orientations to learning.

It is suggested (Deci & Ryan, 1985) that a person’s reason for performing an activity can be understood in terms of the degree to which it is perceived as being freely chosen and endorsed by themselves and therefore self-determined. According to Ryan and Deci, ‘people can be motivated because they value an activity or because there is strong external coercion’. (Ryan & Deci, 2000, p. 69). Motivation can be intrinsic, where the motivation comes from the inherent pleasure in the activity, or extrinsic, where the motivation comes from the desire for an outcome of performing the activity. Self-determination theory offers a framework of motivation where a variety of motivational orientations are arranged along a continuum which range from least to most self-determined orientations (Deci & Ryan, 1985; Ryan & Deci, 2000). Figure 4 illustrates the self-determination continuum, adapted from Ryan and Deci (2000).

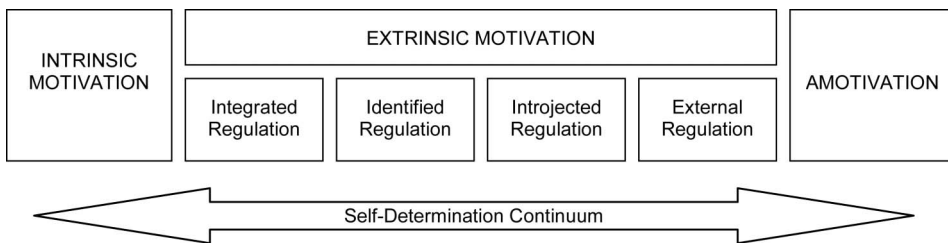


Figure 4. Self-determination continuum.

The most self-determined orientation is intrinsic motivation where the motivation to perform the activity comes from the enjoyment and pleasure derived from participation in the activity. Intrinsic motivation has been further defined by three interrelated types of motivation: knowledge (the pleasure of learning new things), accomplishments (the pleasure from undertaking new challenges) and stimulation (the enjoyment gained from the experience) (Vallerand, 1997). Extrinsic motivation is divided into four regulations. Integrated regulation refers to a state where a person undertakes an activity because it supports a valuable component of the person's self-identity. Identified regulation refers to the state where the person undertakes the activity because they believe it will help them to achieve a personal goal. Introjected regulation refers to the state where the person undertakes an activity in order to avoid guilt or anxiety or to pursue an ego enhancement. External regulation refers to the state where the person undertakes the activity to receive a reward, satisfy a demand or avoid punishment (external to themselves). Amotivation refers to the absence of motivation or 'the state of lacking the intention to act' (Ryan & Deci, 2000, p. 72).

Based on previous research which adopted the self-determination continuum to the study of language learning motivation (Noels, Clément, & Pelletier, 1999; Noels, Pelletier, Clément, & Vallerand, 2000), a 16-item motivation questionnaire was developed for this evaluation which included statements for each of the motivation types discussed above. The questionnaire was created in Likert format in order that participants could rate the intensity of their agreement to any particular statement. For each of the above motivation types, two statements were included in the questionnaire. Table 2 details these statements together with the corresponding

Table 2. Motivation statements per type.

Motivation type	Statement
Intrinsic – knowledge	I get a satisfied feeling when I learn new words and phrases in French. I learn French for the enjoyment I get from learning about other people and cultures.
Intrinsic – accomplishment	I enjoy the challenge of speaking in French. I feel very satisfied when I make progress in French.
Stimulation	I get a good feeling when I speak in French. I get a good feeling when I can understand French.
Integrated regulation	I learn French because I want to be able to communicate with French speaking people. I learn French because it is important in my life to interact with people who speak French.
Identified regulation	I learn French because I think it will be good for my personal development. I learn French because I choose to be the kind of person who can speak a second language.
Introjected regulation	I learn French because it is important for me to show others I can speak a second language. I learn French because I would feel ashamed if I couldn't speak in a second language.
External regulation	I learn French because it is expected of me. I learn French so I can get a good grade.
Amotivation	I don't see the point in learning French. I take French because it is compulsory. I will drop it as soon as I can.

motivation type. In contrast to the usability questionnaires, the motivation questionnaire was not balanced for positively and negatively worded statements, as the sentiments in the statements did not readily lend themselves to being reworded.

The questionnaire aims to highlight how the students in the experiment are motivated to learn the target language. The questionnaire was administered at the beginning of the session to gather data on students' general attitudes towards learning the target language. At the end of the session, the questionnaire was administered again to investigate whether using the CALL program had any effect on their motivation towards learning the target language.

Although the motivation questionnaires are designed to be used as a within-subjects tool, that is to compare motivation before and after using the CALL system for the two language groups separately, some comparisons were also made between the two language groups on the first motivation questionnaire data. The results are described below.

Results

The following details the results of the usability questionnaires, motivation questionnaires, verbal interviews and recognition accuracy rates.

User attitude results

The Likert questionnaires contained items which focused on affective issues, engagement issues and issues relating to the interaction with respect to the dialogue itself and with respect to the content within the interaction. Participants completed a Likert questionnaire following each of the scenarios: one-to-one scenario 'about train times' (O-O 1), one-to-one scenario 'about journey details' (O-O 2) and the INT. The user attitude results for each of the scenarios are detailed here, and are provided for the two language groups separately.

User attitude results – EFL

The EFL group gave an overall mean score of 5.03 (on a seven-point scale) for one-to-one scenario 'about train times', an overall mean score of 5.43 for one-to-one scenario 'about journey details', and an overall mean score of 5.58 for the INT. Table 3 details the overall mean scores for the speech-enabled scenarios.

It can be seen that attitude scores for the individual items increased across the three consecutive speech-enabled scenarios, indicating that in the EFL group, participants' attitudes are affected by a learning effect of interacting in the scenarios.

Repeated measures analysis was conducted across the attitude data for the three speech-enabled scenarios. Comparing the O-O 1 against O-O 2, it was found that each of the affective issues scored significantly higher in the second scenario than in the first. Participants felt significantly more in control ($p = 0.033$); they felt highly significantly less embarrassed ($p = 0.000$); they felt highly significantly more relaxed ($p = 0.000$); and they felt significantly less stressed ($p = 0.011$). Additionally, a preference for speaking the language in class, in comparison to speaking with the animated characters was highly significantly less in the second scenario than in the first ($p = 0.000$), and the feeling that the character did not understand them was

Table 3. Overall mean scores speech enabled scenarios – EFL.

Questionnaire statement	'O-O 1' (Mean = 5.03)	'O-O 2' (Mean = 5.43)	Interactive (Mean = 5.58)
I felt in control when talking to the character.	4.50	4.90	5.08
I felt embarrassed when talking to the character.	4.31	5.25	5.52
I felt relaxed talking to the character.	4.52	5.58	5.58
I felt stressed talking to the character.	4.42	4.96	5.50
I enjoyed interacting with the character.	5.15	5.46	5.81
I prefer speaking English in class, rather than interacting with the character.	4.75	5.19	5.31
I would be happy to talk to the character again.	5.54	5.77	6.19
I felt that this interaction was useful for my learning of English.	6.33	6.33	6.31
I felt I always understood what the character said.	6.06	6.23	6.38
I felt I always knew how to respond to the character.	5.31	5.65	6.06
I felt that the character did not understand what I said.	4.49	5.27	5.25
I felt the character was difficult to understand.	5.58	5.84	5.90
I felt that the level of the language was difficult for me to understand.	5.81	6.08	6.13
I felt that this dialogue was too easy for me.	3.63	3.44	3.15

highly significantly less in the second scenario than in the first ($p = 0.009$). The issue of the character understanding them better in O-O 2 compared to O-O 1 will be taken up in the discussion of the speech recognition results later in this article.

Comparing the O-O 2 against the INT (the second one-to-one scenario was completed immediately prior to the INT) found significant differences amongst some individual items in the questionnaire. Participants felt highly significantly less stressed in the INT than in the second one-to-one scenario ($p = 0.002$); participants were highly significantly more happy to try the INT again ($p = 0.005$) and participants were highly significantly more confident that they knew how to respond in the INT ($p = 0.009$).

User attitude results – French

The French group gave an overall mean score of 4.90 (on a seven-point scale) for one-to-one scenario 'about train times', an overall mean score of 4.74 for one-to-one scenario 'about journey details', and an overall mean score of 4.55 for the INT. The mean scores for each of the individual items are given in Table 4.

It can be seen that attitude scores for the individual items did not increase in the same way for the French group as they did for the EFL group across the three consecutive speech-enabled scenarios.

The overall mean score decreased across the three speech-enabled scenarios. Repeated measures analysis was conducted across the attitude data for the three

Table 4. Overall mean scores speech enabled scenarios – French.

Questionnaire statement	'O-O 1' (Mean = 4.90)	'O-O 2' (Mean = 4.74)	Interactive (Mean = 4.55)
I felt in control when talking to the character.	4.54	4.39	4.21
I felt embarrassed when talking to the character.	3.96	4.25	4.54
I felt relaxed talking to the character.	3.89	4.36	4.14
I felt stressed talking to the character.	5.00	5.11	4.56
I enjoyed interacting with the character.	4.79	4.79	4.50
I prefer speaking French in class, rather than interacting with the character.	4.18	4.00	3.96
I would be happy to talk to the character again.	5.14	5.21	5.11
I felt that this interaction was useful for my learning of French.	5.79	5.64	5.79
I felt I always understood what the character said.	6.04	5.07	4.82
I felt I always knew how to respond to the character.	5.21	4.64	4.18
I felt that the character did not understand what I said.	4.04	4.07	3.18
I felt the character was difficult to understand.	6.21	5.64	5.11
I felt that the level of the language was difficult for me to understand.	6.36	5.33	5.32
I felt that this dialogue was too easy for me.	3.43	3.86	4.32

speech-enabled scenarios. Comparing the O-O 1 against O-O 2, it was found that participants felt that they understood the characters less in O-O 2 than in O-O 1 ($p = 0.023$). Highly significant results were also found for two of the interaction-content items. Participants felt that they found the characters in O-O 2 more difficult to understand ($p = 0.002$) and that the level of the language was more difficult ($p = 0.003$). The first one-to-one scenario focussed only on train time information while O-O 2 first asked which destination in the host country the learner would like to go to and proceeded to ask about train time and platform information. There was more variety in this one-to-one scenario and therefore learners would not be able to predict what they would be asked next. This may account for the significant differences between some of the interaction-content items in the questionnaire.

In comparing O-O 2 against the INT, participants felt significantly more stressed in the INT than in the second one-to-one scenario ($p = 0.017$). Participants felt that the characters in the INT understood them less than the character in the one-to-one scenario ($p = 0.035$). Further, participants felt that they found the characters in the INT more difficult to understand ($p = 0.041$) and that the one-to-one scenario dialogue was easier than the INT dialogue ($p = 0.025$).

A comparison of attitude results

The EFL group's attitudes towards the speech-enabled scenarios were different to the French group's attitudes. The EFL group exhibited an increase in attitudes in their progression through the lesson, whereas the French group exhibited a decrease. Additionally, the EFL group exhibited less stress and anxiety the more they progressed through the speech-enabled scenarios. However, the French group felt that the characters were more difficult to understand as they progressed through the scenarios and that the content became more difficult as they progressed.

The lessons are designed to offer a progression through the 'lesson', starting with the observational scenario, practising the language with a series of question and answer dialogues in the one-to-one scenarios and culminating in the interaction scenario where the learner accomplishes the transactional task within the virtual scene. The EFL group appeared to become more relaxed/less anxious as they progressed through the lesson whereas the French group appeared to focus on the dialogues becoming more difficult. It is possible that an inherent difference in attitude to learning the languages is present within the two groups. If the EFL group relish the challenge of the lessons and become more relaxed in their interaction with the characters, this may be in part due to how they feel about learning the language. If the French group notice the difficulty of the target language as the lesson progressed, this may in part be due to their attitudes towards learning the language itself. It is possible that the motivation results could help to explain the differences found here with respect to attitudes towards the program or interacting with the characters.

Motivation results

The results from motivation questionnaires before and after experiencing the CALL system are described separately for each language group. A within-subjects comparison was made for each of the two language groups. Additionally, some analysis was made to compare the motivation of the two language groups before they experienced the CALL system. This highlights some fundamental differences between the two groups of learners on what motivates them to learn the target language.

EFL group

Students completed the motivation questionnaire at the beginning of the session, prior to using the application to gather their attitudes in general to studying English. After using the application, the questionnaire was administered again to gauge their attitudes to studying English with the recent experience of using the CALL program. The mean scores of pre- and post-motivation attitudes are given in Table 5.

As can be seen from this table, each of the individual attributes in the intrinsic motivation section increased after usage with the CALL system. As intrinsic motivation is seen as an indicator of success in foreign language learning, this is a positive result for the CALL system. Many of the extrinsic motivation attribute scores fell slightly after use of the system. In other words, the students were less likely to assign positively the extrinsic attributes to their learning. The external regulation feature of 'getting a good grade' increased after usage of the system. However, this may be expected given the surroundings in which the testing took place (i.e. school

Table 5. Pre- and post-usage motivation scores – EFL.

Questionnaire item	Motivation 1	Motivation 2
<i>Intrinsic motivation</i>		
Knowledge – phrases	6.10	6.23
Knowledge – culture	5.79	5.96
Accomplishment – challenge	5.85	6.12
Accomplishment – progress	6.02	6.15
Stimulation – good feeling	5.94	6.31
Stimulation – understanding	6.15	6.40
<i>Extrinsic motivation</i>		
Integration – communicate	6.25	6.21
Integration – interact	6.02	6.40
Identified – develop	6.13	5.75
Identified – person	5.15	5.13
Introjected – L2	5.08	4.69
Introjected – ashamed	2.45	2.23
External regulation – expected	2.38	2.19
External regulation – grade	3.40	3.90
<i>Amotivation</i>		
Amotivation – pointless	1.83	1.50
Amotivation – compulsory	1.67	1.46

premises). It is perhaps difficult for students to disassociate any activity that takes place in school from the aims of the school system. Finally, both of the amotivation attribute scores fell after use of the system. Although the amotivation scores were low to begin with, this is a very encouraging result as it suggests that for this group of learners, the CALL system could decrease feelings of amotivation.

A within-subjects repeated measures analysis (GLM) was performed on the pre- and post-usage motivation scores. Feelings of accomplishment with the challenge of speaking in English increased significantly after use with the CALL system ($p = 0.036$). Both stimulation attributes were found to have highly significant increases after use with the CALL system: *a good feeling when speaking in English* ($p = 0.000$) and *a good feeling when understanding English* ($p = 0.009$). The integrative attribute *'important to interact with people who speak English'* also significantly increased ($p = 0.007$) after use with the CALL system. The identified attribute *'good for my personal development'* significantly decreased ($p = 0.005$) after usage with the CALL system. This does not suggest that participants feel that learning English is *not* good for their personal development, but rather that learning simply *for* personal development became less of a reason after usage with the system. The introjected attribute *'it is important for me to show others I can speak a second language'* also significantly decreased ($p = 0.036$) after using the CALL application, indicating that this external motivation became less of an issue after students experienced the CALL system.

Upon usage of the CALL application, scores for *'getting a good grade'* significantly increased ($p = 0.030$). As mentioned above, this may in part be due to the environment in which the testing took place and the pressures for school students to pass exams. However, Ellis (1985) warns that 'we do not know whether it is motivation that produces successful language learning, or successful learning that

enhances motivation' (Ellis, 1985, p. 119). Both amotivation attributes significantly decreased after usage of the CALL system. Students were less likely to see the study of English as being pointless ($p = 0.003$) after usage with the CALL system; students were less likely ($p = 0.011$) to feel that they only study English because it is compulsory.

French group

The mean scores of pre- and post-motivation attitudes for the French group are given in Table 6.

As can be seen from this table, most of the individual attributes in the intrinsic motivation section increased after usage with the CALL system. A slight decrease was found for the individual item '*know about culture*'. Also, many of the extrinsic motivation attribute scores fell slightly after use of the system. In other words, the students were less likely to positively assign the extrinsic attributes to their learning. However, the external regulation statement ('grade') increased after usage of the system, as with the EFL group. Again, this may be due to the school environment in which the testing took place. It was also interesting to find that the introjected statement of '*shame*' increased with the post-usage questionnaire. This may have been influenced by the participants' feelings of anxiety when interacting with the system if they felt they did not know how to respond. Finally, both amotivation attribute scores decreased after use of the system.

A within-subjects repeated measures GLM analysis was performed on the data from the two motivation questionnaires. Overall, only one item in the questionnaire produced any statistically significant differences for the pre- and post-usage motivation questionnaire. The identified attribute '*good for my personal development*' significantly decreased ($p = 0.013$) after usage with the CALL system. Again, this

Table 6. Pre- and post-usage motivation scores – French.

Questionnaire item	Motivation 1	Motivation 2
<i>Intrinsic motivation</i>		
Knowledge – phrases	5.82	5.82
Knowledge – culture	4.89	4.64
Accomplishment – challenge	5.46	5.79
Accomplishment – progress	5.85	6.04
Stimulation – good feeling	5.21	5.43
Stimulation – understanding	6.21	6.21
<i>Extrinsic motivation</i>		
Integration – communicate	5.79	5.68
Integration – interact	4.82	4.79
Identified – develop	5.64	5.25
Identified – person	4.96	4.68
Introjected – L2	5.36	5.00
Introjected – ashamed	3.43	3.71
External regulation – expected	3.19	3.19
External regulation – grade	4.79	5.07
<i>Amotivation</i>		
Amotivation – pointless	2.54	2.21
Amotivation – compulsory	2.93	2.71

may not suggest that participants feel that learning English is *not* good for their personal development, but rather that learning simply *for* personal development became less of a reason after usage with the system.

Comparison of motivation results

In order to make comparisons between the two language groups prior to using the system, data from the motivation questionnaires were analysed for the pre-usage scores of the two language groups. These compare the motivation of the two groups of participants prior to their use of the system, and therefore give a general comparison of the two groups, given their prior experience with studying the target language.

A division was found with the two language groups between the two sides of the self-determination continuum. Excepting one statement (stimulation – understanding), the EFL group score higher mean scores in all the intrinsic motivation categories and the two internally regulated sub-category items, whereas the French group score higher in the externally regulated sub-categories and the amotivation items. This suggests an overall division in motivating regulation between the two learner groups, which is perhaps a cultural one. These results are depicted graphically in Figure 5.

A between-subjects independent samples *t*-test was performed on the data in order to find any significant differences between the two groups for each of the individual items in the pre-usage motivation questionnaire. The EFL group had significantly higher ratings for two intrinsic motivation items: they indicated a tendency towards their reasons for learning the target language *for the enjoyment*

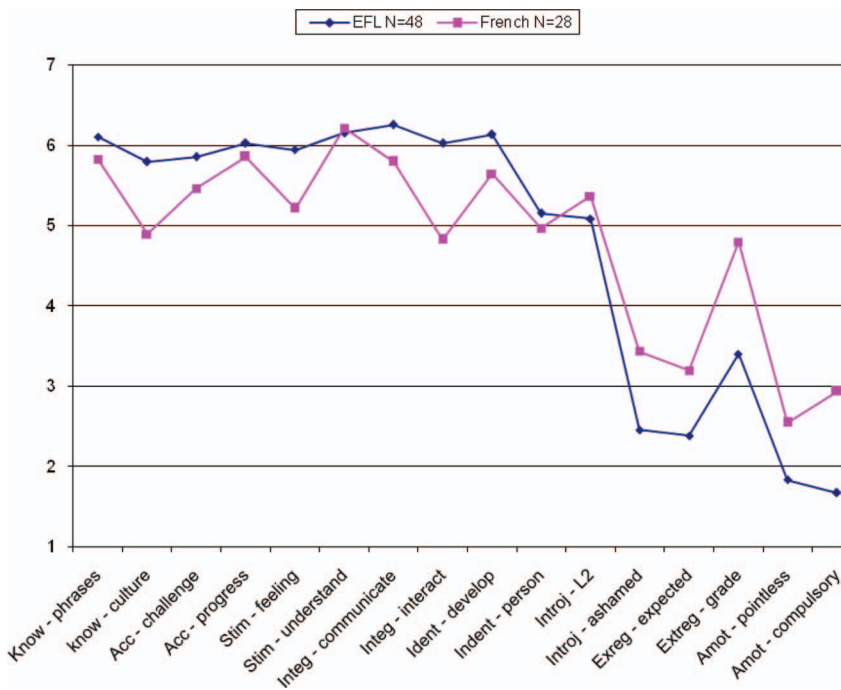


Figure 5. Pre-usage motivation scores EFL vs. French.

they get from learning about other people and cultures ($p = 0.002$); the EFL group also showed significantly stronger results towards the intrinsic motivation item of *getting a good feeling when I speak in the target language* ($p = 0.001$). Additionally, the EFL group expressed significantly stronger tendencies towards learning the target language because *'it is important in my life to interact with people who speak' the target language* ($p = 0.000$). This item is part of the extrinsic motivation, belonging to the integrated regulation on the self-determination continuum. Integrated regulation is the most self-determined of the extrinsic motivation categories. According to Deci and Ryan (2000, p. 73), 'actions categorised by integrated motivation share many qualities with intrinsic motivation, although they are still considered extrinsic because they are done to attain separable outcomes rather than for their inherent enjoyment'. However, it can be seen that the EFL group had significantly higher results in the more self-determined, or autonomous sides of the continuum.

In contrast, the French group had significantly higher scores for those items at the opposite end of the continuum; the French group scored significantly higher results for the more externally regulated motivation than the EFL group. The French group were significantly more motivated to learn French because they *'would feel ashamed if they couldn't speak in a second language'* ($p = 0.007$). External pressures of rewards or punishment were also significantly more important to the French group than the EFL group. The French group were significantly more likely to learn the second language because it was *'expected of them'* ($p = 0.019$) or in order to *'get a good grade'* ($p = 0.000$) than the EFL group.

These results do not show that the French group were more extrinsically motivated than they were intrinsically motivated. The results for both groups show that they indicated stronger feelings towards the intrinsic motivation items than the extrinsic motivation items. However, the EFL group was more strongly intrinsically motivated than the French group, whereas the French group was more strongly extrinsically motivated than the EFL group.

Additionally, the French group showed statistically stronger feelings of amotivation than the EFL group. Again, the French group did not rate the amotivation items highly in comparison with the other items in the motivation questionnaire, and the overall means for the amotivation scores were low; however, in comparison to the EFL group, their attitudes towards learning the target language are striking. A significantly higher result was found for the item *'I don't see the point in learning <English / French >'* ($p = 0.006$); similarly, the French group were significantly more sensitive to having to study French because there was no other choice. The item *'I take <English / French > because it is compulsory. I will drop it as soon as I can'* was significantly higher for the French group than the EFL group ($p = 0.000$).

These results must be considered in light of the individual differences between the two groups of students involved in the study. The students involved are studying different foreign languages, in different schools, with differing methods and expectations. It is also unclear whether the students in the EFL group were more susceptible to a response acquiescence bias, although their lower scores towards one of the introjected items and both external regulation items would suggest that they were not.

In addition, as the study was primarily interested in the students' use and perceptions of the CALL system, the motivation questionnaire was designed to be succinct while containing items within each of the categories on the self-determination continuum. To draw any firm conclusions, a more in-depth study of students'

motivations would have to be made. However, the results presented above, prior to any use of the CALL application, indicate that there is a difference in general motivating attitudes towards learning the target language between the two participating groups. This difference in motivation towards learning the target language may impact on the participants' attitudes towards using the CALL program.

User response data

Participants' utterances when interacting with the system were recorded and later transcribed for analysis of response type as well as recognition accuracy. The system also logged the recognition results at each stage of the dialogue.

Participants' utterances were categorised into four response types. As the interaction between characters and user is a series of question and answer pairs, the shortest response type that facilitates the conversation is 'answer only'. This is often a one-word answer (e.g. *two*). The second response type used is 'phrase', which constitutes a number of words but does not contain a main verb (e.g. *two tickets to Oxford*). The third response type employed is 'sentence' which contains a main verb (e.g. *I'd like to buy two tickets to Oxford*). The fourth response type is 'verbal non-answer'. This final category constitutes responses where the user has made an utterance (which triggers the recogniser), but does not answer the question. For example, mutterings, thinking aloud, verbal hesitations and non-lexical noises (e.g. coughs) are included in the 'verbal non-answer' category.

Tables 7 and 8 detail the response types for the participants in the evaluation. The response types are given for the individual dialogues within the lesson for both language groups.

As can be seen from these tables, the participants from the French group gave a higher percentage of 'answer only' responses than the EFL group overall, 52.7% for the French group compared to 44.1% for the EFL group. The EFL group gave full

Table 7. User response type – EFL.

Interaction	Participants	Utterances	Answer only (%)	Phrase (%)	Sentence (%)	Verbal non answer (%)
O-O 1	48	341	30.5	34.0	29.6	5.9
O-O 2	48	316	39.6	25.3	32.6	2.5
INT	48	452	57.5	22.3	19.2	0.9
Total		1109	44.1	26.8	26.2	2.9

Table 8. User response type – French.

Interaction	Participants	Utterances	Answer only (%)	Phrase (%)	Sentence (%)	Verbal non answer (%)
O-O 1	28	176	41.5	18.8	32.4	7.4
O-O 2	28	156	53.2	14.1	21.2	11.5
INT	28	353	58.1	26.1	12.7	3.1
Total		685	52.7	21.5	19.7	6.1

sentence responses in 26.2% of the utterances, while the French group attempted full sentence utterances in 19.7% of all utterances.

The French group attempted full sentence responses more frequently in the first one-to-one scenario than in the second one-to-one scenario, whereas the EFL group attempted full sentence responses more frequently in the second one-to-one scenario. Both groups of participants gave more one-word responses in the INT than in the one-to-one scenarios. This pattern of shorter responses in the INT is consistent with that found in a previous experiment (Morton et al., 2008).

As can be seen from Table 8 (French), there was a high percentage of ‘verbal non-answer’ response types for the second one-to-one scenario. Further investigation of this found that the majority of these utterances were the participant speaking to themselves in English, often stating that they did not understand the question. The confusion of some participants in this one-to-one scenario is consistent with the significantly lower attitude results to the interaction-content statements in the attitude questionnaires described earlier. Significant differences were found for the questionnaire items ‘*always understood what the character said*’, ‘*character difficult to understand*’ and ‘*level of the language difficult*’ between the first and second one-to-one scenarios for the French group.

Speech recognition analysis

The accuracy of the speech recognition component is analysed by comparing the transcriptions of user utterances with the output from the recogniser. The utterances are then grouped into in-grammar (IG) and out-of-grammar (OOG) responses. IG responses have been defined in the grammar recognition files (that is, the system developer has predicted and programmed the exact word-for-word response). OOG responses are utterance strings which have not been included in the recognition grammar files.

In-grammar/out-of-grammar rates

Tables 9 and 10 detail the IG and OOG responses for the participants in the evaluation. The results are given for the two language groups separately.

Table 9. In-grammar and out-of-grammar user input – EFL.

Interaction	Utterances	IG (%)	OOG (%)
O-O 1	341	47.2	52.8
O-O 2	316	50.3	49.7
INT	452	72.1	27.9
Total	1109	58.3	41.7

Table 10. In-grammar and out-of-grammar user input – French.

Interaction	Utterances	IG (%)	OOG (%)
O-O 1	176	50.0	50.0
O-O 2	156	40.4	59.6
INT	353	58.6	41.4
Total	685	52.3	47.7

Although overall there was a slight majority of utterances that were IG across all participants, there was a higher percentage of utterances which were IG for the EFL group than the French group. The above tables also show that the participants in both language groups produced more IG responses in the INT than in the one-to-one dialogues across both languages. This is consistent with the findings from a previous evaluation (Morton et al., 2008) and may in part be due to the more transactional nature of the interactive dialogue or because the INT produces shorter utterances from the participants than the one-to-one scenarios.

In-grammar utterances

Further analysis of the recognition accuracy of the system was made. Investigation was made of the system recognition accuracy for the IG utterances. The system recognition output of these utterances was analysed for both word-for-word recognition and semantic value recognition. As the interaction in the dialogues follows a series of question and answer routines, a semantic value is logged for each of the user's utterances. For example, the question from the system 'What time does the train to Oxford leave?' might elicit the answer 'It leaves seven thirty'. In this response, the word-for-word recognition is the string 'it leaves seven thirty'; whereas, the semantic value is 'seven thirty'. If the system recognises this utterance as 'it leaves at seven thirty', the word-for-word recognition is inaccurate, but the semantic value is correct. The semantic value recognition rates include the word-for-word utterances and in addition those utterances where the semantic value of the participant's utterance has accurately been recognised. Tables 11 and 12 detail the recognition accuracy for the IG utterances across the two groups.

Overall, 66.7% of all IG utterances for the EFL group and 50.3% of all IG utterances for the French group had an accurate word-for-word recognition. Reflecting participants' views on the characters understanding them in the experiment, the recognition accuracy of the French group was lower than the recognition accuracy for the EFL group. In contrast to a previous evaluation

Table 11. In-grammar recognition accuracy – EFL.

Interaction	IG utterances	Word for word (%)	Semantic value (%)	Misrec. semantic (%)	Reject (%)
O-O 1	161	62.1	71.4	8.1	20.5
O-O 2	159	67.3	79.2	3.1	17.6
INT	326	68.7	81.0	1.8	17.2
Total	646	66.7	78.2	3.7	18.1

Table 12. In-grammar recognition accuracy – French.

Interaction	IG utterances	Word for word (%)	Semantic value (%)	Misrec. semantic (%)	Reject (%)
O-O 1	88	34.1	72.7	9.1	18.2
O-O 2	63	57.1	74.6	14.3	11.1
INT	207	55.1	61.4	6.3	32.4
Total	358	50.3	66.5	8.4	25.1

(Morton et al., 2008) which suggested that word-for-word accuracy was higher for the group that gave shorter responses, in this evaluation the recognition accuracy was higher for the EFL group who on the whole produced longer utterances. One possible explanation for this is that the commercial recogniser used for the experiment may have more advanced models for their English package than for the other language packages.

There was a slight increase for accurate recognition of the semantic value; overall 78.2% of IG utterances of the EFL group and 66.5% of IG utterances of the French group had accurate recognition of the semantic value. Accurate recognition of the semantic value allows the dialogue to continue effectively between the system and the user; the system has 'understood' the user's response correctly. For example, in the data, to the question '*Where would you like to go?*' one participant answered '*I want go Oxford*'. The system recognised this, incorrectly, as '*I want to go Oxford*'. In this case, the system was able to respond to the participant's answer appropriately (by then asking questions relating to the departure time of the train to Oxford); however, one of the errors in the participant's response would not be accurately recognised. The correct recognition of the semantic value is useful for the facilitation of the dialogue between the characters and the learner; however, it does not always indicate that the learner's errors have been identified. The word-for-word accuracy category includes those utterances where an error had been made by participants that were correctly recognised. However, the accuracy rates for the word-for-word recognition are low, particularly for the French group. This suggests that although such a dialogue system is useful for learners practising the target language, it is not reliable as a tool for accurate identification of learners' errors.

The remainder of the utterances were either misrecognised or erroneously rejected by the recognition component. A higher proportion of IG utterances in both groups were rejected by the system. A total of 18.1% of the EFL IG utterances and 25.1% of the French IG utterances were rejected by the system. The effect of a system reject at any stage in the dialogues is that the character repeats or reformulates the initial question and the user has another opportunity to respond to the question. Although this does not hinder the dialogue between the system and the user, it may contribute to participants' perceptions of the recognition performance of the application.

Investigation was also made on the recasting feedback given to users on their IG utterances in the one-to-one scenarios. The analysis first removed all instances of 'rejections' in the IG data which result in the reformulation design of the dialogue interaction. Of the remaining utterances in the IG data, the analysis investigated correct recasting (where the learner has given a non-full response or a response containing a grammatical error) and incorrect recasting (where the system has not given a recast when one was necessary or where the system has given a recast when one was unnecessary). The results of the analysis are presented in Tables 13 and 14.

In the majority of cases (85.2% for the EFL group and 81.1% for the French group), the participants' responses required a recast and the recast was correctly given. A total of 9.6% of EFL utterances and 3.9% of French utterances correctly did not receive a recast (that is, their utterance was a full grammatical response). However, in 4.4% of EFL utterances and 15% of French utterances, a recast was given when it was not necessary. That is, the participant did not make an error but the system misrecognised the utterance was containing a grammatical error. Finally, in 1.6% of EFL utterances, the system erroneously did not give a recast when one

Table 13. Recasting accuracy – EFL.

	Recast given – correct	Recast not given – correct	Recast given when not necessary	Recast not given when necessary
251 utts	85.2%	9.6%	4.4%	1.6%

Table 14. Recasting accuracy – French.

	Recast given – correct	Recast not given – correct	Recast given when not necessary	Recast not given when necessary
127 utts	81.1%	3.9%	15.0%	0%

was required. As can be seen, the recasting feedback was correctly administered (or not) in 94.6% of EFL utterances and 85% of French utterances. In cases where there was an error made with the recasting facility, it was almost always on the side of caution. That is, the participant received the implicit feedback when not necessary, rather than not receiving when it was necessary.

Whenever a recast was given to a participant's utterance, an error flag would be set for the given error made or that the response was a full sentence. Investigation was also made on the error flagging in the IG utterances for the one-to-one scenarios. A small minority of utterances were full sentence utterances which contained grammatical errors (eight EFL utterances and three French utterances). In each of these cases, the system correctly identified the grammatical error and set the correct error flags. However, it was found that there was a higher percentage of erroneous error flagging where the system detects an error which did not occur in the utterance and therefore sets an unnecessary error flag. This occurred in 12.4% of the EFL group's utterances and 22% of the French group's utterances. For example, a recurring error made by the recogniser was in detecting a missing 'à' (in for example, '*Le train part à neuf heures*') which was in fact present in the learner's utterance. Therefore, as with the recasting facility, the error flag system contains errors in its accuracies, largely in mistakenly detecting errors which are not present.

Out-of-grammar utterances

Investigation was made on the OOG utterances made by the participants. It is expected that OOG utterances should be rejected by the system, as the system is not programmed to 'listen' for these utterances. However although these utterances are OOG, the system may misrecognise some of these utterances for something within the recognition grammars. Analysis of the OOG utterances classifies the results into three categories: correct rejection of the utterance, recognition of the correct semantic value of the utterance or misrecognition of the semantic value of the utterance. For example, an utterance may contain a mid-utterance repetition, which would not be included in the recognition grammars. However, if the system then misrecognises this utterance for an utterance that is in the recognition grammars,

and the value of the recognition is accurate, then the system will proceed appropriately. For example, in the data, one participant answered '*Le train pour Nice part à dix dix heures*'. The system recognised this as '*Le train pour Nice part à dix heures*' thus continuing the dialogue as if the participant's utterance had not contained this disfluency. Note that as with the semantic value recognition in the IG utterances, these correctly recognised semantic value utterances would not necessarily trigger the system to offer feedback to the user in the form of recast. Again, they only indicate those utterances where the dialogue proceeds with the response that the participant intended. Tables 15 and 16 detail the category types for the OOG utterances between the two groups.

The majority of OOG utterances were correctly rejected by the system: 66.5% for the EFL group and 53.8% for the French group. Such utterances include utterances in the participants' first language (thinking aloud), non-lexical responses or hesitation noises as well as responses that are inappropriate to the question asked. A total of 22.5% of OOG utterances in the EFL group and 22.6% of OOG utterances in the French group were recognised with the correct semantic value. These utterances often include short disfluencies in the user's utterance or self-repairs which entailed the utterance was OOG; however, the system recognised the utterance with the intended semantic value. An example of a self-repair in the data which resulted in the recognition of the correct semantic value is the response '*nine fifty in the aft- in the evening*' which was recognised as '*nine fifty in the evening*'.

A total of 11.0% of OOG utterances for the EFL group and 23.5% of OOG utterances for the French group were misrecognised with the wrong semantic value. These are the most problematic as they are likely to cause confusion on the user's part. Analysis of these OOG misrecognitions highlighted that there were some problems (which also occurred in the IG misrecognition results) with similar sounding time responses. For example, in the EFL data there were multiple misrecognitions of '*thirteen*' for '*thirty*', '*fourteen*' for '*forty*' and '*fifteen*' for '*fifty*', and vice versa, and in the French data there were multiple errors with similar sounding time phrases, e.g. '*treize heures*' being misrecognised as '*trois heures*'.

Table 15. Out-of-grammar recognition – EFL.

Interaction	OOG utterances	Correct reject (%)	Recog. semantic (%)	Misrec. semantic (%)
O-O 1	180	67.2	16.7	16.1
O-O 2	157	63.7	26.1	10.2
INT	126	69.0	26.2	4.8
Total	463	66.5	22.5	11.0

Table 16. Out-of-grammar recognition – French.

Interaction	OOG utterances	Correct reject (%)	Recog. semantic (%)	Misrec. semantic (%)
O-O 1	88	58.0	13.6	28.4
O-O 2	93	34.4	29.0	36.6
INT	146	63.7	24.0	12.3
Total	327	53.8	22.6	23.5

Discussion

The results of the evaluation indicate that speech interactive CALL systems are potentially very useful for language learners, despite misrecognitions by the speech recognition component. Analysis of the speech recognition component found an IG word-for-word accuracy of 66.7% for the EFL group and 50.3% for the French group with corresponding semantic value accuracy rates of 78.2% and 66.5%, respectively. Regardless of recogniser inaccuracies, user attitude results indicate a high level of engagement and enjoyment with using the system. This is in accordance with research into the use of automatic speech recognition in CALL applications (Holland et al., 1999), which found that despite the limitations of the speech recogniser and the misrecognitions it generated, end users enjoy the interactions with the system and would prefer a speech interactive component to be included in the CALL application.

The CALL program was able to boost the motivation levels of the participants in the ELF group; however, a similar boost was not found in the French group. It was found that the French group was more motivated than the EFL group by external rewards and the EFL group were more motivated by intrinsic rewards. A comparison of the motivation attributes of the two groups prior to use of the system found a distinct divide between motivation types of the two groups. Given that there was such a divide, it appears that this type of language learning activity was unable to boost the motivation of the group whose motivation is externally regulated to the same degree as it could boost the motivation of more intrinsically motivated learners.

Motivation is an important aspect of success in language learning. Each of the attributes in the extrinsic motivation are thought of as belonging on a sliding scale indicator of success, such that the high scores in attributes of integration and identification are stronger indicators of success than high scores in introjected and external attributes (Deci & Ryan, 1991).

The group whose motivation was intrinsically motivated had an increase in their motivation levels after using the system. As this system is more akin to a real-world experience, those students who already expressed that they were intrinsically motivated expressed that their motivation was boosted by using the system. Those students who were more extrinsically motivated did not appear to increase their motivation to learning the target language after using the system. Such students are motivated by exam or grade results. However, as this system is perhaps different to their current learning practices, they did not view it as beneficial to the same extent.

It has been pointed out (Weiner, 1990) that behavioural theories tended to focus on extrinsic motivation (i.e. rewards) while cognitive theories focus on intrinsic motivation (i.e. goals). Local learning cultures may prioritise one learning theory over the other. The design of the CALL program described in this article is based on cognitive learning theories. Therefore, it may be that learners who are accustomed to teaching methodologies based on cognitive theories were more motivated by this program, whereas learners who are accustomed to teaching methodologies based on behavioural theories were not motivated to the same degree.

The local cultures of the two participant groups in this study may have influenced the students' motivation. The students in the French group may perceive learning French as another school subject rather than a means of communicating with speakers of the target language (in a way that the EFL cohort seemed to). In this

way, the CALL activities were perhaps not viewed as being beneficial or motivation-enhancing for them. There may be also more of a perceived need for the EFL group to learn English, than the French group to learn French. Intrinsic and extrinsic motivation are not mutually exclusive and students can be motivated by different types of motivation. However, the EFL group showed stronger signs of intrinsic motivation than the French group. The reverse was true of the extrinsic motivation attributes. It would be interesting to measure any changes to intrinsic and extrinsic motivation over an extended period of time in order to investigate whether 'getting used to' the CALL program and the methodological approach had any effect on learners' motivations. However, that is beyond the scope of this article.

Note

1. Nuance v8.0.

Notes on contributors

Dr Hazel Morton holds an MA in applied linguistics and a PhD in speech interactive CALL. Her research interests are in the use of new technologies for eLearning and eCommerce applications, in particular the use of speech recognition technology and embodied conversational agents.

Professor Mervyn Jack is the Lloyds Banking Group Professor at the University of Edinburgh. A Fellow of the Royal Society of Edinburgh, Mervyn leads a multi-disciplinary team of researchers investigating usability engineering of eCommerce and eBanking services. With 35 years' experience of research in his field, his long-term research interests are in dialogue engineering and virtual reality systems design for advanced eCommerce and consumer applications. He is author of some 250 scientific papers and five textbooks.

References

- Allum, P. (2001). Principles applicable to the production of CALL-ware: Learning from the field of human computer interaction (HCI). *ReCALL*, 13(2), 146–166.
- Chapelle, C.A. (1998). Multimedia CALL: Lessons to be learned from research on instructed SLA. *Language Learning and Technology*, 2(1), 22–34.
- Deci, E.L., & Ryan, R.M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Plenum Press.
- Deci, E.L., & Ryan, R.M. (1991). A motivational approach to self: Integration in personality. In R. Dienstbier (Ed.), *Nebraska symposium on motivation: Vol. 38. Perspectives on motivation* (pp. 237–288). Lincoln, NE: University of Nebraska Press.
- Ellis, R. (1985). *Understanding second language acquisition*. Oxford: Oxford University Press.
- Garcia-Carbonella, A., Rising, B., Montero, B., & Watts, F. (2001). Simulation/gaming and the acquisition of communicative competence in another language. *Simulation and Gaming*, 32(4), 481–491.
- Gass, S. (1997). *Input, interaction, and the second language learner*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Holland, V.M., Kaplan, J.D., & Sabol, M.A. (1999). Preliminary tests of language learning in a speech-interactive graphics microworld. *CALICO Journal*, 16(3), 339–359.
- Johnson, W.L., Rickel, J.W., & Lester, J.C. (2000). Animated pedagogical agents: Face-to-face interaction in interactive learning environments. *International Journal of Artificial Intelligence in Education*, 11, 47–78.
- Lester, J.C., Converse, S.A., Kahler, S.E., Barlow, S.T., Stone, B.A., & Bhogal, R.S. (1997). *The persona effect: Affective impact of animated pedagogical agents*. In Proceedings of the conference of human factors in computer systems (CHI-97) (pp. 359–366), Atlanta, GA.
- Likert, R. (1932). *A technique for the measurement of attitudes*. New York: Columbia University Press.

- Long, M.H. (1996). The role of the linguistic environment in second language acquisition. In W.C. Ritchie & T.K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). New York, NY: Academic Press.
- Masgoret, A.M., & Gardner, R.C. (2003). Attitudes, motivation, and second language learning: A meta-analysis of studies conducted by Gardner and Associates. *Language Learning*, 53, 167–210.
- Massaro, D.W. (1998). *Perceiving talking faces: From speech perception to a behavioral principle*. Cambridge, MA: MIT Press.
- Morton, H., Davidson, N., & Jack, M.A. (2008). Evaluation of a speech-interactive CALL system. In F. Zhang & B. Barber (Eds.), *Handbook of research on computer-enhanced language acquisition and learning* (pp. 220–240). Hershey, PA: Idea Group Publishing.
- Morton, H., & Jack, M.A. (2005). Scenario-based spoken interaction with virtual agents. *Computer Assisted Language Learning*, 18(3), 171–191.
- Noels, K.A. (2001). New orientations in language learning motivation: Towards a model of intrinsic, extrinsic and integrative orientations. In Z. Dörnyei & R. Schmidt (Eds.), *Motivation and second language acquisition* (pp. 43–68). Honolulu: University of Hawai'i Second Language Teaching and Curriculum Centre.
- Noels, K.A., Clément, R., & Pelletier, L. (1999). Perceptions of teachers' communicative style and students' intrinsic and extrinsic motivation. *Modern Language Journal*, 83(1), 24–34.
- Noels, K.A., Pelletier, L., Clément, R., & Vallerand, R.J. (2000). Why are you learning a second language? Motivational orientations and self-determination theory. *Language Learning*, 50(1), 57–85.
- Pica, T. (1994). Research on negotiation: What does it reveal about second-language learning conditions, processes, and outcomes? *Language Learning*, 44, 493–527.
- Ryan, R.M., & Deci, E.L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1), 68–78.
- Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235–253). Rowley, MA: Newbury House Press.
- Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.), *Principle and practice in applied linguistics: Studies in honour of H. G. Widdowson* (pp. 125–144). Oxford, UK: Oxford University Press.
- Vallerand, R.J. (1997). Toward a hierarchical model of intrinsic and extrinsic motivation. In M.P. Zanna (Ed.), *Advances in experimental social psychology* (pp. 271–360). New York, NY: Academic Press.
- Wachowicz, K., & Scott, B. (1999). Software that listens: It's not a question of whether, it's a question of how. *CALICO Journal*, 16(3), 253–276.
- Weiner, B. (1990). History of motivational research in education. *Journal of Educational Psychology*, 82(4), 616–622.