Web-based mini-games for language learning that support spoken interaction

Helmer Strik [1], *Luigi Palumbo* [1], *Febe de Wet* [2] *and Catia Cucchiarini* [1]

[1] Centre for Language and Speech Technology, Radboud University, Nijmegen, the Netherlands
[2] Human Language Technology Research Group, CSIR, Meraka Institute, Pretoria, South Africa

`w.strik@let.ru.nl`, `luigi.palumbo@student.ru.nl`, `FDWet@csir.co.za`,
`c.cucchiarini@let.ru.nl`

Abstract

The European 'Lifelong Learning Programme' (LLP) project 'Games Online for Basic Language learning' (GOBL) aimed to provide youths and adults wishing to improve their basic language skills access to materials for the development of communicative proficiency in Dutch, French, and English through web-based mini-games. These mini-games were tested in four countries: The Netherlands (Dutch), Belgium (French), United Kingdom and South-Africa (English). Four types of mini-games were developed, and in two of them users can use 'automatic speech recognition' (ASR) to support spoken interaction. In the current paper we will focus on the English versions of these two games that were tested in the United Kingdom and South-Africa. The analyses that are presented in this paper were conducted to determine what users' perceptions are about mini-games with and without speech input and ASR and which aspects of the speech-enhanced games are strongly related to each other.

Index Terms: second language learning, language and speech technology, speaking practice

1. Introduction

The GOBL (Games Online for Basic Language learning) project [1, 2] was set up with the aim of providing youths and adults who wish to improve their basic language skills access to materials for the development and/or improvement of basic foreign-language communicative proficiency through web-based mini-games that support spoken interaction.

Educational mini-games are small and self-contained games which focus on specific well-defined learning topics, which are highly reusable and cost-effective, and which are highly motivating. Mini-games are particularly fit for the development of language skills at the lower end of the proficiency scale (e.g. A2 or B1 level of the CEFR [3]), and allow to focus on aspects which tend to receive little attention in language classrooms nowadays, such as explicit grammar and vocabulary teaching (e.g. [4, 5, 6]). Moreover, there is evidence that disadvantaged language learners seem to profit most from such mini-games [7].

Over the last few years, games have been used in a number of programmes to motivate and emancipate disadvantaged citizens, such as people with health problems [8, 9], deaf people [10], people with dyslexia [11], and young boys 'at risk' who drop out of formal education programmes [12]. Clearly, the motivational aspects of mini-games are being used to address a wide range of socially relevant issues. For the teaching of foreign languages, a number of fully immersive games exist (e.g. [13, 14]), but these often require expensive hardware and target more advanced language learners.

The main objective of the GOBL project was to apply the motivational elements of mini-games to the teaching of foreign language grammar, vocabulary and basic communicative skills in order to cater for the needs of low-skilled language learners in secondary schools and adult education. Learning materials have been developed for Dutch, English and French. Additionally, speech recognition technology has been implemented for speaking activities in Dutch and English.

In the ASR versions of the games students practice lexical and syntactic skills in the spoken modality. It might be argued that ASR has no added value in this case because these skills could just as well be practiced without resorting to ASR, i.e. in the written modality or through drag and drop. Although a certain amount of transfer can take place from one modality to another, this usually applies to declarative knowledge, while for procedural knowledge skill-specific practice is required [18]. In other words, although one could train lexical and syntactic skills in the written modality, then not all knowledge (esp. procedural knowledge) will transfer to the spoken modality, and thus the acquisition of vocabulary and syntax in the spoken modality benefits from practice in the same modality.

In [15, 16] we reported on the initial stages of the GOBL project, explaining the background, the aims, the needs analysis, and the development of the mini-games. In this paper we report on the subsequent stages that concern the use and evaluation of the mini-games.

2. The GOBL project

In the GOBL project three evaluation stages were envisaged:
1. an initial needs analysis;
2. a mid-term evaluation;
3. a final evaluation.

In each stage target users (A2 or B1 learners) and teachers were involved. In stage 1, the initial needs analysis, mock-ups of the games were used, while in stages 2 and 3 the users played with the first and second versions of the games, respectively. During the evaluations of stages 2 and 3, we used questionnaires and focus group discussions to capture user feedback.

The results of the needs analysis yielded ideas for the design of the mini-games [15, 16]. Mid-term evaluations with language learners were conducted in May-June 2013 in the Netherlands, Belgium, the United Kingdom (UK), and South-Africa, and results were presented at the SLaTE 2013 workshop.

The mid-term evaluation revealed that some games were experienced as being either too fast or too slow, that there were some difficulties in understanding how to play the games, that sometimes the goals of the games were not clear,

Figure 1: *Example of a Fingerprints (FP) exercise. The user could select one of the shown answers (fingerprints). The wrong answer was chosen. The top bar provides (left to right) a) a brief explanation of what to do; b) the score multiplier (assigning extra-points after three consecutive right answers); c) the current score, and d) a pie chart-like indicator showing how much time is left (in green).*

Figure 2: *Example of the overview provided at the end of a Fingerprints (FP) game. Larger-sized words indicate the answers chosen. A green tick mark after a word denotes a right choice, a red cross highlights a wrong one. The user's final score is shown in the top-right corner.*

and that some games were perceived more as exercises than as games. Students also complained about poor graphics and the lack of immediate feedback.

The games attracted positive comments when they were easy to play, easy to understand, when sentences were short, when the games were not too fast and when they were suitable for learning as well as having fun. Little support was found for comparing scores with other learners, as students were mainly interested in comparing their own current score with their previous scores.

Valuable recommendations were to include more topics relevant to the needs of the learners, to improve the introduction to the games, and to have the language of instructions on the screen match the target language or the L1 of the learners.

The results and the feedback from the mid-term (stage 2) evaluation were taken into account, and used to develop a second, improved version of the mini-games, which was then evaluated in 2014. The results presented here are based on analyses conducted on the quantitative information drawn from the questionnaires that were part of the final (stage 3) evaluation round.

3.   Material and method

Within GOBL there are two modes: story mode and individual games. In the first mode, a detective story, the games are presented in a fixed order. At certain places in the story, the learner has to play games, and after playing the games the story continues. In the 'individual games' mode the learner can choose which games to play. Four types of mini-game were developed within the GOBL project: Fingerprints (FP), Roof-surfing parrot (RP), Lie detector, and Line-up. We are presently reporting results pertaining to the first two games (FP & RP), for which users could choose to play either with or without ASR.
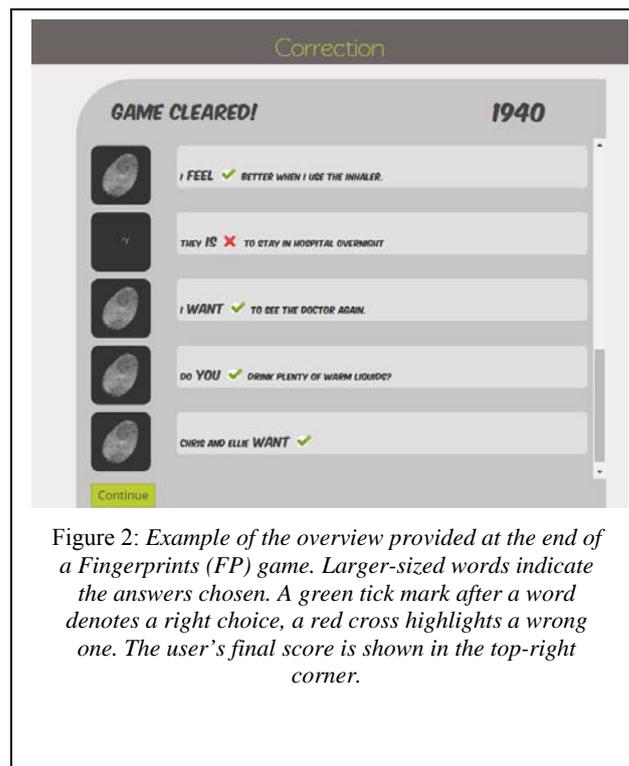
3.1. Fingerprints

In the FP game, users are provided with a series of scenarios where several blank-spaced sentences ─ i.e., incomplete sentences concerning a given topic, such as diseases ─ need to be filled in with one of the word alternatives offered in the form of fingerprints (see Figure 1). As a hint, each sentence is accompanied by a relevant picture aimed at adding visual context users could possibly benefit from in order to infer the right answer. According to whether the ASR is turned on or off, users need to either utter their chosen answer, (ASR on), or manually select it by clicking on the fingerprint (ASR off).

FP's 'playing' goal is for users to collect as many fingerprints as possible within the game's time limit. As for its 'language learning' goal, the mini-game is intended for learning new lexical items and, when the ASR is deployed, for stimulating oral production of the new words.

Users get immediate feedback on their answers. If the chosen word/fingerprint is right, they are shown explicit, positive feedback in the form of a green tick immediately followed by a 'winning' ring and they gain points. Conversely, if their answer is wrong, they get explicit, negative feedback in the form of a red cross followed by a 'losing' ring and points are deducted. The time-out feedback informs the student that time is up and he/she should proceed to the next item. At the end of the mini-game, an overview is presented of all items with the corresponding responses (Figure 2).

3.2. Roof-surfing Parrot

In the RP game, users are shown a blue parrot that can jump from the top of one skyscraper to the next one (see Figure 3). A dark cloud approaches the parrot, and the learner should try to move quickly to make sure that the parrot stays ahead of the cloud.

Figure 3: *Example of a Roof-surfing parrot (RP) exercise. The user chose the right answer and the parrot jumped from the 'question' skyscraper to the skyscraper with the right answer, thus distancing himself from the approaching dark cloud. The top bar provides (left to right) a) a brief explanation of what to do; b) the score multiplier (assigning extra-points after three consecutive right answers); c) currently gained points, and d) a pie chart-like indicator showing how much time is left (in green) before the dark cloud reaches the parrot.*
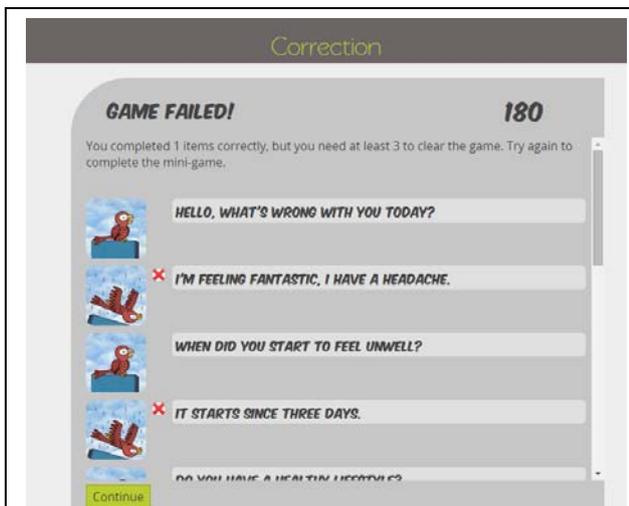

Figure 4: *Example of the overview provided at the end of a Roof-surfing parrot (RP) game. Each question provided by the system is followed by the answer chosen by the user. The red crosses indicate a wrong choice. The user's final score is shown in the top-right corner.*

First there is one skyscraper with a question. Immediately to the right are 3 or 4 skyscrapers with different answers. The learner should direct the parrot to the correct one. Explicit feedback is provided immediately after an answer is given, which is positive if the parrot was directed to the right response, or negative if not. In the latter case, the parrot automatically jumps to the correct answer. Then the parrot goes to the skyscraper with the next question, etc.

Unlike the isolated prompts in the FP mini-game, here the game comprises a dialogue on a given topic, e.g., dealing with a phone call to book a visit with a medical specialist. From a language learning point of view, users are required to choose the grammatically right answer. As above, the playing dynamic changes according to whether the ASR is turned on or off, in that users need to either utter the whole sentence or manually select it.

At the end of the RP mini-game, the learner is also presented with an overview of all the items in the mini-game, together with their responses. An indication is given of whether the responses were right or wrong (see Figure 4).

3.3. Participants

Evaluations of the English version of the GOBL mini-games took place in the UK and South-Africa. The demographic details of these two sub-groups are as follows:

a) the first sub-group was formed by GOBL users tested at Nottingham's Central College and Newcastle's Westgate Community College, UK (N = 47, very diverse nationalities, age range: mostly in their 20s or 30s and two of them in their 50s, about 2/3 women);

b) the second sub-group was formed by GOBL users tested at Stellenbosch University's Language Centre, South-Africa (N = 11, mostly from Korea and Mozambique, all in their 30s, only one woman).

In the following sections we sometimes present results for the two sub-groups separately, and sometimes for the combined groups, i.e. all 58 participants that tested the English version. In general, if we do not explicitly mention that it concerns a sub-group (UK or SA), then about the results pertain to the whole group (UK + SA).

3.4. Experimental Procedure

The students played the games in two sessions, one in 'free play' and the other in 'story mode'. At the beginning of the sessions they received instruction form the session leaders. A PowerPoint introduction was first shown and the students were told that they would be asked to play four games and that through the games they would receive practice in vocabulary, grammar and phrases that could be useful in real life situations such as going to the doctors, using public transport and job interviews. It was explained to the students that they would have to choose correct answers as quickly as possible to collect fingerprints and find the truth and save the parrot, otherwise they would not be able to solve the mystery.

During each session the learners completed questionnaires at the beginning of the session, after playing each mini-game, and at the end of the session. In addition, there were also focus group discussions at the end of each session.

3.5. Questionnaires

Participants could answer the questions on a seven-point scale, with the extreme values being 1 = 'not at all' and 7 = 'very much'. Here we present results related to the following sets of questions.

a) General issues concerning the overall experience with the mini-games:
Q1. Was it clear how to play the game?
Q2. Was the game easy to play?
Q3. Did you like the game?
Q4. Did you learn some English from the game?
Q5. How good do you think you were at the game?

b) Questions focussing more on issues related to speaking:

Q6. Was it clear how to play the game with speaking?
Q7. Was the game easy to play with speaking?
Q8. Did you like being able to practise speaking with the game?
Q9. Did you learn some English from the game?
Q10. How good do you think you were at the game?
Q11. Did you find speaking your answers useful?
Q12. Did you prefer the speaking game to the version of the game where you don't speak?
Q13. Did the game understand everything you said?

Questions Q1 – Q5 and Q6 – Q10 are related, they are similar questions for the ASR on and ASR off versions of the same games. In the following section the answers to these related questions will be analysed. Note that they concern the opinions of the learners captured in the questionnaires which reflect their perceptions of different aspects of the mini-games, e.g. the perceived clarity (Q1 and Q6) and the perceived quality of the ASR (Q13).

4. Analyses

A number of statistical analyses were carried out on the participants' quantitative answers using IBM® SPSS®. Here we present results of t-tests and correlation analyses. We are currently carrying out additional statistical analyses. If we obtain more interesting results, we will present them at the workshop and on our websites [1] [2].

As for the probability, i.e., p-values, of getting some given results if the null hypotheses were true, our thresholds for accepting the alternative hypotheses were $p < .05$ for a statistically significant result and $p < .01$ or even $< .001$ for more significant ones. Any $p > .05$ indicated non-significant results.

We performed a series of bivariate correlational analyses according to the different aspects that were investigated by means of questionnaires. We especially focussed on the results relevant to when the ASR was turned on with regards to the whole English group. We used Pearson's product-moment correlation coefficient, r, as a measure of the strength of the relationships between the considered variables. As a quantifier of the experimental effect size, the correlation coefficient accounted for a) 1% of the total variance at the value of .10, i.e., a small-sized effect; b) 9% of the total variance at the value of .30, i.e., a medium-sized effect; and c) 25% of the total variance at the value of .50, i.e., a large-sized effect [17]. Additionally, we report the correlation coefficient's 95% Confidence Intervals (CI).

It is worth noticing that, whenever we attempted to infer any conclusions from our results, we always bore two principles in mind: a) the *tertium quid* or 'third-variable' problem, i.e., taking into account the presence of a third measured or unmeasured variable that potentially affected the relationships between the ones being presently under observation; and b) the 'direction of causality' problem, i.e., the fact that we could not determine — at least in statistical terms — which of the two variables caused the other to change. Rather, we tried to deduce the most plausible — at least from a logical point of view — conclusion on the basis of the results we obtained.

A first series of dependent/paired-samples t-tests were conducted within the same sub-group gathering together participants from the same countries. At the same time, a second series of independent-samples t-tests were carried out between the two sub-groups together, i.e. UK vs. South-Africa. Within those two categories of t-tests, the mean differences (*M*) and their corresponding standard error means (*SE*) (arising from the aforementioned sub-groups being each time considered either dependently or independently) were tested according to whether the two ASR-supported mini-games, namely Fingerprints (FP) and Roof-surfing Parrot (RP), were played with the speech recognition facility being turned on or off. In this case, Pearson's correlation coefficient, r, was manually computed as an indicator of the effect size of the t-tests [17]. We used the same quantifiers of the experimental effect size for these analyses.

5. Results

Before presenting the results below, we would like to emphasise that, in the present paper, we focus on results related to speaking and ASR. The data concern the answers to Q6 – Q13 of the questionnaires and reflect the learners' opinions. Therefore the scores on variables such as enjoyment, quality of the ASR, usefulness, and amount of English learned, are indications of the way these game elements were perceived by the learners.

5.1. Users' perceptions of the games

The results in this section give an indication of how users' perceptions of different aspects of games with and without speech input and ASR differ.

5.1.1. Clarity on how to play (Q1 – Q6)

On average, only users from the UK practising English with the FP mini-game judged its 'ASR on' version less clear to be played (*M* = 4.91, *SE* = .36) than the 'ASR off' mode (*M* = 5.71, *SE* = .32). This difference of .80, 95% CI [.25, 1.33], is significant $t(23) = 2.80$, $p = .01$, $r = .50$. In all the other cases no significant differences were observed regarding clarity.

5.1.2. Difficulty of the mini-games (Q2 – Q7)

On average, users practising English with the FP mini-game judged its 'ASR on' mode to be harder to play (*M* = 4.88, *SE* = .35) than the 'ASR off' one (*M* = 5.23, *SE* = .35). However, this difference, .35, 95% CI [-.34, 1.00] was not significant $t(25) = 1.03$, $p = .314$.

We subsequently hypothesised that the same applied to the RP mini-game, i.e., its ASR mode-on version would have been considered the hardest one. Our hypothesis was confirmed as the mini-game played with ASR was harder to play (*M* = 4.73, *SE* = .33) than without ASR (*M* = 5.38, *SE* = .29). In this case the difference, .65, 95% CI [.04, 1.30], was significant, $t(25) = 1.94$, $p = .032$, $r = .36$. In all other tested cases concerning this variable no significant differences were observed.

5.1.3. Self-perceived skill with the mini-games (Q5 - Q10)

On average, users practising English with the FP mini-game judged themselves to be worse at playing with its 'ASR on' version (*M* = 4.57, *SE* = .28) than with the ASR turned off (*M* = 4.88, *SE* = .27). However, this difference, .31, 95% CI [-.15, .77], was not significant $t(25) = 1.22$, $p = .117$.

The same result was observed for RP, for which, on average, users felt they were worse at playing with ASR (*M* = 4.65, *SE* = .27) than without (*M* = 5.00, *SE* = .26). Again, this difference, .35, 95% CI [-.15, .92], was not significant $t(25) = 1.30$, $p = .102$. We also examined the results of the sub-groups UK & SA for this variable, but no significant differences were observed.

5.1.4. Appreciation of the mini-games (Q3 - Q8)

On average, users from the UK practising English with the FP mini-game enjoyed its 'ASR on' version less ($M = 5.10$, $SE = .30$) than the 'ASR off' version ($M = 5.70$, $SE = .32$). This difference, .60, 95% CI [.15, 1.05], was significant $t(19) = 2.45$, $p = .024$, $r = .49$.

Similarly, on average, users from the UK practising English with the RP mini-game liked the 'ASR on' version less ($M = 5.05$, $SE = .28$) than the version without ASR ($M = 5.65$, $SE = .31$). This difference, .60, 95% CI [.15, 1.05], was significant $t(19) = 2.56$, $p = .019$, $r = .50$.

Between the two sub-groups, on average, users from the UK practising English with the 'ASR on' version of RP liked being able to practise spoken language more ($M = 5.00$, $SE = .27$) than users from South-Africa ($M = 3.66$, $SE = .71$). This difference, 1.34, 95% CI [-.04, 3.00], was significant $t(25) = 2.09$, $p = .047$, $r = .38$. All of the other tested cases concerning this variable showed non-significant differences.

5.1.5. Amount of English learned (Q4 - Q9)

On average, users from the UK practising English with the FP mini-game felt to have learned less English with its 'ASR on' version ($M = 4.90$, $SE = .26$) than without ASR ($M = 5.52$, $SE = .30$). This difference, .62, 95% CI [.09, 1.14], was significant $t(20) = 2.28$, $p = .034$, $r = .45$.

As for the differences concerning the ASR version between the two countries, we found that, on average, users from the UK practising English with the ASR version of FP felt to have learned much more English ($M = 5.04$, $SE = .24$) than users from South-Africa ($M = 2.86$, $SE = .40$). This difference, 2.18, 95% CI [1.32, 3.06], was very significant $t(29) = 4.34$, $p < .001$, $r = .62$. All of the other tested cases concerning this variable showed non-significant differences.

5.2. Relationships between aspects of the games

The results presented in this section give an indication of the relationships between different aspects of the speech-enabled version of the games (ASR on). The relationships are quantified in terms of correlation coefficients.

5.2.1. Clarity on how to play the game

For clarity on how to play the games (Q6), the following correlations with other variables were observed.
Q7 – it is easy to play the games:
 $r = .58$, 95% CI [.382, .768], $p < .001$;
Q8 – level of enjoyment:
 $r = .55$, 95% CI [.359, .731], $p < .001$;
Q9 – amount of English learned:
 $r = .27$, 95% CI [.033, .467], $p < .05$;
Q10 – perceived skill of playing the game:
 $r = .59$, 95% CI [.439, .737], $p < .001$.

These values indicate that, if it is clear how to play the games, users find it easier to play the games, enjoy them more, think they learn more and are more skilful in playing the games.

5.2.2. Amount of English learned

For perceived amount of English learned when the ASR was deployed (Q9) we found the following correlations with other aspects of the mini-games.
Q8 – degree to which users liked to practise spoken English:
 $r = .69$, 95% CI [.411, .856], $p < .001$;
Q11 – usefulness of speaking practice:
 $r = .66$, 95% CI [.378, .831], $p < .001$;

Q12 – learners' preference for the ASR mode-on:
 $r = .55$, 95% CI [.285, .775], $p < .001$;
Q13 – quality of ASR:
 $r = .46$, 95% CI [.092, .739], $p = .001$.

The results seem to suggest that, the more users think they learn with the game, the more they experience it to be useful, want to use it, are positive about the quality of the ASR, and want ASR enabled.

5.2.3. Quality of ASR

Answers to Q13 "Did the game understand everything you said?", are an indication of the perceived quality of the ASR. We observe a number of significant, positive correlations between perceived quality of the ASR (Q13) and other variables, which are listed below.
Q8 – how much users liked to practise speaking:
 $r = .68$, 95% CI [.431, .871], $p < .001$;
Q11 – usefulness of speaking practice:
 $r = .74$, 95% CI [.522, .863], $p < .001$;
Q12 – the extent to which they preferred ASR to be on:
 $r = .59$, 95% CI [.204, .857], $p = .001$.

Thus, if the perceived quality of ASR is good, learners are also positive about other aspects of the mini-games (see above), such as the amount they have learned, the usefulness of speaking practice, and whether they like to use ASR or not.

6. Discussion and conclusions

In our t-test analysis we have made comparisons between 'ASR on' and 'ASR off' versions of two mini-games for five pairs of questions (see section 5.1). The comparisons were made for all data together (UK and SA, for the two mini-games, i.e. everything collapsed), and for the two sub-groups separately. Significant differences are presented in section 5.1. It should be noted that in the majority of the cases that were analysed, no significant differences were observed, so the results mainly reveal trends rather than fixed patterns.

In general, we observe a tendency for more positive results for the 'ASR off' versions of the games compared to the 'ASR on' versions, with some significant differences, especially for the FP game in the UK. For FP - UK we also found significant differences for clarity (Q1-Q6), appreciation (Q3-Q8), and amount of English learned (Q4-Q9).

Similarly, we observe a tendency for the UK users to have more positive feelings towards (aspects of) the mini-games in comparison to the South-African users. Here we find only two significant differences for specific games: for the RP game on appreciation (Q3-Q8), and for the FP game on amount of English learned (Q4-Q9).

If we then look at the correlations, we observe many significant, positive correlations between the different aspects of the games that were investigated. The correlations are positive because of the way in which the questions 6 to 13 were formulated, i.e. a higher score on one question goes hand in hand with higher scores on other questions. The fact that many of these correlations are significant indicates that this covariation in the data is strong, and that learners have similar opinions about the relations between these aspects, e.g. that clarity how to play the games and quality of ASR are very important for how learners enjoy the games and perceive them to be useful.

Combining these two types of findings (from t-tests and correlations) seems to suggest that in the present experiments the way in which the quality of the ASR was perceived by the learners had a considerable impact on the results of the evaluations, which were not very positive with respect to the

'ASR on' mode. This is something that deserves further attention.

The advantages of multi-media language learning environments are indisputable. Stimulating users with appropriate visual and audio cues may be challenging from a design point of view, but fairly simple to accomplish technically. However, creating environments in which users can also generate audio responses has proven to be an enormous technical challenge. Even in unresponsive settings where audio interaction is limited to 'record and playback', users often struggle with hardware issues like microphone settings, playback volume, etc.

The challenge becomes even more daunting in responsive systems where ASR is used to provide some form of feedback on speech produced by users. In addition to the hardware issues mentioned before, ASR systems are sensitive to changes in acoustic channels (e.g. different microphones) and environments (e.g. classrooms) and it is not always possible to maintain recognition performance at an acceptable level under different conditions.

In a project like GOBL where the mini-games were deployed in many different language schools all over the world, it is impossible to anticipate all the conditions in which the ASR would have to function. In addition to the technology itself, small practical things like internet access, browser versions, audio settings and noisy classrooms also had an impact on how the games functioned and, as a consequence, were experienced.

For example, the data that was captured during the ASR-enabled GOBL games was manually annotated and the transcribers were instructed to mark events where speaker noise, background noise or other acoustic events occurred instead of speech. It was found that more than 30% of the files that were sent to the ASR for processing did not contain any speech at all, but music or noise. The types of noise observed included speech from other users, the background music of the mini-games, speaker-generated noises like lip smacks and filled pauses and noise generated by incorrectly connected microphones.

The majority of the files that did contain speech also contained some noise, like speaker noises, the background music of the games or speech produced by other students playing speech-enabled games at the same time. An analysis of the ASR results showed that the presence of substantial background and speaker noise often resulted in incorrect recognition. The shorter utterances produced during the Fingerprints game were also more difficult to recognise correctly than the longer utterances associated with the Roof-surfing Parrot game. This was anticipated, and therefore we instructed users to start with the longer RP utterances, during which the ASR adapted to the voice of the user; and we also advised them to wear a head-set. However, users not always followed these, and other, instructions.

Creating interactive, ASR-enhanced language learning environments therefore requires further development to ensure predictable and stable operating conditions as well as more robust ASR technology. At present it can be argued that much of what users perceive as interaction with ASR systems actually is interaction with other, more practical issues and quite a bit of what ASR systems are required to process is not speech at all.

7. Acknowledgements

8. References

[1]   http://hstrik.ruhosting.nl/gobl/

[2]   http://www.gobl-project.eu/

[3]   Council of Europe, *A Common European Framework of Reference for Languages: Learning, Teaching, Assessment.* Cambridge: Cambridge University Press, 2001.

[4]   http://www.digitaldialects.com/

[5]   http://learnenglish.britishcouncil.org/en/games

[6]   http://beta.visl.sdu.dk/games_gym.html

[7]   M. E. Herselman, "South African resource-deprived learners benefit from CALL through the medium of computer games," *Computer-Assisted Language Learning*, vol. 12, no.3, pp. 197–218, 1999.

[8]   P. M. Kato, "Video games in health care: Closing the gap," *Review of General Psychology*, vol. 14, no. 2, pp. 113–121, 2010.

[9]   L. Gamberini, G. Barresi, A. Majer, and F. Scarpetta, "A game a day keeps the doctor away: A short review of computer games in mental healthcare," *Journal of CyberTherapy and Rehabilitation*, vol. 1, no. 2, pp. 127–145, 2008.

[10] T. M. G. Saliés, and P. Starosky, "How a deaf boy gamed his way to second-language acquisition: Tales of intersubjectivity," *Simulation & Gaming*, vol. 39, no. 2, 209–239, 2008.

[11] I. Smythe, *Dyslexia in the Digital Age. Making IT Work.* London: Continuum, 2010.

[12] C. Steinkuehler, and E. King, "Digital literacies for the disengaged: Creating after school contexts to support boys' game-based literacy skills," *On the Horizon*, vol. 17, no. 1, 47–59, 2009.

[13] http://www.alelo.com/tactical_language.html

[14] http://www.lost-in.info

[15] F. Cornillie, A. Grenfell, S. Windeatt, P. Desmet, "Challenges in specifying and evaluating a conceptual design for a task-based mini-game environment for language learning," in *EuroCALL 2012 – CALL: Using, Learning, Knowing, August 22–25, Gothenburg, Sweden*, 2012.

[16] H. Strik, P. Drozdova, C. Cucchiarini, "GOBL: Games Online for Basic Language Learning," in *SLaTE 2013 – Workshop, August 30–31 – September 1, Grenoble, France, Proceedings*, 2013, pp. 48–53.

[17] A. P. Field, *Discovering Statistics Using IBM SPSS Statistics*. London: SAGE, 2013.

[18] R. M., DeKeyser, Practice in a second language: perspectives from applied linguistics and cognitive psychology. New York: Cambridge University Press, 2007.