

Research Article

User Experiences from L2 Children Using a Speech Learning Application: Implications for Developing Speech Training Applications for Children

Maria Uther ^{1,2}, **Anna-Riikka Smolander**³, **Katja Junntila**^{3,4}, **Mikko Kurimo**⁴, **Reima Karhila**⁴, **Seppo Enarvi**⁴ and **Sari Ylinen**^{3,5}

¹Department of Psychology, University of Winchester, Sparkford Rd., Winchester SO22 4NR, UK

²Department of Psychology, University of Wolverhampton, Faculty of Education, Health and Wellbeing, MC Building (Room MC305), University of Wolverhampton, Wolverhampton WV1 1LY, UK

³Cognitive Brain Research Unit, Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki, P.O. Box 9, 00014 Helsinki, Finland

⁴Department of Signal Processing and Acoustics, School of Electrical Engineering, Aalto University, P.O. Box 13000, 00076 AALTO Espoo, Finland

⁵Cicero Learning, Faculty of Educational Sciences, P.O. Box 9, 00014 Helsinki, Finland

Correspondence should be addressed to Maria Uther; m.uther@wlv.ac.uk

Received 31 May 2018; Accepted 1 October 2018; Published 1 November 2018

Guest Editor: Rafal Rzepka

Copyright © 2018 Maria Uther et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We investigated user experiences from 117 Finnish children aged between 8 and 12 years in a trial of an English language learning programme that used automatic speech recognition (ASR). We used measures that encompassed both affective reactions and questions tapping into the children's sense of pedagogical utility. We also tested their perception of sound quality and compared reactions of game and nongame-based versions of the application. Results showed that children expressed higher affective ratings for the game compared to nongame version of the application. Children also expressed a preference to play with a friend compared to playing alone or playing within a group. They found that assessment of their speech is useful although they did not necessarily enjoy hearing their own voices. The results are discussed in terms of the implications for user interface (UI) design in speech learning applications for children.

1. Introduction

This study investigated child user experiences from a game-based language learning application that used automatic speech recognition (ASR) technology. The application, called 'Say it again, kid!' (SIAK) [1, 2], is designed to assist foreign language learning (vocabulary and production) in children. The technology behind the application is designed for both computers and tablets and uses ASR components for the assessment of children's speech produced while learning new words in a new, nonnative language. Thus far, the use of ASR engines in language learning has primarily been aimed at assisting language learning in native context (e.g., reading tutors such as listen [3], tball [4], space [5], and flora [6]).

In itself, the use of automatic speech recognition engines in children is challenging [7, 8], but the use of ASR to aid foreign language learning is a venture that poses even further challenges and is still a field that is underdeveloped [7–11].

Nonetheless, the SIAK project sought to address this research gap by developing and implementing a foreign language learning application. The application uses a Hidden Markov Model (HMM) segmenter to find phoneme boundaries in players' utterances. Phoneme segments are individually evaluated by a Recurrent Neural Network (RNN) bilingual phoneme classifier. The classifier results are then fed into a Support Vector Machine (SVM) scoring regressor which outputs a 0-100 score, which is further mapped into a rejection or a 1-5 star score. The latency for scoring an

utterance is 2-3 seconds. The scoring mechanism is trained with in-game second-language (L2) utterances that were collected in previous experiments and scored by a human expert. The game is designed to produce a score computed using speech recognition technology after each utterance. We designed the trial of the game such that we compared performance and user reactions to a nongamified pronunciation learning environment. We used gamification since it has been suggested that this lends itself to more situated learning and a more immersive, engaging experience, which can be helpful within the language learning context (see [12–14] for a review).

The goal of the SIAK project was broadly to develop and test a novel, automatic speech recognition application to assist children in learning a new language. We have already published details of the algorithms and implementation itself [1]. In this paper, the research questions were as follows:

- (i) What are the child's affective reactions to the SIAK?
- (ii) What are their perceptions of pedagogical utility of gamified applications?
- (iii) Where there any user aspects relating to the use of audio and interaction with speech in such applications that were problematic? This final point is crucial since the implementation of spoken language learning applications necessarily requires audio input and output.

Answering these research questions with a robust data set on user experience from actual child language learners is clearly needed in this underresearched and underdeveloped area.

As we were measuring child user experience in our application, we followed the framework for steps in usability testing according to Markopoulous & Bekker [15], namely, (1) develop assessment criteria (goals); (2) develop usability testing measurements (contexts, tasks); and (3) consider child characteristics that may constrain the design of the measurements/tasks (e.g., knowledge, age, and language ability). To this end, we focused on three areas in terms of assessment: (1) affective reactions and user engagement, (2) the perceived pedagogical value, and (3) audio interaction issues. As SIAK was an application which necessarily involved audio, we were interested in the users' perceptions of sound quality, which may then in turn affect their learning experience [16]. We were also interested in the use of speech production scoring. It should be emphasized that our goal here was not to present the actual learning outcomes (these will be presented in due course separately), but rather to report on the user experience feedback that would in turn inform the design of any future iterations of this application (and indeed speech training applications for children in general).

To measure the areas we wished to assess, we included questions regarding basic affective reactions (e.g., 'did you like this game?') as well as questions related to the elements which were pertinent to using speech-based applications (e.g., 'how clear was the sound?' and 'did you like hearing your own pronunciation during the game?'). For these questions, we used the smiley-o-meter method from the 'fun toolkit' techniques (see [17] for a review) with reference to

Likert-scale answers to agree to the statement with 5 points ('not at all' to 'very much' at the extremes). There were also items from an 'again again' table (also from the 'fun toolkit'), where the player was asked questions comparing game and nongame versions of the application. The 'again again' method was chosen for comparison of the game and nongame version as it allowed freedom for the children to express a preference (or dislike) for both versions simultaneously. The 'again again' items included questions on (a) Would you like this game for yourself (yes/no or maybe)? or (b) Would your teacher like this (yes/no/maybe)? The latter question (on the teacher's presumed preferences) was selected as it has been shown to tap into the child's sense of pedagogical utility more easily than a direct question of education value. Within other studies [18], it has been shown that the perception of whether a teacher chooses an application is related to the child's perception of how good it was for learning.

In line with our research questions, we hypothesized that

- (a) The user experience (in particular, affective reactions) to our new SIAK software in the target audience of 8 to 12 year olds would be positive and especially so for the game-based version of the software.
- (b) Although gamification of the application would presumably add positively to the user experience, the perceived pedagogical value may be rated less positively. In addition, there may be further desire for collaborative learning as games are often enjoyed as social endeavours.
- (c) Participants rating of sound quality may be affected by application or device type [16] and although not studied explicitly before to our knowledge, we also wished to explore the children's reaction to their voice being scored by computers, which may cause self-consciousness for example.

2. Materials and Methods

2.1. Participants. We recruited 117 children (59 females and 58 males) aged between 8 and 12 years (mean=9.5; SD=1.2) from the Helsinki area. The children were recruited via local schools in Helsinki, with the consent of schools to recruit and consent of the children and parents to participate. There were also some children that were recruited directly via social and community networks.

2.2. Materials. The speech learning application (SIAK) is implemented as a computer board game that runs on an Android tablet or a Windows PC (laptop) and a headset [1]. Following the testing period, the children also were given a questionnaire that modeled itself on the 'fun toolkit'—namely using the 'smiley-o-meter' scale and the 'Again again table' [17]—where they were asked to respond to a question such as 'Would you like to use this game again' responding with either 'yes', 'no', or 'maybe'. The again again items were especially used to compare game and nongame versions. There were 15 questions in total (7 smiley-o-meter items, 7 again again items and one qualitative open-ended question asking whether they

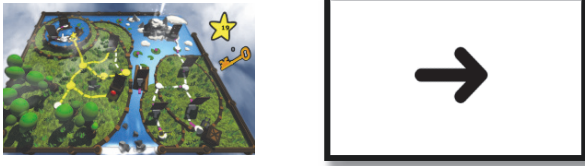


FIGURE 1: Screen shot of game and nongame versions of SIAK. Left panel shows example game ‘world’ that the learners explore. Right panel shows a nongame simple interface.

would like any particular item more (e.g., certain characters, sounds, etc.). All questionnaire items were given to the children in Finnish, which was their native language (any reference to questions in this paper are a translation into English of the original Finnish).

2.3. Procedure. The children were given the SIAK application which functioned to improve their pronunciation and broaden their vocabulary in English by introducing new English single words. As the children progressed in the game, they then encountered sentences which contained the new words. Children heard the word in Finnish and in English (produced by different native English speakers) and saw a related picture. The child was required to repeat the pronunciation of the word aloud. The children then received feedback on their pronunciation as a numerical score. The child’s own and native English speaker’s utterances were played again for comparison, and they received a one to five star rating based on the utterance score.

While testing, there were elements of the program that were not implemented as a game (instead of an immersive experience, they were shown a simple white background, forced order of stimulus presentation, no feedback) although the stimulation and the speech production task were the same as the game version which is described in [19]. Figure 1 shows the comparison of the screen between game and nongame version. A video of child participants playing SIAK is also available at: <https://www.youtube.com/watch?v=-cgyJFV8-58&feature=youtu.be>

All children evaluated the game and asked to compare game and nongame versions using the ‘again again’ question items (note that for 21 participants in the light user group, they did not evaluate the nongame version). The sample was divided into two groups: light users and experienced users. This was done as because we wanted a large sample of game players to give feedback on the application for user acceptance testing (UAT) and user experience (UX) reasons. For UAT/UX testing, we did not need the participants to test for many weeks at a time, but a minimum of one week – hence a ‘light usage’ group. On the other hand, to judge *efficacy* of the intervention, we also tested a set of ‘experienced’ users who have been using the application for at least 4 weeks. For this latter sample, we tested educational outcomes and brain measures as a result of training (those data will be reported separately), as training effects are typically seen over a minimum 1 month period. But it was not necessary for all

TABLE 1: Affective reactions from the SIAK application using the smiley-o-meter rating scale (NB: some items had lower N due to participants not filling in every question).

Question item	N	Range of scores	Mean rating	SD
How much do you like the application? (1=not at all; 5=very much)	116	1-5	4.07	.85
How easy did you find the application? (1= very difficult; 5= very easy)	116	2-5	3.86	.78
Did you have any problems with the game? (1=very much; 5=no problems)	114	1-5	3.62	.94

participants to be tested over such a long period, which is why we had two groups. Nonetheless, the experienced user group’s user experience was also tested to determine whether length of time of use might have had an impact on the user ratings.

For the light users, the children played approximately 10-15 min per day, 3-4 days a week. The testing period was either 1-3 weeks for light users (n=50) or 4-5 weeks for experienced users (n=67). Out of these participants, 52 used Windows laptops and 65 used Android tablets. Following the testing period, they were given the questionnaire of 15 items in their native Finnish language, which covered the breadth of three key areas: overall affective reactions; value and interest in game/nongame versions and perceived pedagogical value (indexed by question on whether their teacher would like the game for the students) and finally a set of questions relating to audio and speech interaction (e.g., perceived sound quality, utility and affective reactions to having their speech samples tested).

3. Results

3.1. Affective Reactions. From Table 1, it can be seen that the children had a generally positive reaction to the software (mean scores around 4 on a scale of 1-5, with 5 being a positive rating). There were no differences between light users and extensive users on any of these measures.

Children were also asked about affective reactions in relation to the game and nongame versions of the program using the ‘Again again table’—i.e., “Would you like to use this program again?”—and they could say either ‘Yes’, ‘No’ or ‘Maybe’. Data from this method was analyzed for counts in each category using chi-squared analysis and showed that there was a significant difference between the game and nongame version ($\chi^2=11.89$, $p<0.05$, $df=4$). For the game version, children were less ambivalent and generally more positive (63 out of 101 said yes to the question, 32 said maybe and only 6 said no). By contrast, for the nongame version, children were less positive and more ambivalent (only 28 out of 101 said yes, 36 said maybe and 37 said no).

TABLE 2: Assessment of whether teacher or self would like the game version.

	yes	maybe	no
Like game for self	59	40	13
Teacher would like the game ¹	38	72	2

¹NB: this question was worded slightly differently in the first 50 participants but had wording changed later to ensure that the children would not misunderstand this question, from (a translation from Finnish) ‘Would my teacher like this?’ to ‘Do you think your English teacher would like pupils to use this game?’ We analysed the results from both wording versions and there was no statistically significant difference.

TABLE 3: Assessment of whether teacher or self would like the nongame version.

	yes	maybe	no
Like program for self	23	35	37
Teacher would like the program	22	59	14

TABLE 4: Rating of whether the player would prefer to play alone, with a friend or in a group.

	yes	maybe	no
Like to play alone	41	44	48
Like to play with friend	57	40	16
Like to play in a group	35	44	35

3.2. *Perceived Pedagogical Utility and Collaboration Preferences for Game and Nongame Versions.* The children were asked specifically about the perceived pedagogical utility by asking their views on whether their teacher would like the game versus whether they would like the game for themselves.

There was a tendency for the children to rate themselves as certainly liking the game for themselves (more ‘yes’ judgements) and a more ambivalent rating (more ‘maybe’ judgements) for the teacher ($\chi^2=22.21$, $p<0.01$), see Table 2.

For the nongame version, the children appeared to rate the self and teacher liking the game differently ($\chi^2=32.22$, $p<0.01$). In particular, they felt proportionally less ambivalent (when considering their yes/no responses) compared to the ratings the teachers who they felt would be more ambivalent and proportionally less negative, see Table 3.

Finally, when the children were asked separate questions as to whether they preferred to play alone, with a friend or with a group (results in Table 4).

It appears that in general the children appear to prefer playing with a friend more than they preferred to play alone or in a group. The differences between playing with a friend versus group might be due to the fact that ‘group’ may mean a group of people not known to the children and therefore they may be more ambivalent. Nonetheless, the finding that a greater proportion of positives and fewer negatives for playing with friends (compared to playing alone) would suggest that this age group tend towards preferring to play with known peers.

TABLE 5: Reactions from the SIAK application relating to speech and audio elements using the smiley-o-meter rating scale (NB: some items had lower N due to participants not filling in every question).

Question item	N	Range of scores	Mean rating	SD
How clear was the sound for the Finnish words? (1=not at all clear; 5=very clear)	116	2-5	4.13	0.97
How clear was the sound for the English words? (1=not at all clear; 5=very clear)	116	1-5	3.71	1.00
Did you like hearing your own pronunciation? (1=not at all; 5=very much)	114	1-5	3.83	1.21
Did you like getting feedback regarding your pronunciation? (1=not at all; 5=very much)	114	1-5	4.16	0.96

3.3. *Reactions to Elements Related to Speech Training Software.*

In this category, there were two sets of questions posed to the children that are useful to consider when designing speech training software. The first set was in relation to the rating of speech quality. Here, the children were asked two separate questions: ‘How clear was the sound for the Finnish-speaking words?’ and ‘How clear was the sound for the English-speaking words?’. The second set of questions related to the use of speech pronunciation feedback. As the program involved not only getting scoring as feedback, but also a replay in comparison to the native speech, they were asked two questions: ‘Did you like getting feedback regarding your own pronunciation?’ and ‘Did you like hearing your own pronunciation during the game?’

Interestingly, the children rated the Finnish samples as having better sound quality than the native English samples ($F_{1,112}=7.703$, $p<0.01$), see Table 5. Although the sampling rate, microphone frequency responses were the same, they were not collected in identical labs and hence there may have been subtle differences. However, it appears that experience with the samples may have also played a part in the rating. The more experienced users rated samples better compared to the less experienced users of the application ($F_{1,112}=4.739$, $p<0.05$). This would suggest that the tendency to rate English sounds as worse quality than the Finnish ones might be due to exposure to that particular language.

With respect to the other questions regarding pronunciation, we looked at the aspects of perceived helpfulness of pronunciation training versus the actual experience of hearing their own voice. As one might have predicted, the children liked getting feedback more than the process of actually hearing their own voice. This is probably due to the children feeling self-conscious about their voices, but yet seeing the value of feedback.

4. Discussion

The results from our extensive user trial show that children in general had a positive user experience from the speech training SIAK game, answering our first question regarding the affective responses of children in this age group to the application. There was evidence that they in general were positive about the application (scoring over 4 out of 5 on a scale of 1-5) and they had found the application fun and helpful. They also did not report major difficulties with the application and found it easy to use. This is encouraging as it is helpful to have a tool that is perceived to be easy to use and elicits positive user feedback in this group.

With respect to our second research objective, the game version was perceived more positively in this age group compared to the nongame version. Interestingly though, the children ranked the perceived pedagogical value (marked by the question of whether the teacher liked the game) as being more definitely more negative to the nongame version than the game version. This contradicts our initial hypothesis that the game-based version may be perceived as having less 'educational value' and therefore be perceived by the children as being less favourably rated by their teachers. On the other hand, this was coupled with the finding that the children rated their teachers as more ambivalent towards the game (and nongame) versions than the children, suggesting that children may not necessarily be clear as to what their teachers would think. When interpreting these results, we need to be mindful that the children thought the question was meant to ask whether the teachers would choose the application for themselves, whereas instead, the intended focus was to ask whether the teachers would choose the application for the child. As [20] states, usability testing in children can often lead to unexpected results and raise more issues in testing than originally envisaged. However, this explanation is unlikely given that an elaboration of the wording was made in a later stage of data collection to clarify understanding, which did not change the results. Of course, further research is needed to definitively tell whether the children understood the question in the way intended. For example, one possibility could be that 'theory of mind' might not be fully operational in the children at the lower end of the age range of the sample [21]. However, we did not see any age differences in the way these questions were answered either.

With respect to the issue of collaboration, it appeared that this age group slightly preferred interaction with friends. Although more data would be needed to confirm this, the trend seen in these data accord with the results of other studies (e.g., Heikkinen et al.'s JamMo implementation [22]) which showed that children of this age group report liking working in pairs or very small groups, particularly on mobile devices.

With respect to the third research question, we also sought to investigate whether there were any auditory interaction issues (perceptions of sound quality, experience of pronunciation feedback) that may impact on the user learning experience or learning outcomes. With respect to the listening experience: it appears that there were some perceptions of difference in sound quality between native and

nonnative speech. Although at the time of writing, detailed acoustical analysis on the speech was not available to know whether the perceived differences were the result of actual real subtle differences, the effect seemed to be moderated by experience (in other words, the heavy users of the program rated the nonnative speech better in sound quality than the more naïve users of the program). That is, the perception of sound quality may be affected by the difficulty to map speech input into existing mental representations. Further research is needed to determine whether sound quality is being affected by language experience.

With respect to pronunciation feedback, it was clear that this age group found feedback useful, but was less positive about hearing their own voice. These less positive ratings could be potentially mitigated by including assurances for the learners about confidentiality and the value of receiving a replay of their own voice. It is unclear whether this might reflect performance anxiety which may in turn see effects on performance. Further analysis on actual learning outcomes could explore whether there is a relationship as has been seen in other contexts [23].

In summary, future research will need to focus on the following areas:

- (1) Whether user perceptions of voice clarity occur in native versus nonnative language in other samples and differ as a function of nonnative language experience (as we found here). We would also be interested in investigating whether such biases in perceived clarity impacts negatively on learning outcomes in foreign language learning contexts.
- (2) Whether children in general are self-conscious of automated recognition as we found here and it would be useful to know whether such effects are modulated by age.
- (3) Whether the positive affect ratings for gaming result in improved language learning outcomes compared to nongamified versions of the implementation. Such data would be helpful to determine whether the value of gaming in other domains also transfers to automatic speech recognition.
- (4) Finally, we would be also interested in exploring whether children in general prefer to work with their friends in all learning contexts – or are there some situations where they may prefer to work alone (e.g., when they are being assessed and could be self-conscious).

5. Conclusions

In conclusion, these data serve a starting point of observations around speech training elements that are useful and well received for this 8-12-age group. Further work around the perception of speech quality in nonnative language and preference for collaboration is needed. What is clearly confirmed from the data is that the gaming aspect of the application is well-received and serves well as a positive tool to deliver speech training in this group. Children also interestingly

appear to rate their teachers as being more ambivalent but also less negative for the game version. This may suggest that they do not necessarily perceive the game aspect to have less pedagogical value than the nongame version, although further research is needed to clarify.

Data Availability

The anonymised, de-identified, raw data used to support the findings of this study are available from the corresponding author upon request.

Disclosure

A subset of preliminary data from this study were disseminated in an oral presentation and abstract form at Eurocall 2018, August 2018, Jyväskylä, Finland.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors gratefully acknowledge funding from Academy of Finland (Grant no. 1274058) and the funding from Visiting Professor Grant awarded to Professor Uther by Nokia Foundation to complete the work on this evaluation while on a visit hosted by Cicero Learning.

References

- [1] K. Reima et al., "SIAK - A Game for Foreign Language Pronunciation Learning," *Interspeech*, vol. 2017, 2017.
- [2] A. Rouhe, R. Karhila, K. Heini, and M. Kurimo, "A pipeline for automatic assessment of foreign language pronunciation," *ISCA Workshop on Speech and Language Technology in Education (SLaTE)*, vol. 2017, 2017.
- [3] J. Mostow, S. F. Roth, A. G. Hauptmann, and M. Kane, "Prototype reading coach that listens," in *Proceedings of the 12th National Conference on Artificial Intelligence. Part 1 (of 2)*, pp. 785–792, August 1994.
- [4] M. Black, J. Tepperman, S. Lee, P. Price, and S. Narayanan, "Automatic detection and classification of disfluent reading miscues in young children's speech for the purpose of assessment," in *Proceedings of the 8th Annual Conference of the International Speech Communication Association, Interspeech 2007*, pp. 941–944, Belgium, August 2007.
- [5] J. Duchateau, Y. O. Kong, L. Cleuren et al., "Developing a reading tutor: Design and evaluation of dedicated speech recognition and synthesis modules," *Speech Communication*, vol. 51, no. 10, pp. 985–994, 2009.
- [6] D. Bolan-Os, R. A. Cole, W. Ward, E. Borts, and E. Svirsky, "FLORA: Fluent oral reading assessment of children's speech," *ACM Transactions on Speech and Language Processing*, vol. 7, no. 4, 2011.
- [7] M. Gerosa, D. Giuliani, S. Narayanan, and A. Potamianos, "A review of ASR technologies for children's speech," in *Proceedings of the 2nd Workshop on Child, Computer and Interaction, WOCCI '09, USA, November 2009*.
- [8] J. Proença et al., "Mispronunciation Detection in Childrens Reading of Sentences," *IEEE/ACM Trans. AUDIO*, vol. 26, no. 7, pp. 1203–1215, 2018.
- [9] E. M. Golonka, A. R. Bowles, V. M. Frank, D. L. Richardson, and S. Freynik, "Technologies for foreign language learning: A review of technology types and their effectiveness," *Computer Assisted Language Learning*, vol. 27, no. 1, pp. 70–105, 2014.
- [10] T. Cincarek, R. Gruhn, C. Hacker, E. Nöth, and S. Nakamura, "Automatic pronunciation scoring of words and sentences independent from the non-native's first language," *Computer Speech and Language*, vol. 23, no. 1, pp. 65–88, 2009.
- [11] S. M. Abdou, S. E. Hamid, M. Rashwan et al., "Computer aided pronunciation learning system using speech recognition techniques," in *Proceedings of the INTERSPEECH 2006 and 9th International Conference on Spoken Language Processing, INTERSPEECH 2006 - ICSLP*, pp. 849–852, USA, September 2006.
- [12] F. Cornillie, S. L. Thorne, and P. Desmet, "ReCALL special issue: Digital games for language learning: Challenges and opportunities," *ReCALL*, vol. 24, no. 3, pp. 243–256, 2012.
- [13] T. M. Connolly, E. A. Boyle, E. MacArthur, T. Hainey, and J. M. Boyle, "A systematic literature review of empirical evidence on computer games and serious games," *Computers & Education*, vol. 59, no. 2, pp. 661–686, 2012.
- [14] B. DaCosta and S. Seok, "Developing a Clearer Understanding of Genre and Mobile Gameplay," in *Handbook of Research on Immersive Digital Games in Educational Environments*, Advances in Educational Technologies and Instructional Design, pp. 201–231, IGI Global, 2019.
- [15] P. Markopoulos and M. Bekker, "On the assessment of usability testing methods for children," *Interacting with Computers*, vol. 15, no. 2, pp. 227–243, 2003.
- [16] M. Uther and A. P. Banks, "The influence of affordances on user preferences for multimedia language learning applications," *Behaviour & Information Technology*, vol. 35, no. 4, pp. 277–289, 2016.
- [17] G. Sim and M. Horton, "Investigating children's opinions of games," in *Proceedings of the 11th International Conference*, p. 70, Bremen, Germany, June 2012.
- [18] G. Sim, S. MacFarlane, and J. Read, "All work and no play: Measuring fun, usability, and learning in software for children," *Computers & Education*, vol. 46, no. 3, pp. 235–248, 2006.
- [19] R. Karhila, "SIAK - A game for foreign language pronunciation learning," in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, vol. 2017, p. 2017, 2017.
- [20] G. Sim, J. Read, and M. Horton, "Practical and ethical concerns in usability testing with children," in *Games User Research: Study Approach, Case*, A. Peters, Ed., pp. 1–33, CRC Press, 2016.
- [21] C. I. Calero, A. Salles, M. Semelman, and M. Sigman, "Age and gender dependent development of theory of mind in 6- to 8-years old children," *Frontiers in Human Neuroscience*, no. MAY, 2013.
- [22] K. Heikkinen, J. Porras, J. Read, and G. F. Welch, "Designing Mobile Applications for children," in *User Requirements for Wireless, No. September*, L. T. Sorensen and K. E. Skouby, Eds., pp. 501–512, River publishers, Aalborg, Denmark, 2015.
- [23] D. E. Callan and N. Schweighofer, "Positive and negative modulation of word learning by reward anticipation," *Human Brain Mapping*, vol. 29, no. 2, pp. 237–249, 2008.

