Text-to-speech enhanced eBooks for emerging literacy development

Febe de Wet[1]*, Laurette Marais*[1] *& Daleen Klop*[2]

[1]Human Language Technology Research Group, CSIR, Meraka Institute, Pretoria, South Africa
[2]Division Speech-Language and Hearing Therapy, Department of Interdisciplinary Health Sciences, Faculty of Medicine and Health Sciences, Stellenbosch University, South Africa

fdwet@csir.co.za, laurette.p@gmail.com, dk@sun.ac.za

Abstract

The purpose of this study was to measure the efficacy of an eBook to improve the vocabulary and word recognition skills in an Afrikaans speaking group of lower socio-economic status of 6- to 7-year old children with poor vocabulary. The main goals were to investigate if exposure to an interactive eBook would result in the acquisition of new vocabulary and sight word reading in the study participants. A randomised pre-test/post-test between-subjects design was used. An experimental group that received an intervention was compared to a control group before the control group received a delayed intervention. Both groups were reassessed, eight weeks after the interventions to assess the retention of their newly acquired skills. Results show a significant improvement in recognition and vocabulary skills in the experimental group compared to their initial assessments, as well as compared to the control group.

Index Terms: text-to-speech, emerging literacy, eBooks

1. Introduction

Emerging literacy refers to the level of literacy of children who are only starting to read. In the South African context this specifically refers to the literacy skills of children in Grade R and Grade 1[1]. Literacy skills are critical to academic success and survival in a modern, industrialised, knowledge-driven society.

Unfortunately, low literacy levels of South African children are already apparent in the foundation phase, i.e. in the first three years of primary school. The Department of Basic Education's (DBE) Report on the Annual National Assessments of 2011 indicates that 53% of South African children in Grade 3[2] did not pass the literacy test and that the average score in literacy for a typical Grade 3 learner in South Africa was 35%.

An estimated 80% of state-run schools lack a library. Children therefore do not have access to supplementary reading material or literacy teachers. Children from disadvantaged backgrounds often also do not have access to literacy development tools or family members to assist them. As a consequence, they often start their formal schooling with limited experience with books and literacy and often display limited early literacy skills, such as phonological awareness, print awareness and knowledge of story structure.

This challenge is further exacerbated by the fact that South Africa has 11 official languages and that many children learn to read and write in a language that is not their first language (i.e. the language which they understand the best) due to issues such as having to attend the school nearest to their homes, and exposure to more than one language in the home and community. As a result, they are under-prepared for the formal curriculum and struggle to learn to read.

Evidence of this can also be found in the DBE's 2011 report, which indicates that Quintile 1 children (the poorest children) score on average 31% for literacy, while Quintile 5 children (the least poor) score on average 49% for literacy. Once the lower income group children have acquired basic reading skills, they often fail to read with comprehension in the later grades. Poor reading skills (and consequently poor language skills) have a negative impact on children's ability to acquire knowledge and skills in all subjects. This is compounded by education policy which dictates that in certain schools, African languages are used as medium of instruction up to Grade 3, at which time there is a switch to English as medium of instruction (Grade 4 and up).

Any attempts to improve emerging literacy in South Africa must therefore be easily extendable to applications in various languages. Given the widespread disparities in socio-economic situations of many South Africans, which may influence academic needs and performance, solutions must also be flexible enough to allow targeted intervention, geared towards each group of children in their immediate context.

Mobile technology and mobile development platforms have advanced to the point where computer based solutions can be deployed on mobile devices, such as smart phones and tablets. Such platforms provide a rich ecosystem within which language- and speech-related technologies can be developed. Specifically, text-to-speech (TTS) technology provides the opportunity for personalised and interactive reading of content as it converts written text into synthetic speech. In a society that depends increasingly on digital technology, exposing children to mobile technology is an added advantage.

The aim of this study, therefore, was to develop and test the effects of using a mobile application addressing aspects of emerging literacy using TTS technology in a South African context. Although similar commercial mobile applications that are geared towards literacy development are available, none are available in the 11 South African languages and none have had content developed specifically for the South African environment. Furthermore, not all available applications make use of TTS technology. Only a few have been developed with the aim of having teachers and literacy experts customise the content and technology to suit their requirements.

The application that was developed for the study consists of an Afrikaans eBook enhanced with TTS that can run on a mobile platform. It was designed to improve early literacy skills of children in Grade R and 1 and aims to address the problem of

[1]Grade R: 5 years old, turning 6; Grade 1: 6 years old, turning 7
[2]8 years old, turning 9

poor literacy teaching skills, poor access to supplementary reading material and poor access to literacy development support at home.

In addition, the eBook includes locally developed and culturally relevant content and graphics. The focus is on recognition and vocabulary skills, such as highlighting of sentences as they are read, and interactive activities where drag-and-drop gestures must be used to match words to picture referents. The application also aims to enhance human narrated books, by providing the possibility to interact with story content.

2. Background

Lumsden et al. [1] provide some guidelines for designing applications for "mobile experiential language-learning technologies", which include assessment of the users' needs and abilities, and then adapting the approach to take this into consideration. The ability to follow these guidelines therefore depends on a high degree of flexibility, since content must be designed to suit a specific target population. Text-to-speech technology provides much flexibility for generating audio content, offering "some of the benefits of a personal reader" [2], since any text may be read aloud on the fly. Although there are some limitations to TTS, such as the possibility of inappropriate intonations, an inability to personify characters, or occasional mispronunciations, high quality TTS technology is able to convey meaning very effectively [3], and it is by far more flexible and less expensive than pre-recorded speech.

It has been shown that students with special needs due to disability are more interested, pay more attention to detail and are able to learn more effectively when using electronic multimedia devices [4]. Such students also benefit from being more independent during the learning process. It has also been found that such devices are particularly useful for developing aspects of literacy such as phonological awareness [5], text visualisation and supported reading [6]. Children generally enjoy touching things, and therefore a learning environment that encourages touch gestures allows for "natural interaction" [4].

While this study does not focus on disabled students, many poor South African students have similarly specific educational needs because of the lack of support mentioned above. The application is therefore designed to incorporate multimedia content by making use of graphics, as well as audio in the form of a high quality synthetic voice, and so that it may be interacted with in an intuitive way via touch gestures and interactive feedback.

Presenting material in this way is not sufficient to ensure that children interact successfully with the content of the application. Two specific challenges arise. Firstly, the application must be designed to prevent a situation where children focus on producing certain multimedia effects, instead of interacting with the content itself. Children must be encouraged to think and react to the concepts and questions presented to them. This is mainly addressed by designing the flow of interaction and the feedback given by the application to encourage thoughtful responses. Secondly, the concepts and questions presented must be suitable to the children's context. More specifically, in designing the content, the background knowledge and learning environment of children must serve as a guide [7].

When the focus is on vocabulary learning, as in our case, the words and concepts children are likely to be familiar with must be considered. Furthermore, the story and characters in the eBook must also be suitable. Given the poor support for literacy for many South African children, the story should have a simple structure, recognisable themes and characters that children are able to relate to. Consequently, it is crucial that children have access to content in their own language that may be customised for their specific needs. Related studies published by recognised leaders in the field of electronic books to support language and literacy, such as [8, 9, 10, 11, 12], were consulted to aid in the conceptualisation of the study and the development of the methodology.

3. Content

Content planning and interface design was a combined effort between the emerging literacy expert and the development team. Principles for successfully incorporating TTS into the application were established, such as deciding on the conditions under which elements of the application would respond to touch gestures with TTS, deciding on strategies to keep the interaction with the application focused on the content as opposed to mere interaction with the TTS effects of the application, and developing a sequence of activities designed to assist the children in their progress.

The solution that was developed provides two kinds of activities: the first closely resembles reading of an eBook, while being enhanced with TTS and simultaneous highlighting, and the second focuses on interaction with pictures, words and sentences taken from the context of the book.

The purpose of the eBook is to expose children to general reading conventions, as well as introduce the content and vocabulary around which the activities are built. The purpose of the question and answer activities is to provide more in-depth interaction with specific words, while also introducing certain mobile technology conventions, such as the use of touch gestures.

The eBook was created to suit the developmental levels of the target groups, i.e. a repetitive storyline with themes and vocabulary suitable for 6- to 7-year old children. In the story a young girl walks with her dog in the wild; they encounter different animals (rabbit, frog, bird, mouse, and bee); the dog chases all the animals that disappear in their homes; finally, the dog chases a bee and the bee stings him on the nose.

The story content was created to comply with the goals of the intervention programme, i.e. the acquisition of expressive and receptive vocabulary and recognition of target words. The story content is therefore controlled for the number of word occurrences and variety of word classes (nouns, verbs, adjectives, and adverbs), sight words at the targeted reading level, and more advanced words for novel word learning.

4. Application development

The application was developed specifically for Android tablets and was enhanced through the use of TTS. There are three categories of outputs of the application:

1. The home page with choices for activities.
2. Activities consisting of interaction with the eBook content.
3. Multichoice-style quizzes with questions that are answered by clicking/dragging an object to the correct answer.

The home page of the application provides buttons for accessing six different activities. These activities range from reading through the eBook content and interacting with the eBook content, to multichoice-style activities for answering questions

about the eBook content. Activities 1 and 2 are eBook activities, while activities 3 to 6 are multichoice-style activities.

Activities 1 and 2 consist of a succession of pages that allow different levels of interaction with the content. Activity 1 allows the user to select the text in order for it to be read aloud by the TTS function, while sentences are highlighted as they are read. Activity 2 is similar, but allows the user to touch the characters on each page to hear additional information about them.

Figure 1: *Screenshot of the application in eBook mode.*

Activities 3 to 6 include various multichoice-style games with questions (read and highlighted) that are answered by clicking on or dragging an object to the correct answer. These activities allow for interaction with the content of the eBook. Using both dragging and click gestures, users can match answers with questions. Figures 2 and 3 provide examples of questions for the click and drag activities respectively.

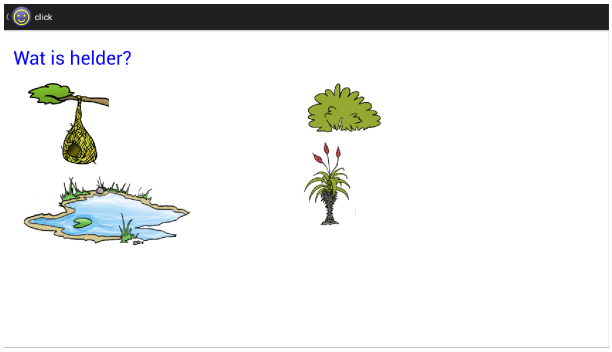

Figure 2: *Example question (*"What is bright?"*) in a "click" activity.*

On each screen, a question is presented to be read out aloud. When the gesture of the specific activity is correctly performed, the text of the answer appears at the bottom of the screen and it is read aloud.

5. Application implementation

5.1. Android

In Android, each screen extends the Activity class. The main activity of this application is the home page, while three other classes of activities are defined for the eBook and multichoice-style screens: one for the eBook screens, and two classes of activities for the multichoice-style quizzes, corresponding to the two gestures used, i.e. dragging and clicking.

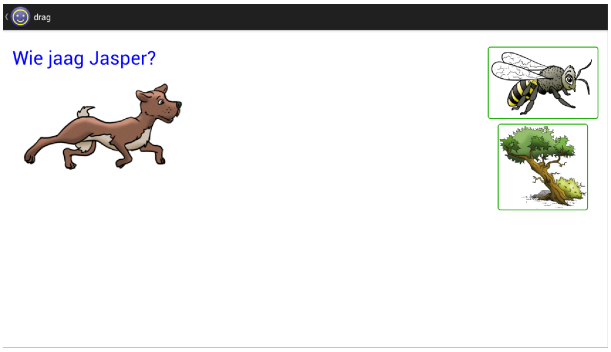

Figure 3: *Example question (*"Who does Jasper chase?"*) in a "drag" activity.*

TTS is used to read aloud all text used in the application, allowing for a larger degree of flexibility with regards to content and related activities than would be the case with human narrated audio. The application structure is modular to ensure that this flexibility is capitalised on effectively.

An XML-style structured format was developed for specifying all aspects of the content of the eBook and related activities, including images, text and information needed to generate multichoice screens. The multichoice activites are generated randomly from a list of "matches" - objects containing a question, its answer and alternative answers. The format allows specifying a number of levels for each activity, where each level presents the user with an increasing number of alternative answers. The XML document is parsed into a BookTree object, which is used by a Book object in the application. The Book object is, in turn, used by all classes extending Activity to populate various elements of the screen, depending on the kind of activity.

5.2. Text-to-speech

The application uses the Qfrency[3] Afrikaans voice to convert text to speech. Qfrency is a TTS system that was developed specifically for South Africa's official languages.

Each page in the eBook carries a list of sentences. As soon as a user clicks on the text block displayed on the screen, each sentence is read aloud and highlighted in turn. Internet access was not available in the deployment environment. As a consequence it was not possible to access a remote TTS server via the application. Instead, for the purpose of the intervention, the text-to-speech audio for the custom content was pre-rendered and installed as part of the application. A lookup function matched each sentence in the Book tree with its corresponding audio, and the Android API was used to keep track of the list of audio tracks to be played for each page. Callbacks were used to synchronise the highlighting of the sentences as they were read aloud.

A similar approach was followed for the multichoice-style quizzes, especially with regard to the question presented on each page. In this case, the appearance of the answer text and the playback of its audio is triggered whenever the correct answer is given by the user via clicking or dragging gestures.

[3] http://www.qfrency.com/

6. Intervention

Participants were recruited from the Grade 1 class ($n = 42$) of an Afrikaans-medium farm school in a lower socio-economic rural community. For inclusion in the study, children had to pass otoscopic examinations, pure tone hearing and optometric screening tests, and had to have below average language skills and non-verbal intelligence skills within normal limits.

The Afrikaanse Reseptiewe Woordeskattoets (ARW) [13] was used to assess the children's receptive language skills. All the children obtained standard scores ranging from 55 to 85. This is one to two standard deviations below the mean standard score, indicating that they have language impairments [13].

Their non-verbal skills were assessed by means of the Test of Nonverbal Intelligence-4 (TONI-4) [14]. Their TONI-4 index scores ranged from 82 to 104 (scores below 90 indicate below average performance). Twelve children who did not meet the selection criteria were excluded from the study and referred for further testing.

The final study population comprised 28 children. After matching for gender, the children were randomly assigned to the experimental ($n = 14$) and control ($n = 14$) groups. The two groups were comparable in terms of age. One-way ANOVAs showed no differences between the groups for ARW standard scores, $F(1, 27) = .06202, p = 0.81$, and TONI-4 index scores $F(1, 27) = .05306, p = .82$. Both groups comprised of seven boys and seven girls.

Participant questionnaires were used to determine their previous exposure to digital technology. None of the children reported that they had computers at home and none had cell phones or were allowed to play games on their parents cell phones. They had, however, been exposed to educational computer programmes since they started formal schooling.

The experimental group ($n = 14$) received the eBook intervention programme in groups of 3 to 4 children, 3 times per week in 20 minute sessions, for a period of 2 weeks; i.e. 6 sessions in total under supervision. Two different approaches were taken during the intervention. In the first approach, children were exposed to new words, to determine if the application could assist them in learning to use and recognise new words. In the second approach, children were also exposed to representations of new words, but the aim was to determine whether the application could assist them in learning to read words they already knew. A facilitator was used to oversee the use of the application by children in small groups, but was not allowed to help them with the content of the application. The control group received no intervention apart from normal classroom activities.

After the experimental group completed the intervention programme, the control group ($n = 14$) and the children who did not comply with the selection criteria ($n = 12$) received the eBook intervention programme in exactly the same way as the experimental group.

After a period of 8 weeks, the assessments were repeated on the experimental and control groups to determine their retention of newly acquired skills.

7. Results

The results of the screening and the pre- and post-intervention assessments were coded and analysed by the researchers. In addition, they were independently coded by a third research assistant who was blinded for the group status of the participants. A commercial software package, Statistica 12, was used for the statistical analyses of the data by a biostatistician.

Intra-class correlations (ICC) were calculated to determine inter-rater reliability. For both word definition and word recognition, the ICC was found to be > 0.99 which indicated near perfect inter-rater reliability. In order to examine the differences between the participants in the experimental and control groups, mixed model repeated measures ANOVAs were used. A 5% significance level ($p < 0.05$) was used as guideline for determining significant effects of variables.

The results of the pre-, post- and follow-up assessments for the experimental and control groups are summarised in Table 1. The table shows that the control and experimental groups had similar pre-test scores, but the experimental group performed significantly better than the control group after the intervention on the receptive and expressive vocabulary measures.

The follow-up results, eight weeks after completion of the intervention programmes, show significant improvement in the control groups receptive and expressive vocabulary scores after they had also received the intervention. The follow-up measures for the experimental group indicate that they had retained these gains. With regard to the word recognition scores, however, no differences were observed between the groups before or after the interventions took place.

7.1. Receptive vocabulary

The receptive vocabulary assessments measured the children's receptive knowledge of the 15 target word meanings after they had been exposed to the words in the story context and intervention activities. A significant interaction between group and time was found, $(F(2, 52) = 24.11, p < 0.05)$.

7.2. Expressive vocabulary

7.2.1. Sentence completion

The sentence completion task assessed the children's ability to complete sentence cues with corresponding pictures by using the target vocabulary words. A significant interaction between group and time was found, $(F(2, 52) = 8.6605, p < 0.05)$.

7.2.2. Word definitions

This subtest assessed the children's ability to provide definitions of the target vocabulary words. A significant interaction between group and time was found, $(F(2, 52) = 8.2109, p < 0.05)$.

7.2.3. Word recognition

The word recognition test measured the children's recognition of selected target words from the text in the intervention programme. No significant interaction between group and time was found, $(F(2, 52) = .09851, p = 0.91)$. The two groups had similar pre-, post- and follow-up test scores. A gradual improvement in the scores for both groups over time was observed, probably as a result of schooling rather than their exposure to the eBook.

7.3. Reception of TTS voice

The children did not seem affected by the TTS voice. When asked to comment on the voice, some said that the story was read by a "computer", which suggests that they recognised it as being a synthetic voice.

Table 1: *Means (and ranges) for receptive vocabulary, sentence completion, word definitions and word recognition scores for participants in the experimental (Exp, n=14) and control (Control, n=14) groups for the pre-, post- and follow-up assessments.*

Assessment task	Pre-test		Post-test		Follow-up	
	Exp	Control	Exp	Control	Exp	Control
Receptive vocabulary (max score =15)	6.6 (3-10)	6.4 (3-9)	12.3 (8-15)	6.1 (2-9)	11.9 (9-15)	9.6 (3-15)
Sentence completion (max score =15)	0.9 (0-4)	1.4 (0-9)	5.6 (1-15)	1.3 (0-3)	4.1 (1-13)	3.4 (0-10)
Word definitions (max score =45)	8.6 (3-12)	7.2 (1-13)	17 (4-32)	8 (2-14)	15.6 (4-28)	12.3 (3-25)
Word recognition (max score =30)	11.8 (0-24)	10.7 (0-27)	14.6 (0-25)	13.9 (0-26)	19.6 (2-29)	18.1 (5-30)

8. Conclusions & Future work

The goal of this study was to determine if exposure to a short-term interactive eBook intervention programme would improve the language and literacy skills of an Afrikaans speaking group of lower socio-economic status of 6- to 7-year old children with language impairments and little previous experience in the use of digital technology. The data-analysis of pre-, post- and follow-up test scores indicated that the children's receptive and expressive vocabulary improved as a result of the intervention. All the participants acquired new words at different levels of semantic representation, although some children showed only modest gains. The small sample size of this study limits the generalisation of the findings, but the results indicate that interactive eBooks may be effective tools in improving children's language skills.

After the successful completion of the intervention with Afrikaans speaking children, an isiXhosa version of the application was developed. It will be used during a similar intervention in the Eastern Cape province of South Africa during the second quarter of 2015.

The initial concept of the eBook included word-level synchronised reading as a feature. However, in the Android TTS API the audio playback is asynchronous, and the Android operating system only provides callback notifications to utterance start and utterance end events. Finer grained audio playback notification is therefore not available through the operating system. This limitation will be addressed in future work to implement word-level synchronised reading in the application.

9. References

[1] J. Lumsden, R. Leung, D. D'amours, and D. McDonald, "Alex©: a mobile adult literacy experiential learning application," *International journal of mobile learning and organisation*, vol. 4, no. 2, pp. 172–191, 2010.

[2] E. Balajthy, "Text-to-speech software for helping struggling readers," *Reading Online*, vol. 8, no. 4, pp. 1–9, 2005.

[3] P. Taylor, *Text-to-Speech Synthesis*. Cambridge: Cambridge University Press, 2009.

[4] Álvaro Fernández-López, M. J. Rodríguez-Fórtiz, M. L. Rodríguez-Almendros, and M. J. Martínez-Segura, "Mobile learning technology based on iOS devices to support students with special education needs," *Computers and Education*, vol. 61, no. 0, pp. 77 – 90, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0360131512002199

[5] M. C. McKenna, L. Labbo, D. Reinking, and T. Zucker, "Effective use of technology in literacy instruction," *Best practices in literacy instruction*, vol. 2, pp. 307–331, 2003.

[6] N. Mana and O. Mich, "Design of customizing applications to support dyslexic children in reading."

[7] P. Kim, T. Miranda, and C. Olaciregui, "Pocket school: Exploring mobile technology as a sustainable literacy education option for underserved indigenous children in Latin America," *International Journal of Educational Development*, vol. 28, no. 4, pp. 435–445, 2008.

[8] C. A. Kegel and A. G. Bus, "Online tutoring as a pivotal quality of web-based early literacy programs." *Journal of Educational Psychology*, vol. 104, no. 1, p. 182, 2012.

[9] D. J. Smeets and A. G. Bus, "Interactive electronic storybooks for kindergartners to promote vocabulary growth," *Journal of experimental child psychology*, vol. 112, no. 1, pp. 36–55, 2012.

[10] O. Korat, "Reading electronic books as a support for vocabulary, story comprehension and word reading in kindergarten and first grade," *Computers & Education*, vol. 55, no. 1, pp. 24–31, 2010.

[11] A. K. Moody, L. M. Justice, and S. Q. Cabell, "Electronic versus traditional storybooks: Relative influence on preschool childrens engagement and communication," *Journal of Early Childhood Literacy*, vol. 10, no. 3, pp. 294–313, 2010.

[12] M. J. Verhallen and A. G. Bus, "Low-income immigrant pupils learning vocabulary through digital picture storybooks." *Journal of Educational Psychology*, vol. 102, no. 1, p. 54, 2010.

[13] M. Buitendag, "Afrikaanse reseptiewe woordeskattoets," *Pretoria: Human Sciences Research Council*, 1994.

[14] L. Brown, R. J. Sherbenou, and S. K. Johnsen, *TONI-4, Test of Nonverbal Intelligence*. Pro-Ed, 2010.