Comparing Semiliterate and Illiterate Users' Ability to Transition from Audio+Text to Text-Only Interaction

Leah Findlater[1], Ravin Balakrishnan[2], Kentaro Toyama[3]

[1]Department of Computer Science
University of British Columbia
Vancouver, Canada
lkf@cs.ubc.ca

[2]Department of Computer Science
University of Toronto
Toronto, Canada
ravin@dgp.toronto.edu

[3]Microsoft Research India
Bangalore, India
kentoy@microsoft.com

ABSTRACT

Multimodal interfaces with little or no text have been shown to be useful for users with low literacy. However, this research has not differentiated between the needs of the fully illiterate and semiliterate – those who have basic literacy but cannot read and write fluently. Text offers a fast and unambiguous mode of interaction for literate users and the exposure to text may allow for incidental improvement of reading skills. We conducted two studies that explore how semiliterate users with very little education might benefit from a combination of text and audio as compared to illiterate and literate users. Results show that semiliterate users reduced their use of audio support even during the first hour of use and over several hours this reduction was accompanied by a gain in visual word recognition; illiterate users showed no similar improvement. Semiliterate users should thus be treated differently from illiterate users in interface design.

Author Keywords

ICT4D, literacy, semiliteracy, development, accessibility, inclusive design

ACM Classification Keywords

H.5.2 [User Interfaces]: Evaluation/methodology, interaction styles.

INTRODUCTION

An estimated 785 million adults around the world are illiterate [7]. More than one third of this population lives in India, where the literacy rate is 65% according to the 2001 Indian census [6]. As access to online information and software tools is increasingly required for individuals to fully participate in society, we need to consider how these systems might better support different levels of literacy.

Interfaces with minimal or no text that instead utilize one or more of graphical, audio and numerical components (to leverage numeracy skills) have been shown to be useful for users with low levels of literacy [1, 3, 10, 15, 16, 20, 23, 25,

Figure 1. Setting for Study 1, showing the translator observing the participant.

26]. Previous research, however, has largely not distinguished between illiterate individuals and the vast number of semiliterate individuals, those who have some familiarity with letters and words but may in fact have difficulty reading even a simple text passage; by one estimate, 75-80% of the literate population in India may be only semiliterate [14]. As an illustrative context, Medhi et al.'s [16] text-free employment search engine matches domestic labourers with jobs in Bangalore, India. The search engine has been shown to be useful for low literacy users in general, but the target population has from 0 to 6 years of education, encompassing a range of literacy skills.

Grouping together all low literacy users runs the risk of overlooking potential benefits of text for semiliterate individuals in terms of both performance with the interface and broader impact on reading skill acquisition. Everyday exposure to text provides important incidental learning opportunities [13], which are important for reading skill reinforcement and the maintenance of rudimentary skills. This is the theory behind including same language subtitles in local-language television programs, a technique which has been shown to improve script decoding in Gujarat, India [14]. From an interaction perspective, text offers a fast and precise mode of communication. Augmenting a text-based interface with another modality such as audio or images could provide non-textual support for low literacy users when needed but could ease semiliterate users into interacting more with the text as they gain experience.

Although some studies have shown that text-free interfaces are preferred by low literacy users in comparison to fully text-based interfaces [16], there is some indication, albeit in a study with just one user, that text can help a person with low literacy to disambiguate images over just a few sessions with the interface [3]. As well, users with strong literacy skills are more likely to prefer an interface with at least some text over none [24]. However, there have been no formal evaluations to isolate the preference and performance differences of fully illiterate users from semiliterate users. It is unclear whether previous findings on minimal or no text interfaces apply directly to semiliterate users, or whether semiliterate users, even with very little education, can benefit from text in the interface.

To address this question, we conducted two controlled experiments using an experimental interface and search task provided in Kannada, the predominant language in the state of Karnataka, India. In the first study, 12 semiliterate participants (Figure 1) showed performance improvements that were coupled with reduced use of audio support. The second study measured longer-term learning with the interface and changes in visual word recognition, and compared the performance and preferences of 12 participants with different levels of literacy (illiterate, semiliterate, and literate in other languages). All user groups improved their task performance, while both the semiliterate participants and those literate in other languages reduced their reliance on audio support and significantly improved their ability to recognize written words over the course of the study.

The primary contribution of this paper is empirical evidence to show that, in contrast to the needs of fully illiterate users, a text-based user interface augmented with audio can be beneficial for semiliterate users. For these users, an improvement in performance on a simple search task was coupled with an improvement in visual word recognition, even though this improvement was incidental to the main task goal; this benefit was also seen with users literate in languages other than the target language. From a designer's point of view, these findings suggest that illiterate and semiliterate audiences may need to be considered as very different user populations. The addition of audio annotations to an existing text-based interface can significantly support semiliterate users and should be easier than creating a new, text-free one.

BACKGROUND
We summarize related work on interface design for low literacy users and background on reading skill acquisition. Although not directly related to our work, we note that multimodal interfaces have also been used to support individuals with language impairments (e.g., [17]).

User Interfaces for Low Literacy Users
One guideline for interface design for low literacy users is to include little or no text [16]. Research prototypes with minimal text have been developed for domains such as employment search [16], financial management [20], agriculture [22], video email [23], an electronic bulletin board [24], web page creation [10], and healthcare [12, 25]. These prototypes have been evaluated to varying degrees; we discuss the more generalizable findings.

Medhi et al. [16] compared text-free interface designs for an employment search application and a map application to corresponding text-based versions. An evaluation with low literacy users (0 to 6 years of education) showed that the text-free versions were preferred and increased task accuracy. In contrast to our study, their text-based interface was purely textual and was not augmented with audio. More recently, in a study comparing participants' understanding of health information represented as text, drawings, photographs, videos or animation, all with and without audio, the conditions with audio resulted in higher understanding and task accuracy than conditions without audio [15]. Some participants also reported the text condition to be confusing. While this highlights the importance of audio for low literacy users, the authors acknowledge that some of the accuracy benefit may have been due to participants mimicking the audio rather than understanding the material.

In the design of a financial management system for rural micro-credit groups in India, Parikh et al. [20] conducted iterative design sessions with 32 women. The final design included numbers to leverage numeracy skills, icons, audio, and text in the local language. Audio augmentation was found to be useful for disambiguating items. Although the impact of different levels of literacy was not the focus of their research, the findings provide preliminary support for our work: most participants were able to use the final design, except for those with the lowest levels of literacy. Parikh et al. [21] have also confirmed the importance of audio in the context of a cell phone application for capturing paper-based information. Rural literate users who tested a text-only version of the interface after using it with both audio and text preferred the version with audio.

Finally, Shakeel and Best [24] designed a community bulletin board catering to a range of literacy levels by combining audio, images and text. An evaluation compared the preferences of low literacy users to users who had passed a basic literacy test. Only 2 out of 5 of the low literacy users wanted at least some text, in comparison to all 5 of the users with stronger literacy skills. The evaluation was not designed to measure performance or to assess longer-term, broader impact (e.g., reading skill acquisition) of working in the different versions of the interface.

Reading Skill Acquisition
The long history of research into reading includes many theories, sometimes conflicting, on how beginner readers learn to read and how best to teach these skills [8]. While designing a system to teach users how to read was not the primary goal of our research, a general understanding of prevailing theories can help to interpret our results.

The stages through which beginner readers progress are varyingly defined by different researchers (for a summary see [2]), but a commonality among these theories is that beginning readers first use visual cues (including contextual cues), before progressing to the use of alphabetic or phonemic cues (mapping symbols and syllables to sounds). During the first stage, characteristics such as the shape of the first letter can act as visual cues for remembering a word. The surrounding context can also be used to facilitate recognition, for example, the consistent location of a word on a package of food. During the alphabetic or phonemic cue stage, readers can map sounds to some symbols and use that as a less arbitrary cue to retrieval [2]. More complex strategies are used by skilled readers, such as reading new words by drawing analogies from known words.

Over 35 million people speak Kannada [18], the predominant language of the state of Karnataka, India. Kannada is an alphasyllabary, in contrast to the alphabetic orthography of English [19]. This means that each print unit, or *akshara*, represents a syllable but may contain features that indicate subsyllabic information, such as the vowel sound. There is a one-to-one mapping between aksharas and phonological syllables. One difference between English and Kannada is that the development of phonological awareness (mapping of symbols to sounds) is faster in English [19]. The reading skill acquisition research summarized above was done with English, but the general phases should still apply with Kannada.

Incidental Language Learning and Subtitling
Incidental learning plays an important role in language acquisition, and may even be better for long-term vocabulary growth than explicit learning [13]. Through its multimodal nature, subtitling offers an opportunity for incidental learning and has been shown to be beneficial for novice learners of a second language (e.g., [9]). Multimedia software, though not necessarily multimodal, has also been explored in this context [4]. However, in the context of illiteracy, incidental learning through multiple modalities has not received as much attention.

One exception is a large-scale, long-term study by Kothari et al. [14] on the effect of adding subtitles to a local language music video program broadcast on television in the state of Gujarat, India. Over a six month period the show was broadcast for a total of 8 hours, but viewers could also request hard copies of the song lyrics by mail. Pre-test and post-test reading assessments were administered to 358 viewers and results showed an improvement in the ability to read syllables and short words. In the case of same language subtitling, the main "task" was to watch the music videos; the learning of script characters was incidental. One goal of our research is to investigate if this phenomenon can also occur during interaction with a software user interface.

STUDY 1: USE OF REDUNDANT AUDIO AND TEXT
We designed an experimental system that included redundant audio information for each text component. The

goal of Study 1 was to investigate the hypothesis that although illiterate and semiliterate individuals would almost certainly need to use the audio to complete a basic word search task, having text in the interface would lead to a performance benefit over time. This would be manifest as a decreased reliance on audio as they gained experience with the interface and hence familiarity with the text.

Experimental Methodology

Task
Using the experimental system shown in Figure 2, the task was to search for words among a list of 40 Kannada words. These words were chosen from a set of abstract words provided in the MRC Psycholinguistic database,[1] such as "government", "general" and "nature". Abstract words are more difficult than concrete words to represent in graphical format, hence text is particularly useful for disambiguation in these cases.

At the beginning of each trial (selection), the system played an audio recording of the word the participant needed to find. If the participant felt comfortable reading the words, he/she could visually scan the list and select the correct word by tapping on it with the pen. Otherwise, tapping with the pen on an audio button paired with each word on the list would cause the system to read that word aloud. Participants could exclusively use visual search, audio search, or a combination of the two. If the participant

Figure 2. Screenshot of experimental system showing text and audio pairs; control buttons allow the participant to continue to the next trial upon completion of the current one, and to replay the audio prompt for the current trial. English text labels are for illustration only, and were not shown on screen.

tapped an incorrect word, a red 'X' appeared next to that word and the participant had to select the correct word before continuing to the next trial. We required a pen tap to start audio clips because we wanted participants to explicitly decide whether or not to use the audio; another option would be to have it play when the pen hovered over the audio icon but we felt this would have led to too many inadvertent audio playback activations. Playback of an audio clip was interrupted if the user made a selection (i.e., tapped a word) or tapped another audio clip.

After the correct word was selected, the set of 40 words shown to the participant was reordered randomly to isolate visual pattern recognition from spatial learning. A task block consisted of 10 trials and participants completed 10 task blocks in a row, for a total of 100 trials per participant. The particular set of 10 words to be tested in each task block was randomly pre-selected from the same list of 40 words at the beginning of the study for each participant, and the order of their appearance in trials was randomizd for each task block.

Apparatus
The experiment ran on a Toshiba Tecra M4 Tablet PC with a 14" display at 1400x1050 resolution. Participants used the pen to interact with the Tablet PC in "slate mode", with the keyboard inaccessible. We chose the pen over the mouse or touchpad because it is typically easier for novice users to manipulate. The system recorded all performance data.

Procedure
At the beginning of the session, participants were asked to read a short sentence in Kannada aloud, to filter out those potential participants who did not have any difficulty reading. After answering several background questions, participants were introduced to the experimental system and completed a short practice block of 5 trials. This was followed by the 10 test blocks of 10 trials each. Participants were given the opportunity to take a short break between blocks if desired.

Design
The experimental design was a single-factor (task block) within-subjects design.

Measures
We considered three main dependent variables:
1. **Task completion time**. Time to complete all 10 selections in a task block.
2. **Error rate**. Number of incorrect words selected in a task block; note that because participants could not proceed to the next trial until they had correctly completed the current trial, this number could potentially be higher than the total number of trials.
3. **Number of audio invocations**. Number of times the participant clicked on an audio button in a task block.

Participants
14 participants who self-idenfitied as being semiliterate were recruited from the community by word of mouth. All but two were able to read at least some Kannada; none were fully fluent readers in any language. More than half of them spoke at least one other language, including Tamil, Telugu and Hindi. There were 9 females and 5 males, with an average age of 30 (range 19 to 37) and average of 2.3 years of schooling (range 0 to 6 years). None of them had any computer experience.

Participants were each given a gift worth approximately 60 rupees (about US$1.50) for their time. Sessions were conducted in a location of the participant's choice, including our research lab, a local NGO office, and the home of one of our research assistants. Note that from an experimental control perspective, it would have been preferable to have conducted all sessions in the same locale; however, the ground realities in conducting research in the developing world necessitated some compromises in order to accommodate the schedule and travel constraints of participants.

Hypotheses
Our main hypotheses were that task completion time, audio use and errors would all decrease across the 10 task blocks.

Results
We ran a one-way repeated measures ANOVA (task block) for each of the three measures. A Greenhouse-Geisser adjustment for non-spherical data was applied to all analyses. Two participants were excluded from the analysis because they did not need the audio at all during at least 2 of the first 5 task blocks (i.e., they were clearly literate), thus we report on data from only the 12 participants who clearly required the audio. All three of our main hypotheses were supported: there were significant learning effects for each dependent measure.

Participants used less audio as they became more familiar with the task. There were 40 words in the interface and these were randomly reordered for each trial, so a linear search using only the audio to find a word would result on average in 200 audio invocations per task block (20 per trial). Measured audio usage was lower than this: the number of audio invocations dropped from an average of 189.5 uses ($SD = 125.2$) in task block 1 to 105 uses ($SD = 99.5$) in task block 10 (main effect of task block: $F_{3.19, 35.1} = 4.25$, $p = .010$, $\eta^2 = .279$).

Task completion times and error rates also decreased. The reduced use of audio may also have had some impact on task completion times, which fell from an average of 476.7 seconds ($SD = 228.8$) in block 1 to 264.6 seconds ($SD = 147.0$) in block 10 ($F_{3.55, 39.1} = 7.66$, $p < .001$, $\eta^2 = .410$). Finally, error rate was also impacted by task block, dropping from an average of 7.4 errors in block 1 ($SD = 5.95$) to 2.8 errors in block 10 ($SD = 5.41$) ($F_{3.31, 36.5} = 3.10$, $p = .034$, $\eta^2 = .220$).

Observations and Discussion
The combination of subjective feedback and significant reduction in audio use even in a short 1.5 hour study session
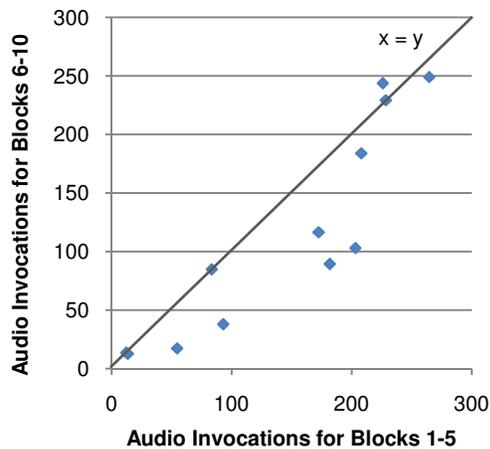
Figure 3. Audio usage in Study 1 during first half versus second half of task blocks, where each point represents one participant's data ($N = 12$).

is encouraging. Participants picked up the experimental task quickly, even though none of them had previously used a computer. Feedback on the combination of text and audio was also positive: when asked afterward if they would prefer to use an interface with sound only, sound and text, or text only, 9 of the 12 participants chose the combination of sound and text; the remaining 3 participants chose the sound only option. The 3 participants who used the least audio still felt it was important to have because it gave them confidence and allowed them to confirm that they had found the right word before selecting it. 6 other participants reported their strategy changed over time and they started using cues such as word length to help identify words.

All participants had self-identified as being semiliterate and we did not further distinguish participants based on literacy level. However, the rate of learning and comfort with the text and/or audio should be impacted by different levels of literacy. Figure 3 shows a graph of average audio use per participant during blocks 1-5 versus blocks 6-10. Points on the diagonal represent instances where participants used the same number of audio invocations in both halves of the study, while points below the diagonal indicate instances where audio invocations dropped in the second half of the study. The pattern suggests that those participants who started off in the middle of the group in terms of audio use reduced their reliance on audio the most during the second half of the study. Those participants who relied on it extensively did not show as marked an improvement. We more systematically explore this pattern in Study 2.

STUDY 2: MEASURING READING IMPROVEMENT AND IMPACT OF LITERACY LEVEL

Study 1 was not designed to identify the specific causes for the reduction in audio use over time. This reduction and corresponding improvement in task completion time could have been due to several factors, including an improved ability to visually decode words, reliance on short-term memory versus longer-term learning, and a general increase in familiarity with the experimental task and setup.

The goal of the second study was twofold: (1) to determine whether the improvement seen in Study 1 was at least partially related to improvement in longer-term word recognition (measured over 4 days of participation), rather than only based on general task familiarity or short-term memory, and (2) to understand how different levels of literacy among participants impacts interaction.

Experimental Methodology
Since the methodology is similar to Study 1, we only highlight the differences here.

Task
The experimental task was again to search for a set of 10 target words in turn out of a total set of 40 words but, unlike in Study 1, the target words were randomly pre-selected only once and were the same for all participants to allow for better comparison between participants.

Procedure
Participants performed 5 task blocks on each of 4 consecutive days, for a total of 50 trials per day, and 200 trials over the entire study per participant. At the beginning of the first day, as with Study 1, we asked participants to read a sentence aloud in Kannada (the initial reading assessment), followed by a set of background questions. At the end of each day we administered a reading test (explained below) and asked participants to describe the strategy they had used to complete the task that day. On the final day we also conducted a short interview to collect feedback about the combination of text and audio, the search strategies that participants used, and their overall preferences. Study sessions were at most one hour long, with the middle two sessions as short as 30 minutes.

Design
A 2-factor mixed design was used: task block (within-subjects) and literacy group (between-subjects).

Measures
In addition to measuring task completion time, error rate and audio invocations as in Study 1, we administered a reading test at the end of each session. It consisted of 20 words from the experimental application (the 10 target words from the search task and 10 other words that were randomly chosen from the remainder). Participants were asked to identify and read aloud as many words as they could from the list.

Participants
Twelve participants (7 males; 5 females) completed this phase of the study. The average age of each of the groups was from 26 to 28 years (range 17 to 36). Literacy groups were defined as follows:

1. **Illiterate:** The 4 participants in this group self-reported as not being able to read at all and on our initial reading assessment, none of them could read a single syllable. One of them had one year of schooling while the remaining 3 had no formal education. Only 1 of them had ever used a computer.

2. **Semiliterate:** These 4 participants self-reported as being somewhat fluent readers in Kannada and not fluent in other languages. All could read some syllables and words with great difficulty on the initial reading assessment. They had from 1 to 6 years of education ($M = 3$). None of them had used a computer.

3. **Literate in another language** (but not in Kannada): These 4 participants all had at least one university degree and all self-reported as being literate in English and at least somewhat literate in one or more other languages (Sinhala, Hindi, Telugu, Konkani or Tamil). Two of them spoke but could not read or write Kannada; the others did not speak Kannada beyond a few words. This group was included to see how literacy in another language might impact learning of a new language, as opposed to those who were only semiliterate.

Upon completion of the study, participants were given a gift worth approximately 150 rupees (about US$3.75, or roughly a day's wages for a casual labourer in the city where the experiment was conducted). The Literate participants were recruited from our research lab, while the Illiterate and Semiliterate participants were recruited through word of mouth from a network of housekeeping staff and from a nearby construction site. Recruitment was especially difficult because of the four day commitment. The construction site participants only had free time late in the evening once they completed their work. The housekeepers often had several jobs during the day and were also pressed for time to participate between jobs. Sessions were conducted either at our research lab or at the home of one of the research assistants based on which location was more convenient for the participants.

Hypotheses
We expected that the Literate group would behave similarly to the Semiliterate group because of knowledge transfer from other languages. Our main hypotheses were:

1. Audio use: Audio use for Semiliterate and Literate participants will decrease with practice (overall trends from Study 1). Audio use for Illiterate participants will not decrease with practice (as suggested by Figure 3).

2. Reading test score: Reading test scores for Semiliterate and Literate groups will improve with practice.

3. Preference: Illiterate participants will prefer audio only (based on previous research [15]). Semiliterate and Literate participants will prefer a combination of audio and text (based on Study 1).

Results
For each of the dependent variables we ran a 2-way mixed factorial ANOVA with Day as a within-subjects variable and blocking on Group (i.e., literacy group) as a between-subjects variable. Bonferroni adjustments were applied for all pairwise comparisons. Where *df* is not an integer, this is because we have applied a Greenhouse-Geisser adjustment

for non-spherical data. We report on all significant results ($p < .05$) and those that represent a possible trend ($p < .10$).

Reading Test Score
Overall, the Illiterate group recognized the fewest words, on average only 0.9 words per day, while the Semiliterate and Literate groups achieved higher average scores of 3.7 and 6.2 words, respectively.

Participants were able to identify and read more words as the study progressed (Figure 4; error bars in all charts show standard error). On average, 1.4 out of 20 words were identified correctly on the first day, up to 5.8 words on the fourth day (main effect of Day: $F_{3,27} = 38.7$, $p < .001$, $\eta^2 = .811$). More interestingly, the three literacy groups also showed different degrees of improvement (main effect of Group: $F_{2,9} = 16.4$, $p = .001$, $\eta^2 = .785$).

Semiliterate and Literate users improved on reading test score but Illiterate users did not. The pattern of change across days differed for the three literacy groups (interaction effect of Day x Group $F_{6,27} = 11.3$, $p < .001$, $\eta^2 = .716$). Both the Semiliterate and Literate groups improved significantly over the course of the study, but no significant improvement was seen for the Illiterate group. By Day 3 for the Semiliterate group there was a significant increase in reading test score over the first day ($p = .020$). The Literate group improved even faster, scoring significantly higher by Day 2 ($p = .001$).

Participants were more likely to recognize target words than distractor words. To understand if participants were more likely to recognize certain words over others, we examined the recognition frequencies for individual words in the test. There was no clear indication that some types of words were recognized more often than others (e.g., short vs. long words). As might be expected, however, participants were much more likely to recognize words that appeared as targets in the search task over words that appeared as distractors (over 6 times as likely). This also differed by Group: only 1 participant in each of the Illiterate and Semiliterate groups recognized any of the distractor words, whereas all 4 Literate participants recognized at least 1 of these words on Day 4.

Task Completion Time and Errors
A decrease in task completion time occurred over the four days, but this differed by Group (interaction effect: $F_{6,27} = $
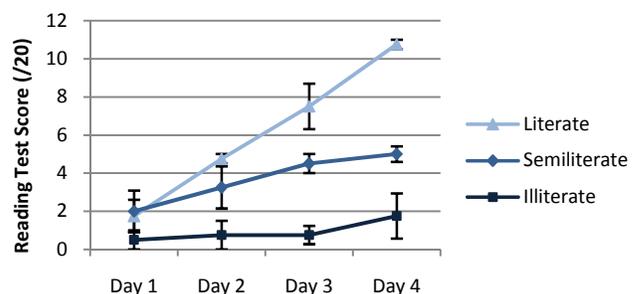

Figure 4. Learning effect on reading test score ($N = 12$).

2.65, $p = .037$, $\eta^2 = .371$). This is shown in Figure 5. While all participants improved in terms of task completion time, the Semiliterate and Literate participants improved faster than the Illiterate participants. Pairwise comparisons showed that Illiterate participants only significantly improved between Day 1 and Day 4 ($p = .037$), whereas both Semiliterate and Literate participants showed improvement already by Day 2 ($p = .004$ and $p = .043$, respectively). Semiliterate participants continued to improve between Day 2 and Day 3 ($p = .009$).

A trend suggested that error rates may have decreased over the course of the study (main effect of Day: $F_{3,27} = 2.85$, $p = .056$, $\eta^2 = .240$), but no other significant main or interaction effects were found on error rate. On average, the number of errors dropped from 16.3 errors on Day 1 to 5.1 errors on Day 4.

Overall Audio Invocations
As participants gained experience with the task they used less audio (main effect of Day: $F_{12.8,11.6} = 19.3$, $p = .001$, $\eta^2 = .682$). On the first day, participants needed the most audio support compared to the other three days ($p < .011$ compared to all other days), but by the last day audio use had dropped considerably, and was lower than each of the preceding days ($p < .040$ for all comparisons). Audio use also differed by group (main effect of Group: $F_{2,9} = 15.5$, $p = .001$, $\eta^2 = .775$). The Illiterate group relied more on the audio support than either the Semiliterate or Literate groups ($p = .032$ and $p = .001$, respectively). No significant interaction effect was found.

The decrease in audio use across Day may also have differed by Group but this was only mildly significant (interaction effect: $F_{2.57,11.6} = 3.31$, $p = .064$, $\eta^2 = .424$).
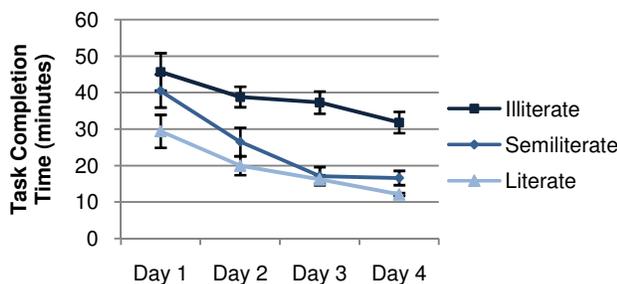
While we did not perform pairwise comparisons, the descriptive data in Figure 6 suggests that this could be due to a lack of improvement or at least a longer learning period before improvement for the Illiterate group.

Breakdown of Audio Use
The overall audio invocations measure encapsulates a range of audio usage, so we performed a secondary analysis to isolate specific types of usage. Since this analysis is a breakdown of the overall audio data it should be interpreted more cautiously; however, it does provide a first step in understanding different types of usage.

Non-target audio use. We first examined audio invocations on distractor words, those that were not the search target for a given trial. As shown in Figure 7, audio use on distractor words differed by Group and showed a main learning effect across Day (main effect of Day: $F_{1.32,11.9} = 18.5$, $p = .001$, $\eta^2 = .673$; main effect of Group: $F_{2,9} = 16.5$, $p = .001$, $\eta^2 = .993$). Semiliterate and Literate participants used less audio than Illiterate participants ($p = .028$ and $p = .001$, respectively). Similar to overall audio usage, a trend also suggested that the learning effect across Day may be differ based on Group (interaction effect: $F_{2.64,11.9} = 3.37$, $p = .060$, $\eta^2 = .428$).

Target audio use. Audio invocations on the target search word for a given trial showed a different pattern from non-target words (Figure 8). While there was a significant reduction in audio use across Day (main effect of Day: $F_{1.53,13.8} = 6.75$, $p = .013$, $\eta^2 = .429$), no other significant main or interaction effects were found. In general, these numbers are more uniform than the non-target audio invocations, and this may be due to *checking behavior*: most of the Semiliterate participants and even the Literate
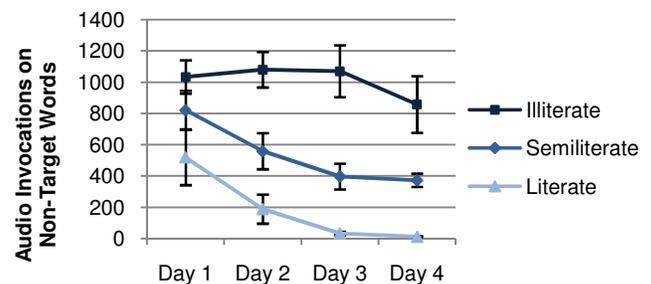

Figure 5. Task completion times across Day ($N = 12$).

Figure 7. Audio invocations on non-target words ($N = 12$).

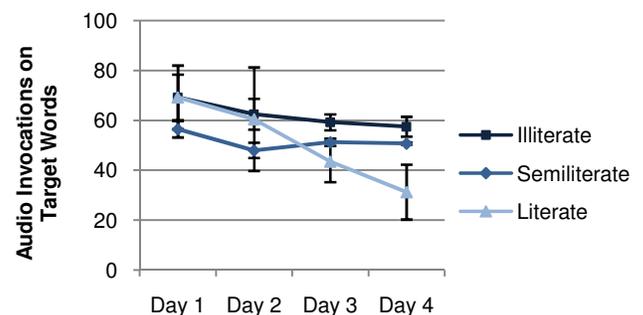**Figure 6. Total audio invocations by day ($N = 12$).**

Figure 8. Audio invocations on target words ($N = 12$).

participants (except on Day 4) used the audio to double-check that they had found the right word, even when they were confident enough to go directly to that word; checking on every trial would account for 50 audio invocations on a given day.

Other audio use. During earlier sessions, participants were more likely to repeatedly play the same audio clip more than once during a trial. The average number of repeated audio clips dropped from 42.9 (SD = 21.6) on Day 1 to 10.9 (SD = 10.3) on Day 4. This may have been due to a general learning effect as participants became more familiar with the words, but may also have been a result of participants adopting more efficient (non-random) search strategies. Participants could also replay the audio task prompt (the word they were meant to search for), and the number of times they did so dropped as they gained more experience with the task.

Finally, participants were more likely to play the audio clips for longer words. There was a significant positive correlation between the length of a word and the number of audio invocations on that word (r^2 = .663, p < .001). This suggests that the shorter words were easier to visually recognize, thus not requiring as much audio support.

Observations and Participant Feedback

We observed several search strategies to complete the task. The most basic was an exhaustive search (often linear) of the audio clips, with no indication that the participant was making any use of the visual cues. This strategy was always used by 3 of the Illiterate participants. The remaining participants reported using strategies with different levels of sophistication: word length, starting/ending syllables, relating the visual pattern to a memory cue (like an emotion evoked by the sound), identifying letters, and over time identifying script modifiers. A representative quote for the Semiliterate group can be seen in one participant's description of her learning process: *"The first day I did not know how to recognize the words, but later on I found out about long words and short words and I would find the words with few [attempts]"* (Participant 5).

When given the option of future Kannada interfaces with sound only, sound and text, or text only, 3 out of 4 Illiterate participants wanted sound only. The only Illiterate participant who preferred sound and text was also the only one in that group who showed a marked decrease in audio use over the four days (see the case of Participant 1, below). In contrast, 7 out of the 8 remaining participants preferred a combination of sound and text. Only 1 Literate participant reported that she would prefer a combination of sound and images without text.

Both the Illiterate participant who showed reading improvement and several of the Semiliterate participants felt that they would see more benefit if they were able to continue the study for longer. Interestingly, the Literate participant who preferred a sound and images option with no text also expressed strong enthusiasm for how the experiment had inspired her to study Kannada: *"I thought I might take the time to learn after this! You need a reason to learn the script, like in English there are lots of things to learn. Now I know some characters and it will make it easier"* (Participant 12). This comment also reflects the lack of incidental language learning opportunities, an important element of language learning as discussed in the Background section.

Participant 1 in the Illiterate group provides a compelling example that audio-augmented text can be useful even for some individuals with no formal education. He was 23 years old and, although illiterate, still had relatively high exposure to computers: he had previously played first person shooters and car racing games in internet cafes at least a few times a month, for a cost of 15 rupees (US$0.38) per hour. He was the only participant in the Illiterate group who showed substantial reduction in audio use and he more than doubled his score on the word recognition test from Day 1 to Day 4. On Day 3 he reported that his search strategy had changed because he was beginning to recognize the words. When asked if he felt he had gained anything from participating in the study he felt that if he could participate over a month he would be able to read and write the words. He clearly benefited from the interaction in a way that would not have been possible with a text-free user interface.

Summary of Results

Both the Literate and Semiliterate participants quickly showed improvement on the reading test after only 2 and 3 days of use, respectively. However, we found no significant improvement for Illiterate participants over the length of the study. In terms of task completion time, Literate and Semiliterate participants improved more quickly than Illiterate participants. The Illiterate group also made the most use of audio. Finally, as expected based on Study 1, there was an overall decrease in audio use over the four days.

DISCUSSION

Illiterate and semiliterate users have differing needs and respond differently to text in the interface. Our results show that text augmented with audio has benefits for semiliterate users: reliance on audio support decreases over time and this corresponds to longer-term evidence of an ability to visually recognize words, even when the visual learning was incidental to the main task. These results are encouraging because they suggest that an augmented text interface provides an incidental learning opportunity and therefore reinforcement of reading skills in semiliterate individuals. Semiliterate individuals can also take advantage of the performance benefits of text over audio (listening to audio takes longer than reading equivalent text) and can benefit from the unambiguous visual representation it offers (especially for abstract words and phrases). Of the 16 semiliterate participants in both studies, a large majority

of them (13) preferred to have both text and sound in the interface rather than only having one of the modalities.

To further support the benefit of working in a text interface augmented with audio, reflection by participants on their search strategies matched the literature on the stages of reading skill acquisition. The reported use of visual cues such as word length, followed by learning of syllables, then finally sub-syllabic information (by two of the literate participants) clearly demonstrate the stages of reading skill acquisition and suggest that working in audio-augmented text interfaces can result in the side benefit of incidental language learning. Although there are differences in implementation, context and exposure, our findings confirm those of Kothari et al.'s [14] same language subtitling work.

Based on both the performance and preference results from our studies, user interface designers should consider the use of audio-augmented text interfaces for semiliterate individuals. For designers, augmenting an existing text-based interface with audio should require much less effort than creating a new text-free version of the interface. When designing minimal-text interfaces there are several visual alternatives to text, such as drawings, photographs and animated images, and it is unclear which option is best [15]. Text-to-speech screen reader systems such as JAWS [11] already provide speech-based interaction with existing applications for blind and low vision users. In contrast, the problem of augmenting the interface with audio for sighted users should be more straightforward because the designer does not need to provide non-visual navigation methods.

We have shown that augmented text interfaces can be beneficial for semiliterate users but we did not find a similar benefit for fully illiterate participants: 3 of the 4 illiterate participants in Study 2 appeared to entirely ignore the text and to focus solely on the audio. Although illiterate individuals may also benefit from redundant text and audio in the interface over a longer time period, our results do not dispute previous work by Medhi et al. [15, 16] on the importance of text-free and minimal-text user interfaces for fully illiterate individuals. These findings highlight the importance of considering the exact makeup of the target population before choosing a text-free versus an augmented, text-based user interface. Since both illiterate and semiliterate target populations have low levels of education, interaction complexity needs to be considered in addition to the choice of whether or not to use text. Finally, individual differences in skills and background, even for users with no formal education (e.g., the one illiterate participant who showed reduction in audio use) should also be considered. One option, for example, is to allow for text to be shown or hidden on demand.

In our experimental system we used a one-to-one mapping from text to audio. Another option is to include tooltip-like audio which would provide a more descriptive alternative to the text. However, the downside of this option is that it will not as effectively support the phonemic cue stage of learning to read (mapping from specific symbols to sounds), so the incidental learning benefit observed in our study may not be as prominent. If maximizing the opportunity for reading skill acquisition is a main design goal, then a karaoke-style approach to simultaneously highlight syllables as words are read aloud (used in same language subtitling [14]) would provide more benefit.

Limitations of the Experiment

The research presented here offers a first step towards understanding how to design interfaces for semiliterate users and how this differs from illiterate users. One limitation, however, is that we did not compare the combination of text and audio to either a text-only or an audio-only condition. For our target users a text-only condition would be an extremely difficult if not impossible option, likely resulting in user frustration and high error rates. On the other hand, an audio-only version could have functioned as a control for general learning effects such as confidence and the development of efficient non-visual search strategies (e.g., linear vs. random search). Performance of the illiterate participants in Study 2 should offer at least some insight into potential performance with an audio only condition: on the first day illiterate and semiliterate participants were roughly equal in terms of task completion time, but by the fourth day semiliterate participants were approximately twice as fast as illiterate participants. This suggests that an audio-only version would be significantly slower than a text and audio version for semiliterate participants over long-term use.

Recruiting and running studies with low literacy individuals in a developing world context presents unique challenges [5]. Our illiterate and semiliterate participants worked extremely long hours and had very little free time for extra activities, so it was a challenge to recruit 8 participants to attend sessions for four subsequent days. Additionally, we did not tell participants ahead of time that they would be receiving a gift at the end, because we did not want this to be the motivating factor for their participation. As a result of these practical constraints we only had four participants in each literacy group in Study 2. It will be important for future work to build on and reinforce our findings via studies with larger numbers of participants.

CONCLUSION

We have shown through two controlled experiments that as semiliterate users become more familiar with a text-based interface augmented with audio, they become less reliant on the audio, a change that likely contributes to the corresponding performance improvement we observed. The second study further showed that semiliterate individuals and those literate in other languages improved on a paper-based word recognition test administered at the end of each study session, as quickly as the second day of use. For fully illiterate users, however, our findings add further support to the growing body of evidence indicating that little or no text should appear in the interface [15, 16]. Combined, these

findings suggest that the needs of semiliterate and illiterate users should be treated differently by designers.

The goal of this research was not to teach users how to read. Rather, we were interested in understanding how users with different levels of literacy would choose to use the text and audio representations in the context of completing a timed task. The next step will be to replicate these results in a field study, with an application such as Medhi et al.'s [15] employment search engine. Further work also remains on how to most effectively augment a text-based interface with audio and, if desired, how to more explicitly support users in transitioning from audio to text. Individual differences likely play an important role, in which case options such as showing or hiding text on demand may be appropriate.

ACKNOWLEDGMENTS

REFERENCES

1. Alvarez, J.M.M. Trust building user generated interfaces for illiterate people. In *Proc. IUI4DR 2008*, (2008), 38-42.

2. Barr, R., Pearson, P.D., Kamil, M.L., Mosenthal, P.B. (Eds.). *Handbook of Reading Research: Volume II.* Lawrence Erlbaum, 1996.

3. Blake, E.H., Steventon, L., Edge, J., Foster, A. A field computer for animal trackers. In *Proc. CHI-SA 2001*, (2001). Retrieved January 15, 2008, from www.chi-sa.org.za/CHI-SA2001/cyber-tracker.pdf

4. Brett, P. Using multimedia: A descriptive investigation of incidental language learning. *Computer Assisted Language Learning 11*, 2(1998), 179-200.

5. Brewer, E., Demmer, M., Ho, M., Honicky, R.J., Pal, J., Plauché, M., Sonesh, S. The challenges of IT research for developing regions. *IEEE Transactions 5*, 2(2006), 15-23.

6. *Census of India – India at a glance: Number of literates & literacy rates.* Retrieved April 30, 2008, from http://www.censusindia.gov.in/Census_Data_2001/India_at_glance/literates1.aspx

7. Central Intelligence Agency. *The World Factbook.* Retrieved May 7, 2008 from https://www.cia.gov-/library/publications/the-world-factbook/geos/xx.html

8. Chall, J. *Learning to read: The great debate.* New York: McGraw-Hill. 1983.

9. Danan, M. Reversed subtitling and dual coding theory: New directions for foreign language instruction. *Language Learning 42*, 4(2006), 497-527.

10. Dhakhwa, S., Hall, P.A.V., Ghimire, G.B., Manandhar, P., Thapa, I. Sambad – Computer interfaces for non-literates. In *Proc. HCII 2007*, (2007), 721-730.

11. Freedom Scientific. *JAWS for Windows Overview.* Retrieved September 2, 2008, from http://www.-freedomscientific.com/fs_products/software_jaws.asp

12. Grisedale, S., Graves, M., Grunsteidl, A. Designing a graphical user interface for healthcare workers in rural India. In *Proc. CHI '97*, (1997), 471-478.

13. Hiebert, E.H., Kamil, M.L. (Eds.). *Teaching and Learning Vocabulary.* Lawrence Erlbaum, 2005.

14. Kothari, B., Takeda, J., Joshi, A., Pandey, A. Same language subtitling: A butterfly for literacy? *International Journal of Lifelong Education 21*, 1(2002), 55-66.

15. Medhi, I., Prasad, A., Toyama, K. Optimal audio-visual representations for illiterate users of computers. In *Proc. WWW 2007*, (2007), 873-882.

16. Medhi, I., Sagar, A., Toyama, K. Text-free user interfaces for illiterate and semiliterate users. In *Proc. ICTD 2006*, (2006), 72-82.

17. Moffatt, K., McGrenere, J., Purves, B., Klawe, M. The participatory design of a sound and image enhanced daily planner for people with aphasia. In *Proc. CHI 2004*, (2004), 407-414.

18. *Languages spoken by more than 10 million people.* MSN Encarta. Retrieved September 1, 2008, from http://encarta.msn.com/media_701500404/Languages_Spoken_by_More_Than_10_Million_People.html

19. Nag, S. Early reading in Kannada: The pace of acquisition of orthographic knowledge and phonemic awareness. *Journal of Research in Reading 30*, 1(2007), 7-22.

20. Parikh, T., Ghosh, K., Chavan, A. Design studies for a financial management system for micro-credit groups in rural India. In *Proc. CUU 2003*, (2003), 15-22.

21. Parikh, T., Javid, P., Kumar, S., Ghosh, K., Toyama, K. Mobile phones and paper documents: Evaluating a new approach for capturing microfinance data in rural India. In *Proc. CHI 2006*, (2006), 551-560.

22. Plauché, M., Prabaker, M. Tamil market: A spoken dialog system for rural India. In *CHI Extended Abstracts 2006*, (2006), 1619-1624.

23. Prasad, A., Medhi, I., Toyama, K., Balakrishnan, R. Exploring the feasibility of video mail for illiterate users. In *Proc. AVI 2008*, (2008), 103-110.

24. Shakeel, H., Best, M.L. Community Knowledge Sharing: An internet application to support communications across literacy levels. In *Proc. Technology and Society*, (2002), 37-44.

25. Sherwani, J., Ali, N., Mirza, S., Fatma, A., Memom, Y., Karim, M., Tongia, R., Rosefeld, R. HealthLine: Speech-based access to health information by low-literate users. In *Proc. ICTD 2007*, (2007).

26. Sin, M., Escobedo, M., Best, M. A directory service for multi-literate users. In *Proc. ICME 2004*, (2004), 1295-1298.