

Guiding the Selection of Child Spellchecker Suggestions using Audio and Visual Cues

Brody Downs
Boise State University
Boise, Idaho, USA
brodydowns@u.boisestate.edu

Aprajita Shukla
Boise State University
Boise, Idaho, USA
aprajitashukla@u.boisestate.edu

Mikey Krentz
Boise State University
Boise, Idaho, USA
mikeykrentz@u.boisestate.edu

Maria Soledad Pera
Boise State University
Boise, Idaho, USA
solepera@boisestate.edu

Katherine Landau Wright
Boise State University
Boise, Idaho, USA
katherinewright@boisestate.edu

Casey Kennington
Boise State University
Boise, Idaho, USA
caseykennington@boisestate.edu

Jerry Fails
Boise State University
Boise, Idaho, USA
jerryfails@boisestate.edu

ABSTRACT

Spellchecking functionality embedded in existing search tools can assist children by offering a list of spelling alternatives when a spelling error is detected. Unfortunately, children tend to generally select the first alternative when presented with a list of options, as opposed to the one that matches their intent. In this paper, we describe a study we conducted with 191 children ages 6-12 in order to offer empirical evidence of: (1) their selection habits when identifying spelling suggestions that match the word they meant to type, and (2) the degree of influence multimodal cues, i.e., synthesized speech and images, have in prompting children to select the correct spelling suggestion. The results from our study reveal that multimodal cues, primarily synthesized speech, have a positive impact on the children's ability to identify their intended word from a list of spelling suggestions.

Author Keywords

Spelling; multimedia cues; search engine; user studies.

CCS Concepts

•**Social and professional topics** → **Children**; •**Human-centered computing** → *User studies*; Text input; •**Information systems** → Information retrieval query processing;

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

IDC '20 June 21–24, 2020, London, United Kingdom

© 2020 ACM. ISBN 978-1-4503-7981-6/20/06... 15.00

DOI: <http://dx.doi.org/10.1145/3392063.3394390>

INTRODUCTION

The use of search tools, including popular search engines like Bing or Google, for online information discovery has become a valuable asset for children [2, 29, 34]. Unfortunately, children continue to struggle during their interactions with search engines in their quest for appropriate and relevant information [32]. The primary starting point for information discovery tasks, is typed queries, which frequently contain spelling errors. This is a concern given that the use of correct terminology is a crucial step in formulating queries that a search engine can effectively process [26]. In other words, queries containing spelling errors can lead search engines to retrieve resources that do not accurately respond to the information needs of users; in fact, these misspelled queries can yield no results at all [14]. Given that misspellings are common in a web search environment [12], even more so among children [10, 8], the need to support young users with formulating queries free of misspellings becomes a must.

Information Retrieval and Natural Language Processing researchers, among others, have allocated research efforts to the design and deployment of strategies for spelling corrections targeting web search environments [15, 24]. These strategies, however, overlook the fact that child misspellings often differ from those of an adult [8]. Outcomes from a recent empirical study indeed demonstrate that a spellchecker tailored to the orthographic development of children can generate spelling suggestions that better relate to children's intended formulation [7]. The authors did note that when using spellchecking to aid in formulating queries, children do not always select the correct spelling suggestion (for example, suggestions as depicted in Figure 1a). More specifically, when children are presented with suggested spelling corrections the word children click on in order to fix the misspelling does not always match

the word they intended to type. When presented with spelling suggestions, children tend to select the first option, even if it is not the word that matched their intent [7]. This aligns with behavior seen in young children where they gravitate towards top-ranked alternatives when presented with search engine results pages [25] or query suggestions [1].

Given that different words can be used to express a similar information need, it is not always possible for a spellchecking strategy to position the “correct” spelling at the top of the spelling options. As possible alternatives that can enhance the design of spellchecker interfaces, Druin et al. [10], who explored children’s interaction with search interfaces, advocated for interactive spelling assistance as well as the use of different forms of input for search interfaces (e.g. images) that could help with *visual searchers* (a search role they defined that is characterized by children who prefer to search with a visual context). For those who are visually impaired, or just do not prefer the visual context, audio descriptions have been shown to be a useful guide for children’s visual attention [30]. In general, the less a child knows about the domain of search, the more likely the child is to need recognition aids to prompt his or her memory and develop an appropriate search strategy [6]. Prior research leaves unresolved the question of how to effectively aid children in choosing the word they meant to type from a list of spelling suggestions.

Inspired by the lessons learned that were reported in [7, 10], in this paper we discuss the results of an empirical analysis we conducted to investigate the influence of **multimodal cues** (i.e., images and audio playback) in assisting children (ages 6-12) when selecting suggested spelling alternatives. We use as a starting point the spelling suggestions generated by a child-oriented spellchecker,¹ known to better respond to children’s spelling strategies, and enhance its functionality by incorporating multimodal cues in order to investigate how these cues can help children effectively select the spelling suggestion that meets their original intent. In order to determine the effectiveness of the aforementioned multimodal cues, we implemented a protocol that uses fact-based and open-ended search tasks; these enable us to generate query logs capturing children’s interactions with a search tool offering enhanced (i.e., multimodal) spellchecking functionality. Child participants (n=191), in a between-subject experimental design, used a search tool with four conditions with varying media cues accompanying the spellchecking suggestions: no cues, audio, image, or both audio and image.

Quantitative analysis of query logs, in addition to observations, led to the results and lessons learned which are reported below. In short, we found that there is benefit from the use of different multimodal cues to help children when choosing the intended spelling correction. Additionally, with the use of multimodal cues, we observed that children were more likely to explore their options rather than resort to the first available alternative.

¹We chose this spellchecker, as it is the state-of-the-art for children; yet, any spellchecker could have been used instead. In other words, spellchecker implementation is independent of the research presented in this paper. We will make available a link to the algorithmic implementation of the kid-friendly spellchecker upon publication.

The primary contribution of this work is a comparison of the effectiveness of media cues in a spellchecking interface for children, which provides in-sights on the limitations of existing tools when it comes to handling children’s selection of spelling suggestions.

The rest of our paper is organized as follows: We first discuss related work. This is followed by a description of our method, including the protocol for data collection, baseline, and selection of the synthesized voice required for our audio condition. Thereafter, we present experiments along with an in-depth discussion of results and limitations. We offer at the end some concluding remarks and directions for future work.

RELATED WORK

In this section, we discuss literature pertaining to search engines, spelling, and search interfaces, which offer context for our work.

Search Engines and Spelling Correction. To complete a search task, formulation of a well-defined query is crucial [11]. Since correct spelling contributes to query formulation and the quality of the results, it is imperative to assist the user when they misspell a word. Some search engines have in-built automatic spelling correction feature as spelling mistakes are very common in user-generated text. Many websites have a feature of automatically giving correct suggestions to the misspelled user queries in the form of “*Did you mean?*” suggestions or automatic corrections [12]. Unfortunately, these built-in features are known to be better suited for adult misspellings only.

Children, Search Engines, and Spelling. There are numerous documented barriers when it comes to children interacting with search engines, including their limited vocabulary, lack of knowledge of query formulation or the use of Boolean logic [1, 11, 4]. They emerge from the fact that primary school age children are known to (i) read slowly and are still learning to write [18, 21], (ii) have limited domain knowledge [28], and (iii) experience difficulties typing on a keyboard [36]. These reasons make searching a tedious task and can result in problems with query formulation including spelling errors [5, 19]. As an attempt to help with this concern, the authors in [7] introduced *Kidspell*, an English based spellchecker suited to children’s orthographic development. It emerged from the results of their experiments that despite having spelling suggestions that capture children’s needs better than popular spellcheckers, such as Aspell² or Bing³ Spell Check, most children clicked on the first suggestion even if it was not the most relevant one (i.e., the correct spelling of the intended word was lower in the list). This is what inspired our exploration of incorporating different modalities that could draw attention to the many spelling alternatives spellcheckers tend to offer given a misspelled word.

Search Engine Interfaces and Multimodal Cues. In a pilot study, Bilal investigated the searching behavior and success of 22 seventh grade science students in using the Yahoo!igans! search engine to find information on a research task [4], but

²<http://aspell.net/>

³<https://tinyurl.com/AzureSpellcheck>

children failed in their quest mainly due to their lack of knowledge of how to use the engine. Similarly, Gossen studied several search interfaces designed for children, and concluded that that visual cues were helpful for formulating queries [16]. Zha et al. showed that displaying images relevant to the search query improved query suggestions and query formulation, but their study was limited to adult users only [39]. Gossen et al. studied various information presentation strategies and tested them to see if children selected the correct resource from a group of relevant resources, but in that paper the focus was on making the information more engaging rather than helping with effective search task completion [20].

Moreover, considering audio cues, the authors in [31] explored the degree to which the use of a vocal assistant as an intermediary between a child and a search engine can ease query formulation. An evaluation of a computer-based Picture Exchange Communication System approach investigated how the virtual tutor’s voice influences the children’s participation and performance [27]. Furthermore, young non-verbal children with autism respond to an on-screen “virtual tutor” through the manipulation of picture/symbol cards. Analysis of the children’s responses suggest that they were able to use the system more effectively when the virtual tutor had a synthetic voice, rather than a human voice [27].

These studies suggest that search engines need to be intuitive and engaging for adults and children. Below we highlight how visual and audio cues and lead to improved engagement—two modalities which we use in our work reported here. These findings also align with established theories of learning, particularly Dual Coding Theory which posits that providing information in multiple modalities (e.g., text, images, and audio) aids readers’ comprehension [35]. Therefore, we take inspiration from these works to make use of multimodal cues (i.e., images and audio) to assist children in making correct spelling suggestion selections.

METHOD

In this section, we explain the protocol used in the experiments we conducted to investigate the effects of multimodal cues on children’s ability to find their intended word among a list of spelling suggestions as well as their selection tendencies when interacting with a spellchecker.

Protocol

In order to answer our research question for identifying which multimodal cues, if any, would better guide children’s interactions with spellcheckers, we adapted the protocol defined in [31], which establishes four dimensions for conducting experiments related to search tools: (1) task, (2) user group, (3) strategy, and (4) environment, which, taken together, afford us the ability to systematically compare across experiments. In our case, we specify each of these four areas as follows:

1. **Task.** In order to elicit search queries from children, facilitators gave verbal prompts to serve as a starting point for a typical online search query. For this reason we relied on two types of prompts: *fact-based*, which are less complex and require children to locate specific and quick answers, as well

as *open-ended* prompts, which require more in-depth consideration of the content of the search results. See sample questions in Table 1.

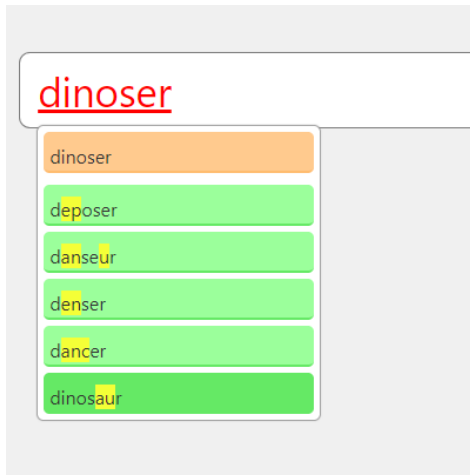
2. **User group.** Participants were children ages 6 to 12. We selected this age group to represent children who have likely developed the basic phonetic skills needed to attempt spelling, but have yet to obtain advanced orthographic skills [3].
3. **Strategy.** We make use of a custom search tool targeting children. Our custom search tool serves as a search interface for submitting queries and provides search results via Google’s web API with safe-search functionality enabled. In addition, the search tool provides a spellchecking utility. During query formulation, when a spelling error is identified, it will be underlined and colored in red. Hovering over the spelling error will provide a list of up to 5 spelling suggestions. The settings can be configured to utilize one of many spelling suggestion algorithms (e.g. Aspell, Bing Spell Check, KidSpell). Dependent on the settings, as seen in Figure 1, hovering over a suggestion can display a relevant image and/or read the word aloud using a speech synthesizer. Clicking on a spelling suggestion replaces the spelling error with the suggested spelling. All user interactions with the interface are automatically recorded.
4. **Environment.** Search tasks were performed by children at local STEM events hosted at three local venues – two elementary schools and a local community building. The stem event held at the community building was organized by state government agencies where children were bussed in from their respective schools and participated as an informal education (i.e., a fieldtrip) experience. Each event contained multiple STEM-related booths and hands-on activities. Children formed their queries based on verbal search prompts assigned by a facilitator (i.e., a member of the research team). Facilitators observed and recorded interactions children made with the search interface and the interface logged children’s queries and interactions with the spellchecker.

Search Task	Type
Fact-based	Who was the first computer programmer?
	What is the state bird of [ANON]?
	How far away is the Sun?
	How tall is a tyrannosaurus rex?
Open-ended	Find me a cool fact about space.
	Find me a difference between Earth and Mars.
	Find me an interesting fact about Albert Einstein.
	Find me a fact about your favorite dinosaur.

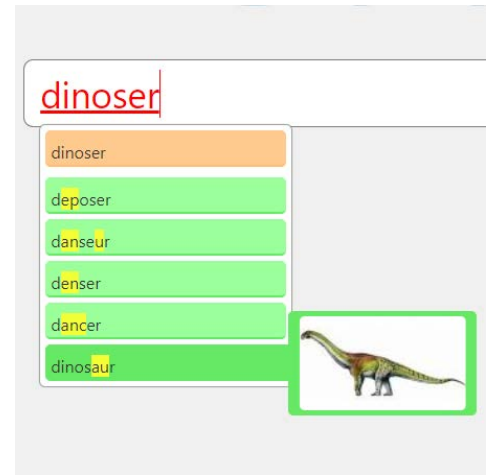
Table 1. Sample tasks prompts presented to study participants, grouped by type.

PRELIMINARY SETUP: ESTABLISHING A BASELINE AND SYNTHESIZED VOICES

In this section we discuss the necessary preliminary setup performed in order to effectively test each of the experimental conditions. To establish a baseline, we setup a condition in which children completed the tasks specified above using



(a) Traditional spellchecking functionality with no visual aid (control and audio playback conditions).



(b) Spellchecking functionality enhanced with images (image and multi-modal cue conditions).

Figure 1. Figures depicting spellchecking functionality offered to children interacting with the custom search tool. Both of the visual aids shown can be accompanied by audio playback of the hovered suggestion.

an interface that when a word was misspelled would allow children to choose from spelling suggestions with no audio or image cues to assist them. The results from this baseline condition serve as a measure for the effectiveness of each of the multimodal cues explored later.

As an additional preliminary setup to the audio cue condition, in which spelling suggestions will be read aloud to users via speech synthesis, we setup an experiment to see which (if any) speech synthesis voices were more suitable than others when considering children as the user base.

Baseline Condition: No Cues

This baseline study was performed at the STEM event of a local elementary school with 51 child participants (ages 6-12). In this particular instance, the search system offered basic spellchecking functionality (i.e., no images or audio to assist children in their spelling suggestion selections). Given the user group, we made use of *Aspell*, a traditional spellchecker, and *KidSpell*, a spellchecker that is explicitly designed to respond to the spelling needs of children [7], with the intent to determine which spellchecker was more suited towards children’s information needs.

KidSpell is a phonetic-based, English spellchecker, designed for the spelling needs of children by addressing the phonological strategies children use to spell. A phonetic similarity technique is used to map words to a phonetic form which is used to find words that are similarly pronounced. For example, the spelling error *afishil* and the correct spelling *official* would map the same phonetic form, making *official* a suitable spelling suggestion despite having few letters in common. Preliminary studies from this experiment have demonstrated that *KidSpell* outperforms traditional spellcheckers, such as *Aspell*, when correcting spelling errors made by children in online search queries [7]. As such, *KidSpell* is solely used in following experiments.

As part of our study to examine children’s ability to find their intended word among a list of spelling suggestions, we observed that despite being given the correct word as a spelling suggestion (meaning it appeared in the list of 5 suggestions provided), children chose their intended word just 68% of the time. In the 84 queries formulated by 51 children, they made an attempt to correct their spelling by clicking on a spelling suggestion 56 times. Out of the 47 selections where children clicked on a suggestion where their intended word was available, they clicked on their intended word just 32 times.

We further examined the position of the suggestions children clicked along with how well those selections matched their intent. Most children selected the word at the first position in the baseline condition when there were no cues (as seen in Table 2). In most cases, this selection matched their intended word. However, there were still several cases when clicking on the word in the first position was not correct. In fact, the number of times children clicked on the first word incorrectly was more than double the times the children clicked incorrectly on words in any other position, as noted in Table 3. Additionally, in the nine scenarios where the intended word was not available on the list of suggestions, children tended to click on the higher ranked suggestions.

These results suggest that children frequently selected the word in the first position regardless of whether it matched their intended word or not, which resulted in many mistakes. The tendency of children to resort to the first available option is analogous with children’s selection tendencies observed in related research [25].

Despite advances in spellcheckers for children, we cannot expect that spellcheckers will be able to fully recover the intent of the child as the first suggestion, which is one reason why spellcheckers provide multiple suggestions. This led to the research question and primary contribution of this paper

K	No Cues			Audio Only			Images Only			Both Audio & Images		
	Clicks	Correct	%	Clicks	Correct	%	Clicks	Correct	%	Clicks	Correct	%
1	28	21	.75	16	16	1.0	16	12	.75	12	9	.75
2	4	2	.50	12	10	.83	13	12	.92	14	13	.93
3	4	3	.75	19	18	.95	7	5	.71	16	14	.88
4	8	5	.63	9	9	1.0	7	7	1.0	21	19	.90
5	3	1	.33	9	7	.78	8	8	1.0	9	9	1.0
Total	47	32	.68	65	60	.92	51	44	.86	72	64	.89

Table 2. Analysis of spelling suggestions. *Clicks* are the number of suggestions clicked at position K; *Correct* is the number of clicks that correctly matched the child’s intent and % is the proportion of clicks that matched their intent.

K	No Cues	Audio	Image	Both
1	7	0	4	3
2	2	2	1	1
3	1	1	2	2
4	2	0	0	2
5	1	2	0	0

Table 3. Position location of spelling suggestion selections when making an incorrect selection.

as to the affect cues can have on helping children select the spelling of words that best match their intent.

Selecting the Speech Synthesis Voice

As part of the setup to the conditions that involved audio playback, it was necessary to select a single synthesized voice to be used as a constant across all experimental conditions. We noticed that speech synthesis systems utter misspelled words with high fidelity to the the misspellings, making them potentially useful for aiding children who can hear how any combination of letters (i.e., non-words) might be pronounced. Given the various choices of speech synthesis voices, we first conducted a preliminary study in order to determine if any speech synthesis voices were more viable than others in terms of having clearer pronunciation or their ability to be understood by children. This study was performed with 8 children aged 6-11 years. The child participants are members of an inter-generational design team that meets in a research lab twice a week [9, 13, 23].

Four American-accented voices were selected for comparison, each from Amazon’s speech synthesis service Polly⁴: Ivy (child female), Joanna (adult female), Justin (child male), and Matthew (adult male). The adult voices were chosen on the basis of having natural speech and neutral accents when compared to the other adult options. The child voices were chosen as options that could be perhaps be more relatable to children. Each of the four voices were used to make recordings of select sentences, examples of which can be seen in Table 4. The sentences were directly chosen from (or modeled after) phrases used in a research-based assessment of children’s ability to identify different phonemes in spoken phrases [37].

⁴Amazon Polly: <https://aws.amazon.com/polly/>

Correct Sentence	Incorrect Sentence
This is my favorite book	This is my <i>favtit</i> book
Throw me the ball	<i>Thow</i> me the ball
Robots have circuits	<i>Robets</i> have circuits
Make a design	Make a <i>desine</i>

Table 4. Examples of sentences used for speech synthesis recordings in the voice selection preliminary study.

Two recordings were created for each sentence: a correct version of the sentence and an incorrect version where one word was altered using a spelling error created by children found in previous research [7]. Using each of the four voices, we tabulated the children’s ability to detect a difference between the two sentences, identify which of the spoken sentences was incorrect, and identify the word that was pronounced incorrectly, the results of which can be seen in Table 5. Although, the *Joanna* voice performed best on each metric, the results were not statistically significant as per McNemar’s statistical significance test. However, given that the use of adult, female, American speech synthesis voices in other existing work showed promise [33, 38], we chose to use the *Joanna* voice for the conditions discussed in the following sections.

	Ivy	Joanna	Justin	Matthew
Detect Difference	0.930	0.972	0.945	0.918
Identify Sentence	0.845	0.877	0.863	0.863
Identify Word	0.915	0.959	0.904	0.877

Table 5. Proportion of participants that were able to detect a difference between the two sentences, identify which sentence was incorrect, and identify which word was incorrect for each speech synthesis voice.

EXPERIMENT

In this section, we describe the procedure as well as the results obtained from each of the three experimental conditions (audio-only, image-only, and both audio and image). The experiments included input from 140 children who performed a total of 297 queries at STEM events located at an elementary school and a local community building.

In general, we follow the protocol outlined earlier in the Method section. The experiments described below were between-subject in design. As such, each child only used

one of the three possible interface conditions. There were three computers setup at each event, each computer had a different conditions setup on it. Children were randomly asked to approach a computer, and as such the condition used by each child was chosen randomly and children were not aware of alternative conditions. Additionally, the order of the spelling suggestions that were provided when a word was misspelled was randomized in the three conditions in order to see what effect a random order might have as compared to the baseline where the order of suggestions were ranked by the spelling suggestion algorithms. Children's spelling intent was determined by a consensus of the event facilitators with help from the notes taken by facilitators, query log information, and search prompts given. Interactions with the spellchecker were analyzed for children's selection tendencies as well as their ability to select their intended word.

Condition 1: Audio Playback Cues

This condition follows the protocol described earlier, but makes use of the *speech synthesis option* in the search tool to read spelling suggestions aloud.

Procedure

Similar to the set up of the study we discussed above for our baseline experiment, here children were given verbal prompts to search and provide answers to queries (sample queries in Table 1) while audio cues were provided for misspelled words to help searchers identify their intended word from a list of spelling suggestions. Whenever a user hovered over a given spelling suggestion (see Figure 1a), synthesized speech using the Joanna voice read the suggestion aloud to the child through over-the-ear headphones.

Results

We observed 57 children who generated 115 queries and made 68 attempts to correct their spelling by selecting a spelling suggestion. Out of the 65 correction attempts where their intended word was available as a suggestion, children selected their intended word 60 times. Compared to our baseline experiment, in which children were given no cues, we observed a significant improvement in their ability to find their intended word, improving from just 68% to 92%. This result was found to be statistically significant using the two-proportion z-test ($p < 0.05$).

We further analyzed the selection tendencies of children based on the ranking position of their selected spelling suggestion. With the use of audio queues, children were able to consistently make correct selections regardless of the position the spelling suggestions clicked, as seen in Table 2. Furthermore, children never made an incorrect selection when selecting the first option and, overall, were more willing to explore other options when compared to the baseline experiment. This is a welcome result that emphasizes the importance of assistive cues for children when interacting with spellcheckers. Additionally this aligns with research that shows that children spell using phonological strategies [22]. We infer the use of audio queues helped reinforce the child's known pronunciation of a word and often led them to a correct selection.

Condition 2: Visual Cues

This condition follows the same protocol as the above condition, but uses visual aids to help searchers identify their intended word from a list of spelling suggestions.

Procedure

When a spelling suggestion is hovered over with the mouse, a relevant image is displayed next to it; an example can be seen in Figure 1b. Images are acquired using Google's Image Search API with safe search enabled. The first image available, using the spelling suggestion as the query, is the image chosen to be displayed alongside each spelling suggestion.

Results

We observed 61 children who generated 102 queries and made 55 attempts to correct their spelling by selecting a spelling suggestion. Among the 51 attempts made to correct their spelling where their intended word was an available suggestion, children selected their intended word 44 times. Compared to our baseline experiment, children were more readily able to find their intended word among a list of spelling suggestions, finding that word in 86% of scenarios compared to just 68% in the baseline. This result was found to be statistically significant using the two-proportion z-test ($p < 0.05$). However, the image cues did not perform as well as the audio cues did, in which children found their intended word in 93% of scenarios, as discussed in the previous section.

We also see as a result of examining the interactions that resulted from this experiment that, when making an incorrect selection, children were more likely to select the higher positioned spelling suggestions. Similar to the baseline condition, despite their intended word being located primarily in the first position, children most frequently made incorrect selections at that position. In the 7 attempts where children made an incorrect selection, 4 of the selections came at the first position, 1 at the second position, and 2 at the third position. Although children still often resort to the first available option, we see an improvement to their ability to find the intended word when compared to the baseline experiment.

Condition 3: Multimodal Cues – Audio Playback & Image

This condition follows the same protocol as the previous conditions, but uses both audio and visual cues to help searchers identify their intended word from a list of spelling suggestions (see example *b* in Figure 1). The audio and images are from the same sources as the first two experimental conditions.

Results

We observed 59 children who generated 80 queries and made 73 attempts to correct their spelling by selecting a spelling suggestion. Among the 72 attempts to correct their spelling where their intended word was available as a spelling suggestion, 64 of those selections were the intended spelling. The correct selection rate of 89% is an improvement from the rate of 68% seen with baseline experiment and was found to be statistically significant using the two-proportion z-test ($p < 0.05$). It notably outperforms the image-only cue, which had a correct selection rate of 86%; yet, less than the rate of 92% seen in the audio-only cue.



Figure 2. Figures depicting the images displayed along side spelling suggestions: left, “bird”, and right a more abstract word, “difference”.

Note that although these cues were tested with different queries on different children, the correct selection rate with both image and audio cues was almost exactly the average between the separate audio and image conditions. This suggests to us that the images may have a direct impact on how helpful the audio was in the selection process. Why images did not contribute as much to children’s spelling suggestion selection as audio is an open question, but we offer possible explanations: (1) many spelling mistakes made by children are generally phonetic, yet images do not provide any feedback on the phonetics of a particular word spelling (whereas the audio does), and (2) many of the words that children first learn are *concrete* in that they denote physical things (e.g., *chair* or *ball*), and will likely have corresponding images that represent them, whereas abstract concepts (e.g., *utopia* or *democracy*) do not. We leave a systematic evaluation of concrete vs. abstract spelling errors for future work.

An interesting observation from the examined incorrect selections is that that when selecting an incorrect spelling suggestion, the children tended to choose suggestions from the top of the list (as shown in Table 3). This same behavior can be seen in Condition 2 (visual only cues) and because of this repeatedly observed behavior; we infer that this is not a flaw of the spellchecker or user interface, but perhaps an assumption made by the users that the first item is most likely to be correct.

ANALYSIS AND DISCUSSION

We discuss below general observations and lessons learned that resulted from our experiments.

Use of multimodal cues. We saw significant improvements to children’s ability to find their intended word among a list of spelling suggestions when using either of the multimedia cues or a combination of the two. Although these improvements were shown to be of statistical significance when compared to the baseline experiment (two-proportions z-test; $p < 0.05$), we did not observe a statistical significance when comparing them to each other (two-proportions z-test with Bonferonni correction; $p > 0.016$).

Image acquisition and comprehension. We postulate that a factor in the difference between the two types of cues is the noise that comes along with the technique used for image acquisition. Although, the images used are retrieved through safe search using Google’s safe search API, they may not

make a concrete connection to the spelling suggestion in question. Image quality through Google’s API is inconsistent, and many words have abstract meanings that cannot be easily communicated through an image. For instance, the frequently misspelled word *difference*, can be especially difficult to convey through imagery and result in retrieval of images that are more vague. An example contrasting a vague image shown to denote the word *difference* that is an abstract concept versus an image of a bird which is a concrete concept can be seen in Figure 2. This finding supports previous research indicating that the concreteness of visuals is directly related to how well they aid comprehension [35].

Though comprehension is important in learning words and their corresponding spellings, the audio modality seems to have optimized on phonetics which more directly affects spelling, whereas the images were tied to the spelling through comprehension of referents.

Ranking. We further analyzed the selection tendencies of children. In the only audio cue playback condition, the most frequent incorrect selections came from words positioned last in the list of spelling suggestions, which was notably different from other experiments, namely, no cues, only image cues and both image and audio cues.

Children were willing to explore options in the audio cue experiment which was evidenced by their tendency to listen to audio suggestions and their minimal incorrect clicks on the first suggestion. However, if they did not find a suitable suggestion by the time they reached the end of the list, they resorted to the last available suggestion. This differed from the other experiments as when looking to correct a misspelling using the spelling suggestions, children tended to click towards the top of the suggestion list, which lends itself to selection behaviors [25], which we discussed above.

Audio set up. Children’s comments on the use of the audio playback were positive. There were some children who commented that the audio cues echoed, which may have resulted from hovering over multiple suggestions in quick succession. Overall, however, many children offered their unsolicited feedback that it helped them pick the right word.

Although audio cues seemed to assist children the most in making accurate spelling suggestion selections, this solution may not always be available. Children may not always use a device with reliable audio. Additionally, the use of audio may not

be appropriate in specific environments, such as classrooms where it may be loud or there may be an expectation to be quiet, or where headphones are not available or allowed. Moreover, effective speech synthesis systems, such as Amazon's Polly incur a cost. Open-source speech synthesis systems are available, but are generally of a lesser quality.

Interface misspelling cues. In some cases, children recognized that the underlined and highlighted word was a spelling error and, as a first step, tried to correct it themselves without the help of the spellchecker's suggestions and consequently also did not utilize the multimodal cues. This could be because these children did not realize they could hover over the spelling error to get a list of spelling suggestions or simply wanted to fix the spelling error themselves without the assistance of spelling suggestions.

Spellchecker. The intention with *KidSpell*, and any spellchecker, is to position the most likely spelling suggestions towards the top of the list of spelling suggestions, making them easier to find. This was the case with the baseline experiment, however, as noted previously, the order was randomized in the multimodal cue conditions so that we could explicitly see how children may explore (or not) the suggestions if they were not in ranked order. We would expect the randomized order in our multimodal experimental conditions to make identifying and selecting the intended term more difficult compared to an order that positioned more likely words towards the top of the suggestion list. Despite this, the experiments using the multimedia cues consistently outperformed the baseline experiment, indicating that these cues impact the usability for children.

LIMITATIONS

We discuss below the limitations that emerged from our work.

Typing. Knowing that children have problems converting sounds to words, our system required the children to do that while also typing the words and thinking of how to spell each word. As a result, children often did not look up from the keyboard while typing. Children would then perform queries with spelling errors without noticing the misspelled word or ask for help instead of figuring out the spelling on their own. This sometimes happened even when they knew that there were suggestions available for the misspelled words.

Technical limitations. Due to the use of an Amazon Polly's Joanna for speech synthesis, we cannot control the way specific words are pronounced. The issue of not being able to directly curate specific audio files for individual words can be an issue for words with a unique phonetic pronunciation.

As noted in the discussion, the images used for spelling suggestions were not curated and were automatically retrieved through Google's API. This resulted in images that could be inconsistent, vague, and not make concrete connections to words. It was noticed on multiple occasions that in such scenarios having audio cues would have helped the children.

Another technical limitation that occurred was due to the limitations of the spellchecker we used. Currently, *KidSpell* only detects non-word errors (i.e., spelling errors that do not result

in the correct spelling of an unintended word). Therefore, not all types of spelling errors were explored in these experiments. For example, the incorrect use of a homonym would not be detected as a spelling error, and if it was, it is unlikely that comparing *to*, *two*, or *too* would be helpful with our speech synthesis audio playback.

Participants. A limiting factor on the choice of participants is that the conditions were all performed in informal educational settings (i.e, various STEM events). This limits the choice of participants to those that have a supported interest in the opportunity.

Knowledge of user interface. As shown in Figure 3, each correctly spelled word is shown in black, misspelled words are red with an underline and stop words are grey in color. Some children did not find the red highlight with underline to be suggestive of an interactive element that could be hovered over to see the generated spellchecking suggestions like we had thought. Instead, they would try to edit the word again to come up with a new spelling or ask the facilitator or parent for help. As such, some children required prompting in order to hover over the misspelled word in order to display a list of spelling suggestions. Some children who had a habit of looking up from the keyboard started re-spelling the word, as soon as it turned red.

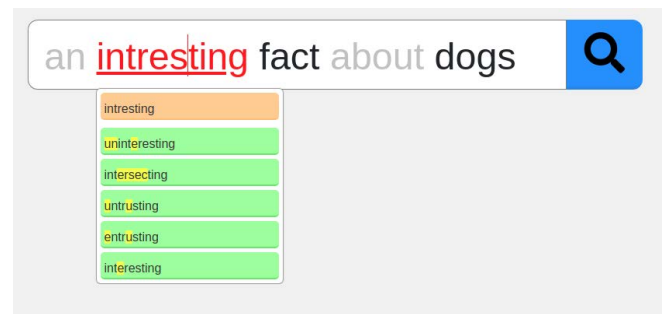


Figure 3. Color code to denote various words in a query: words in red are spelling errors, words in grey are correctly spelled stop words, and words in black are all other correctly spelled words.

For stop words that were faded to grey, e.g. *about* in Figure 3, the children who struggled to spell it and got it right after trying to spell it again, were still not sure that they had corrected the spelling mistake when the color changed from red to grey. They took this as an indication that the word was still spelled incorrectly and tried to re-spell it and as a result could not complete their search query successfully.

CONCLUSIONS AND FUTURE WORK

We conducted a study to explore the impact that image and speech synthesis audio cues have on children when selecting spelling suggestions. Analysis from children ($n=191$) interactions with a search interface, similar to that of Google or Bing, that offered a variety of multimodal cues for spellchecking revealed that those cues helped children identify and select their intended words. Out of the three kinds of multimodal systems - audio cues, image cues, and both audio and images

cues - we saw that all performed better than the system that had strictly text-based cues. The audio-only cues condition led to the best performance, followed by the both cues condition, then the image-only cues condition.

The use of synthesized speech to read aloud spelling suggestions from a phonetic-based spellchecker proved to be an effective strategy to help children successfully select their intended word. It is possible that, because of the method used for automatic image acquisition, the images used did not make a clear connection to the spelling suggestion. Further examination on the effect of image cues could be explored that limits images to those that have concrete connections to words or curating the images that are used. Such curation, however, would require an investment of time and labor.

Although most children understood that a red, underlined word indicated it was spelled incorrectly, some children did not associate that as an interactive element on the screen. To assist with this, future work includes playing a sound indicator to make the spelling suggestions more apparent or automatically opening spelling suggestions without the user needing to hover over the misspelled word.

Providing alternative ways for children to query information, including voice interaction, could enable a child to better explain his or her information need in order to input a query [17]. The work in [31] provides a starting point for the potential a vocal assistant can have in children's search when compared to a traditional keyboard-driven approach. Future work would include a performance analysis between a voice-assisted search engine and a keyboard-driven search engine with enhanced multimodal cues.

We found that multimodal cues helped children select the correct spelling of their intended word with an accuracy of up to 92%. We also noticed that when spelling was misunderstood, the children tended to follow a pattern of selecting options towards the top of the suggestions; however, when prompted an audio cue that helped with the child's spelling understanding, children no longer resorted to this pattern. We intend to explore further the idea of adapting our search tool to help children formulate queries that better capture their information needs.

ACKNOWLEDGEMENTS

This research has been partially supported by the National Science Foundation (Award #1763649). We are thankful to the children who are part of *KidsTeam* for their cooperation and participation in the data collection study. We also thank the organizers of the STEM events at the local schools for inviting us to show our system and collect data for evaluation.

SELECTION AND PARTICIPATION OF CHILDREN

The speech synthesis preliminary study was conducted with 8 children (5 girls, 3 boys; ages 6-10) who were part of an intergenerational design team that meets twice a week after school. Those children were recruited via publicly posted flyers and local social media outputs. The purpose of the

investigations were explained to participants and their parents. Parents signed consent forms to allow their children to participate, and children assented to participating.

The observation data of using the spelling suggestions on the search tool with various modalities were gathered at informal educational STEM events where children (191, ages 6-12) were interacting with computers, robots, math, engineering, and other STEM related items. Only observational data and interaction logs were collected, children were not surveyed regarding their use of the system. Our team had two stations one where children could program robots and another where they could search for STEM-related and local-area information.

Both the intergenerational design team and STEM event observation and logging protocols were approved by the institutional review board.

REFERENCES

- [1] Oghenemaro Anuyah, Jerry Alan Fails, and Maria Soledad Pera. 2018. Investigating query formulation assistance for children. In *Proceedings of the 17th ACM Conference on Interaction Design and Children*. ACM, 581–586. DOI: <http://dx.doi.org/10.1145/3202185.3210779>
- [2] Ion Madrazo Azpiazu, Nevena Dragovic, Maria Soledad Pera, and Jerry Alan Fails. 2017. Online searching and learning: YUM and other search tools for children and teachers. *Information Retrieval Journal* 20, 5 (2017), 524–545. DOI: <http://dx.doi.org/10.1177/0165551515614473>
- [3] Donald R Bear, Marcia Invernizzi, Shane R Templeton, and Francine Johnston. 2015. Words their way. (2015). DOI: <http://dx.doi.org/10.1080/10753569.2013.850461>
- [4] Dania Bilal. 1998. Children's Search Processes in Using World Wide Web Search Engines: An Exploratory Study.. In *Proceedings of the ASIS annual meeting*, Vol. 35. ERIC, 45–53. DOI: <http://dx.doi.org/10.1002/asi.10145.49>
- [5] Dania Bilal and Joe Kirby. 2002. Differences and similarities in information seeking: children and adults as Web users. *Information processing & management* 38, 5 (2002), 649–670. DOI: [http://dx.doi.org/10.1016/S0306-4573\(01\)00057-7](http://dx.doi.org/10.1016/S0306-4573(01)00057-7)
- [6] Christine L Borgman, Sandra G Hirsh, Virginia A Walter, and Andrea L Gallagher. 1995. Children's searching behavior on browsing and keyword online catalogs: the Science Library Catalog project. *Journal of the American Society for information Science* 46, 9 (1995), 663–684. DOI: [http://dx.doi.org/10.1002/\(SICI\)1097-4571\(199510\)46:9<663::AID-ASIA4>3.0.CO](http://dx.doi.org/10.1002/(SICI)1097-4571(199510)46:9<663::AID-ASIA4>3.0.CO)
- [7] Brody Downs, Oghenemaro Anuyah, Aprajita Shukla, Jerry Alan Fails, Maria Soledad Pera, Katherine Landau Wright, and Casey Kennington. 2020. KidSpell: A Child-Oriented, Rule-Based, Phonetic Spellchecker. In *In Proceedings of the Twelfth International Conference*

on Language Resources and Evaluation (LREC 2020). ACL.

- [8] Nevena Dragovic, Ion Madrazo Azpiazu, and Maria Soledad Pera. 2016. Is sven seven?: A search intent module for children. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. ACM, 885–888. DOI : <http://dx.doi.org/10.1145/2911451.2914738>
- [9] Allison Druin. 1999. Cooperative Inquiry: New Technologies for Children. (1999), 8. DOI : <http://dx.doi.org/10.1145/302979.303166>
- [10] Allison Druin, Elizabeth Foss, Hilary Hutchinson, Evan Golub, and Leshell Hatley. 2010. Children’s Roles Using Keyword Search Interfaces at Home. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI ’10)*. ACM, New York, NY, USA, 413–422. DOI : <http://dx.doi.org/10.1145/1753326.1753388>
- [11] Sergio Duarte Torres and Ingmar Weber. 2011. What and how children search on the web. In *Proceedings of the 20th ACM international conference on Information and knowledge management*. ACM, 393–402. DOI : <http://dx.doi.org/10.1145/2063576.2063638>
- [12] Pravallika Etoori, Manoj Chinnakotla, and Radhika Mamidi. 2018. Automatic spelling correction for resource-scarce languages using deep learning. In *Proceedings of ACL 2018, Student Research Workshop*. 146–152. DOI : <http://dx.doi.org/10.18653/v1/P18-3021>
- [13] Jerry Alan Fails, Mona Leigh Guha, and Allison Druin. 2013. Methods and Techniques for Involving Children in the Design of New Technology for Children. *Foundations and Trends® in Human–Computer Interaction* 6, 2 (Dec. 2013), 85–166. DOI : <http://dx.doi.org/10.1561/11000000018>
- [14] Jerry Alan Fails, Maria Soledad Pera, Oghenemaro Anuyah, Casey Kennington, Katherine Landau Wright, and William Bigirimana. 2019. Query Formulation Assistance for Kids: What is Available, When to Help & What Kids Want. In *Proceedings of the 18th ACM International Conference on Interaction Design and Children (IDC ’19)*. ACM, New York, NY, USA, 109–120. DOI : <http://dx.doi.org/10.1145/3311927.3323131>
- [15] Adam Fourney, Meredith Ringel Morris, and Ryen W White. 2017. Web search as a linguistic tool. In *Proceedings of the 26th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 549–557. DOI : <http://dx.doi.org/10.1145/3038912.3052651>
- [16] Tatiana Gossen. 2016. *Search engines for children: search user interfaces and information-seeking behaviour*. Springer. DOI : <http://dx.doi.org/10.1145/3053408.3053428>
- [17] Tatiana Gossen, Julia Hempel, and Andreas Nürnberger. 2013. Find it if you can: usability case study of search engines for young users. *Personal and Ubiquitous Computing* 17, 8 (2013), 1593–1603.
- [18] Tatiana Gossen, Michael Kotzyba, and Andreas Nürnberger. 2017. Search Engine for Children: User-Centered Design. *Datenbank-Spektrum* 17, 1 (2017), 61–67. DOI : <http://dx.doi.org/10.1145/2637002.2637007>
- [19] Tatiana Gossen, Thomas Low, and Andreas Nürnberger. 2011. What are the real differences of children’s and adults’ web search. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*. ACM, 1115–1116. DOI : <http://dx.doi.org/10.1145/2556288.2557031>
- [20] Tatiana Gossen, Marcus Nitsche, and Andreas Nürnberger. 2012a. Knowledge journey: A web search interface for young users. In *Proceedings of the Symposium on Human-Computer Interaction and Information Retrieval*. ACM, 1. DOI : <http://dx.doi.org/10.1145/2391224.2391225>
- [21] Tatiana Gossen, Marcus Nitsche, and Andreas Nürnberger. 2012b. Search User Interface Design for Children: Challenges and Solutions.. In *EuroHCIR*. 59–62. DOI : http://dx.doi.org/10.1007/978-1-4614-2179-5_1
- [22] Daphne Greenberg, Linnea C Ehri, and Dolores Perin. 2002. Do adult literacy students make the same word-reading and spelling errors as children matched for word-reading age? *Scientific Studies of Reading* 6, 3 (2002), 221–243. DOI : http://dx.doi.org/10.1207/S1532799XSSR0603_2
- [23] Mona Leigh Guha, Allison Druin, and Jerry Alan Fails. 2013. Cooperative Inquiry revisited: Reflections of the past and guidelines for the future of intergenerational co-design. *International Journal of Child-Computer Interaction* 1, 1 (Jan. 2013), 14–23. DOI : <http://dx.doi.org/10.1016/j.ijcci.2012.08.003>
- [24] Jai Gupta, Zhen Qin, Michael Bendersky, and Donald Metzler. 2019. Personalized Online Spell Correction for Personal Search. In *The World Wide Web Conference*. ACM, 2785–2791. DOI : <http://dx.doi.org/10.1145/3308558.3313706>
- [25] Jacek Gwizdka and Dania Bilal. 2017. Analysis of Children’s Queries and Click Behavior on Ranked Results and Their Thought Processes in Google Search. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval (CHIIR ’17)*. Association for Computing Machinery, New York, NY, USA, 377–380. DOI : <http://dx.doi.org/10.1145/3020165.3022157>
- [26] Matthias Hagen, Martin Potthast, Marcel Gohsen, Anja Rathgeber, and Benno Stein. 2017. A Large-Scale Query Spelling Correction Corpus. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 1261–1264. DOI : <http://dx.doi.org/10.1145/3077136.3080749>

- [27] Paul Herring, Karen Kear, Kieron Sheehy, and Roger Jones. 2017. A virtual tutor for children with autism. *Journal of Enabling Technologies* 11, 1 (2017), 19–27. DOI: <http://dx.doi.org/10.1109/IJCNN.2010.5596584>
- [28] Hilary Hutchinson, Allison Druin, Benjamin B Bederson, Kara Reuter, Anne Rose, and Ann Carlson Weeks. 2005. How do I find blue books about dogs? The errors and frustrations of young digital library users. *Proceedings of HCII 2005* (2005), 22–27. DOI: <http://dx.doi.org/10.1.1.64.880>
- [29] Simon Knight. 2014. *Finding knowledge—what is it to know when we search?* Institute of Network Cultures. DOI: <http://dx.doi.org/10.1007/978-0-387-89024-116>
- [30] Izabela Krejtz, Agnieszka Szarkowska, Krzysztof Krejtz, Agnieszka Walczak, and Andrew Duchowski. 2012. Audio Description as an Aural Guide of Children’s Visual Attention: Evidence from an Eye-Tracking Study. In *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA ’12)*. Association for Computing Machinery, New York, NY, USA, 99–106. DOI: <http://dx.doi.org/10.1145/2168556.2168572>
- [31] Monica Landoni, Davide Matteri, Emiliana Murgia, Theo Huibers, and Maria Soledad Pera. 2019. Sonny, Cerca! evaluating the impact of using a vocal assistant to search at school. In *International Conference of the Cross-Language Evaluation Forum for European Languages*. Springer, 101–113.
- [32] Ion Madrazo Azpiazu, Nevena Dragovic, Oghenemaro Anuyah, and Maria Soledad Pera. 2018. Looking for the Movie Seven or Sven from the Movie Frozen? A Multi-Perspective Strategy for Recommending Queries for Children. In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval (CHIIR ’18)*. Association for Computing Machinery, New York, NY, USA, 92–101. DOI: <http://dx.doi.org/10.1145/3176349.3176379>
- [33] Joseph E. Michaelis and Bilge Mutlu. 2019. Supporting Interest in Science Learning with a Social Robot. In *Proceedings of the 18th ACM International Conference on Interaction Design and Children (IDC ’19)*. ACM, New York, NY, USA, 71–82. DOI: <http://dx.doi.org/10.1145/3311927.3323154>
- [34] Ian Rowlands, David Nicholas, Peter Williams, Paul Huntington, Maggie Fieldhouse, Barrie Gunter, Richard Withey, Hamid R Jamali, Tom Dobrowolski, and Carol Tenopir. 2008. The Google generation: the information behaviour of the researcher of the future. In *Aslib proceedings*, Vol. 60. Emerald Group Publishing Limited, 290–310. DOI: <http://dx.doi.org/10.1108/00012530810887953>
- [35] Mark Sadoski, Ernest T Goetz, and Joyce B Fritz. 1993. A causal model of sentence recall: Effects of familiarity, concreteness, comprehensibility, and interestingness. *Journal of reading behavior* 25, 1 (1993), 5–16. DOI: <http://dx.doi.org/10.1080/10862969309547799>
- [36] Paul Solomon. 1993. Children’s information retrieval behavior: A case analysis of an OPAC. *Journal of the American Society for information Science* 44, 5 (1993), 245–264. DOI: <http://dx.doi.org/10.1002/asi.10130>
- [37] Katherine A Dougherty Stahl, Kevin Flanigan, and Michael C McKenna. 2019. *Assessment for reading instruction*. Guilford Press.
- [38] Ye Yuan, Stryker Thompson, Kathleen Watson, Alice Chase, Ashwin Senthilkumar, A.J. Bernheim Brush, and Svetlana Yarosh. 2019. Speech interface reformulations and voice assistant personification preferences of children and parents. *International Journal of Child-Computer Interaction* 21 (2019), 77 – 88. DOI: <http://dx.doi.org/https://doi.org/10.1016/j.ijcci.2019.04.005>
- [39] Zheng-Jun Zha, Linjun Yang, Tao Mei, Meng Wang, Zengfu Wang, Tat-Seng Chua, and Xian-Sheng Hua. 2010. Visual query suggestion: Towards capturing user intent in internet image search. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 6, 3 (2010), 13. DOI: <http://dx.doi.org/10.1145/1823746.1823747>