

# Phoneme-based Predictive Text Entry Interface

Ha Trinh  
Northeastern University  
hatrinh@ccs.neu.edu

Annalu Waller  
University of Dundee  
awaller@computing.dundee.ac.uk

Keith Vertanen  
Montana Tech  
kvertanen@mtech.edu

Per Ola Kristensson  
University of St Andrews  
pok@st-andrews.ac.uk

Vicki L. Hanson  
Rochester Institute of Technology  
vlhics@rit.edu

## ABSTRACT

Phoneme-based text entry provides an alternative typing method for nonspeaking individuals who often experience difficulties in orthographic spelling. In this paper, we investigate the application of rate enhancement strategies to improve the user performance of phoneme-based text entry systems. We have developed a phoneme-based predictive typing system, which employs statistical language modeling techniques to dynamically reduce the phoneme search space and offer accurate word predictions. Results of a case study with a nonspeaking participant demonstrated that our rate enhancement strategies led to improved text entry speed and error rates.

## Categories and Subject Descriptors

K.4.2 [Computers and Society]: Social Issues – *assistive technologies for persons with disabilities.*

## General Terms

Performance, Design, Human Factors.

## Keywords

Phoneme-based text entry; prediction; speech impairments.

## 1. INTRODUCTION

It is well documented that people with severe speech impairments often face challenges in literacy acquisition [4]. Without literacy skills, they are unable to effectively use conventional orthographic-based text entry approaches. In addition, speech impairments are often accompanied by reduced motor control, which restricts the affected individuals from accessing traditional input methods such as full-sized keyboards.

Previous research has proposed phoneme-based text entry as an alternative typing approach for people with spelling difficulties [3]. By combining sequences of phonemes (i.e. speech sounds), spoken or written words can be created without knowledge of orthographic spelling. Existing phoneme-based systems,

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 third-party components of this work must be honored. For all other uses, contact the Owner/Author.

Copyright is held by the owner/author(s).

ASSETS' 14, October 20–22 2014, Rochester, NY, USA.  
ACM 978-1-4503-2720-6/14/10.  
<http://dx.doi.org/10.1145/2661334.2661424>

however, suffer from a number of usability issues, including poor entry rate, difficult access methods to select phonemes, and high learning demands [1].

We began to address this issue in our previous work on applying prediction methods to phoneme entry [5]. We developed the iSCAN system [5], which provides two rate enhancement strategies: *dynamic phoneme layout* and *word auto-completion*. iSCAN dynamically rearranges the phoneme layout to allow for faster access to the most probable next phonemes. The word auto-completion offers a single word prediction based on the current phoneme prefix and prior words. Results of our evaluations evidenced the effectiveness of our predictive strategies [5]. However, we also identified the need to provide enhanced support during the phoneme search process and to improve the hit rate of word predictions.

In our current work, we have refined iSCAN to develop iSCAN-2. iSCAN-2 increases the hit rate of the word predictions by offering a *5-word prediction menu* that is seamlessly integrated into the phoneme layout. In addition to the *dynamic phoneme layout*, the system offers a *phoneme set reduction* feature that dynamically narrows down the phoneme search space.

## 2. iSCAN-2 DESIGN

### 2.1 Phoneme Entry Method

Implemented as an iPad application, iSCAN-2 provides access to 42 spoken phonemes from the Jolly Phonics literacy program [2] using a two-layer pie menu interface specifically designed for users with limited dexterity. The phonemes are classified into 7 groups and mapped onto 7 directions on the front layer of the pie menu (Figure 1a). Phoneme with the highest probability of entry in each group (as estimated from a 6-gram phoneme language model) is chosen as the representative of the group and displayed on the front layer. Selecting a phoneme group on the front layer switches the pie menu to the phoneme layer, which displays all phonemes within the group (Figure 1b).

The phoneme entry consists of three steps: (1) selecting the correct phoneme group from the front layer; (2) navigating within the phoneme layer to search for the intended phoneme; (3) moving back to the center circle to confirm the selection.

### 2.2 Rate Enhancement Strategies

#### 2.2.1 5-Word Prediction Pie Menu

Once the user has selected a phoneme and moved back to the center circle, the pie menu switches to the prediction layer displaying up to 5 most probable words (Figure 1c). The word probabilities are estimated from a 3-gram word language model, taking into account the current phoneme prefix and up to two

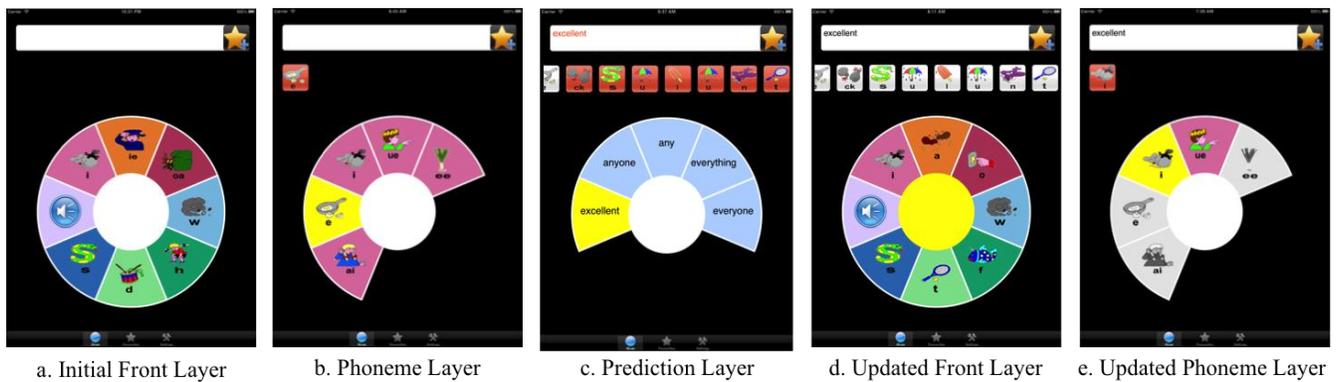


Figure 1. Five stages of the pie menu in the process of creating the word "excellent"

prior words. The user navigates the prediction layer to identify the target word, then moving back to the center circle to confirm the selection and switch back to the front layer. If the target word has not appeared on the prediction menu, the user can touch an empty area (i.e. the black area) to escape from the prediction layer and continue entering the next phonemes. Auditory feedback is provided throughout the navigation process to aid users with limited reading skills.

### 2.2.2 Dynamic Phoneme Layout

After each phoneme or word selection, the system recalculates the probability of entry of each phoneme and rearranges the phoneme layout accordingly. The most probable next phoneme in each group becomes its new representative and appears on the front layer of the pie menu (see Figure 1d). The remaining phonemes in the group are reordered so that phonemes with higher probabilities are closer to the representative phoneme and hence require fewer movements to navigate to from the representative phoneme.

### 2.2.3 Dynamic Phoneme Set Reduction

After each phoneme selection, the system greys out any phonemes *not* in the most 20 probable next phonemes as determined by the 6-gram phoneme language model and a pronunciation dictionary. For example, in Figure 1e, three phonemes in the "Wide Mouth" phoneme group have been greyed out. If all phonemes from a group are eliminated, that group will be greyed out from the front layer. This provides additional visual cues for a more guided phoneme search. Our theoretical evaluation showed that the dynamically reduced 20-phoneme set achieved a very high average hit rate of 97.2%.

## 3. FORMATIVE EVALUATION

We report results of a case study with a nonspeaking adult, comparing the usability of three versions of our phoneme-based text entry interface: *non-predictive* (i.e. all the rate enhancement features are turned off), *iSCAN* (with dynamic phoneme layout and word auto-completion), and *iSCAN-2*.

### 3.1 Participant

Our participant is a 41-year-old male adult who has cerebral palsy and limited literacy as measured by our pre-study literacy tests. The participant had used our previous *iSCAN* system for four months prior to this study.

### 3.2 Procedure

The study consisted of four sessions, including one training session on *iSCAN-2* and three testing sessions. In each testing

session, the participant was asked to transcribe 10 randomly generated phrases as quickly and accurately as possible. The participant used *iSCAN* in Session 2, followed by the non-predictive version in Session 3 and *iSCAN-2* in Session 4.

## 3.3 Results and Discussion

We measured the participant's entry rates in words per minute (WPM). His text entry accuracy was measured as word error rate (WER). The results were: *non-predictive* (1.14 WPM, 50% WER), *iSCAN* (2.54 WPM, 19.26% WER), *iSCAN-2* (4.19 WPM, 0% WER).

The participant expressed a strong preference for the 5-word prediction over the word auto-completion in *iSCAN*. Even with limited reading skills, he showed very little problems selecting desired words from the menu with support of the auditory feedback. This suggests that, for users who rely heavily on the predictions, it is crucial to maintain high prediction accuracy even at the cost of the physical and cognitive workload required to scan a multi-word menu. However, the presentation of the prediction results should be carefully designed to minimize such workload and facilitate the scanning process.

## 4. CONCLUSION

We present the development and formative evaluation of a phoneme-based predictive text entry system. While our system was specifically designed for nonspeaking individuals, it could potentially be applicable to a wider population who experience spelling difficulties.

## 5. REFERENCES

- [1] Goodenough-Trepagnier, C., Prather, P. Communication systems for the nonvocal based on frequent phoneme sequences. *Journal of Speech and Hearing Research*, 24, 322-329.
- [2] Lloyd, S.M. *The Phonics Handbook*. Jolly Learning Ltd., Chigwell, 1998.
- [3] Schroeder, J.E. Improved spelling for persons with learning disabilities. in *20th International Conference on Technology and Persons with Disabilities*, (California, USA, 2005).
- [4] Smith, M. *Literacy and augmentative and alternative communication*. Elsevier Academic Press, 2005.
- [5] Trinh, H., Waller, A., Vertanen, K., Kristensson, P. O., Hanson, V. L. *iSCAN: A phoneme-based predictive communication aid for nonspeaking individuals*. in *ASSETS'12*, (Colorado, USA, 2012).