# iSCAN: A Phoneme-based Predictive Communication Aid for Nonspeaking Individuals

Ha Trinh[1], Annalu Waller[1], Keith Vertanen[2], Per Ola Kristensson[3], Vicki L. Hanson[1]

[1]School of Computing
University of Dundee
{hatrinh,awaller,vlh}@computing.dundee.ac.uk

[2]Department of Computer Science
Montana Tech of the University of Montana
kvertanen@mtech.edu

[3]School of Computer Science
University of St Andrews
pok@st-andrews.ac.uk

## ABSTRACT

The high incidence of literacy deficits among people with severe speech impairments (SSI) has been well documented. Without literacy skills, people with SSI are unable to effectively use orthographic-based communication systems to generate novel linguistic items in spontaneous conversation. To address this problem, phoneme-based communication systems have been proposed which enable users to create spoken output from phoneme sequences. In this paper, we investigate whether prediction techniques can be employed to improve the usability of such systems. We have developed iSCAN, a phoneme-based predictive communication system, which offers phoneme prediction and phoneme-based word prediction. A pilot study with 16 able-bodied participants showed that our predictive methods led to a 108.4% increase in phoneme entry speed and a 79.0% reduction in phoneme error rate. The benefits of the predictive methods were also demonstrated in a case study with a cerebral palsied participant. Moreover, results of a comparative evaluation conducted with the same participant after 16 sessions using iSCAN indicated that our system outperformed an orthographic-based predictive communication device that the participant has used for over 4 years.

## Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation**]: User Interfaces – *input devices and strategies.* K.4.2 [**Computers and Society**]: Social Issues – *assistive technologies for persons with disabilities.*

## General Terms

Performance, Design, Human Factors.

## Keywords

Augmentative and Alternative Communication, Phoneme-based Communication, Phoneme Prediction, Word Prediction.

## 1. INTRODUCTION

The United States Census Bureau has estimated that 2.5 million Americans had difficulty having their speech understood, of which 0.5 million had severe speech impairments (SSI) [3]. Not only are these individuals unable to communicate using natural speech, many of them have motor impairments which restrict access to other communication channels, such as signing or writing. Thus, they often require Augmentative and Alternative Communication (AAC) strategies to meet their communication needs. The majority of existing AAC systems employ graphical symbols to encode a limited set of commonly used words and messages, thereby allowing for quick retrieval of reusable conversational content. However, users of these systems are limited to pre-programmed items rather than being able to create novel words and messages in spontaneous conversation. To overcome this limitation, a number of orthographic-based AAC systems have been developed to enable users to spell out their own messages. Prediction techniques, such as character or word prediction, are often applied to improve the usability and accessibility of these systems. However, these systems are only applicable to people with literacy skills; skills that many children and adults with SSI struggle to acquire [17].

In an effort to empower nonspeaking users to generate spontaneous, unique words and messages without the need for literacy skills, previous research has proposed the use of a phoneme-to-speech approach. This approach allows users to access a limited set of phonemes (i.e. speech sounds). By combining sequences of phonemes, novel conversational items can be generated without knowledge of orthographic spelling. This approach has been used in several communication aids [7] and literacy tools [1]. It has also been adopted as an alternative typing method for people with spelling difficulties [16].

To date, research on phoneme-based AAC systems is very limited. The few published reports on existing phoneme-based systems have highlighted a number of usability issues, including poor communication rate [4, 24], difficult access methods to target phonemes [1, 4, 24], and high learning demands [4, 8]. This past work demonstrates the need for rate enhancement strategies to facilitate phoneme entry and word creation processes. We began to address this issue in our previous work by applying prediction methods to phoneme-based AAC systems [19]. We developed a phoneme-based prediction model, which employed statistical language modeling techniques to perform context-dependent phoneme prediction and word prediction. Theoretical evaluation demonstrated that our prediction model could potentially lead to substantial keystroke savings when applied to a hypothetical 12-key phoneme keyboard [19]. However, we did not conduct any empirical studies on how the prediction model could be integrated into an actual phoneme-based AAC system to improve user performance.

Thus, our current work aims to demonstrate empirical evidence of the potential of our prediction methods. We have developed a


*ASSETS'12*, October 22 - 24, 2012, Boulder, Colorado, USA.

novel phoneme-based predictive communication system. Our system performs prediction at both phoneme and word levels. Our model's phoneme predictions are used to dynamically rearrange the phoneme interface layout to allow for faster access to the most probable next phonemes. The word creation process is further supported by a phoneme-based word auto-completion feature. This feature predicts the word being entered based on the current phoneme prefix and prior words.

Before evaluating our system with representative AAC users, we first wanted to assess the usability of our predictive methods. For this purpose, we tested our system in a three-session study with 16 able-bodied participants measuring their entry rates, error rates, and user experience. We then validated the benefits of our predictive methods in a longitudinal case study with a cerebral palsied adult and report evaluation results of 16 training and practice sessions. In addition, we discuss the results of a study comparing the usability of our phoneme-based predictive system with two orthographic-based predictive communication systems already familiar to the participant. Finally, we propose a number of further studies to extend our current work.

# 2. RELATED WORK

## 2.1 Phoneme-based AAC Systems

The history of phoneme-based AAC systems dates back in 1978 with the development of the HandiVoice by Phonic Ear [7]. The device contained a set of 45 phonemes, each of which was assigned a three-digit code. Users could access these phonemes using a numeric keypad and blend them into synthetic speech. Reports from a few HandiVoice users [4, 24] highlighted the system's slow communication rate as well as the high physical and cognitive efforts required to select target phonemes and produce intended words and sentences.

In an effort to enhance the communication rate of phoneme-based AAC systems, Goodenough-Trepagnier and Prather [8] developed the SPEEC system. The SPEEC system provided users with a combined set of spoken phonemes and frequently used phoneme sequences, each of which was represented by a letter or a letter combination. The size of the selection set ranged from 256 to 400 items. The authors reported that one nonspeaking individual trained in a 400-item version of the system achieved a speed of 8.2 words per minute, a 30% increase over an alphabet system [9]. However, the amount of training was not specified. Results of an evaluation with five cerebral palsied adolescents, including one pre-reader as well as beginning and proficient readers, showed that the participants required from 4 to 8 months of training to achieve some degree of proficiency [8]. This highlights the high learning demands imposed on the users of this system.

Black et al. [1] explored the potential of phoneme-based AAC systems to support language play and phonics teaching for children with SSI. The researchers developed the PhonicStick™ talking joystick, which enables users to access the 42 phonemes used in the Jolly Phonics literacy program [13] using a joystick interface. A prototype of the PhonicStick™, using a subset of 6 phonemes, has been evaluated with seven children without and with SSI. Results of the evaluations showed that the participants could create short words using the phonemes. However, some participants with severe motor impairments experienced difficulties in using the joystick to access target phonemes [1].

Schroeder [16] developed the REACH Sound-It-Out Phonetic Keyboard™, a phoneme-based typing interface for individuals with spelling difficulties. This on-screen keyboard consists of 40 phonemes and 4 phoneme combinations, each of which is represented by a letter or a digraph and optionally a picture. It utilizes a dictionary-based phoneme prediction method to remove improbable next phonemes from the keyboard, thereby aiding users in visually locating the next phoneme in the intended word. The system also employs a dictionary-based word prediction method to present users with a list of the most frequently used words that phonetically match the current phoneme prefix. Evaluations conducted with children and adults both with and without learning disabilities demonstrated that the system led to an increased text input accuracy compared to conventional letter-based keyboards [16]. To our knowledge, REACH Sound-It-Out Phonetic Keyboard™ is the only currently available system that provides phoneme-based predictions. However, these predictions rely on a simple dictionary-based algorithm, which does not take into account contextual information, such as prior text. To date, little research has been done on how more advanced prediction techniques can be employed to improve phoneme-based predictions.

## 2.2 Phonological Awareness

We start with the assumption that in order to use phoneme-based AAC systems without support of predictive features, users must have adequate phonological awareness (PA) skills. PA refers to the explicit attention to the sound structure of language and encompasses a wide range of skills, from rhyming recognition, phoneme blending, to phoneme segmentation and phoneme manipulation [2]. These skills are essential prerequisites for literacy acquisition [2], even in populations of profoundly deaf readers who do not use speech as their primary means of communication [10]. Previous research has shown that individuals with SSI can develop their PA skills despite the absence of speech production [5]. This implies that phoneme-based AAC systems are potentially usable to these individuals. However, many individuals with SSI demonstrate PA deficits compared to their typical developing peers and hence would require focused PA training to acquire these skills [12]. This suggests that users with poor PA skills stand to benefit from phoneme-based predictive interfaces.

## 2.3 Prediction Techniques

Prediction is a rate enhancement strategy widely used in orthographic-based AAC systems [6]. A number of prediction strategies have been developed for AAC users, of which the most commonly used are character prediction and word prediction. Character prediction anticipates probable next characters based on the previously selected characters. Word prediction anticipates the word being entered on the basis of the prefix of the current word and possibly prior words, thereby saving the user the effort of entering every character of a word. Prediction results are typically presented in a horizontal or vertical list, requiring the users to scan the list to select the desired item. While these techniques often result in keystroke savings, previous work has suggested that these savings might not be translated into increased communication rates due to the cognitive and perceptual workload of navigating the prediction list to search for target items [11].

Most existing prediction systems employ statistical language modelling techniques to perform prediction tasks. These techniques often use a large collection of training text to construct n-gram language models, which can be used to predict next most probable items (such as characters or words) based on (n-1) preceding items. A number of advanced language modelling techniques have also been investigated, which utilize

additional information such as word recency, syntactic information, semantic information, and topic modelling [6]. These techniques have the potential to improve prediction performance at the cost of increased computational complexity.

# 3. SYSTEM DESIGN

Little work has been done on how well statistical prediction models can be adapted to phoneme-based AAC systems. In this section, we provide an overview of our phoneme-based prediction model. Our model employs statistical language modeling techniques to perform single phoneme prediction and phoneme-based word prediction. Readers are referred to [19] for a more detailed description of our model. We then present the design of iSCAN (**I**nteractive **S**ound-based **C**ommunication **A**id for **N**on-speakers), which uses the prediction model to implement two predictive features, namely dynamic phoneme layout and word auto-completion.

## 3.1 Phoneme-based Prediction Model

Our prediction model uses a set of 42 phonemes (17 vowels and 25 consonants) used in the Jolly Phonics, a systematic synthetic phonics program widely used in the UK for literacy teaching [13]. By using a literacy-based phoneme set, our model can readily be incorporated into both literacy learning tools (such as the PhonicStick™ joystick [1]) and communication aids.

The prediction model uses a 6-gram phoneme mixture model and a 3-gram word mixture model to provide predictions at both phoneme and word levels. The 6-gram phoneme model is used to predict the next probable phonemes based on up to five preceding phonemes, while the 3-gram word model is used to predict the word currently being entered based on up to two words of prior context. Ideally, these models would be constructed from a large corpus of transcribed conversations of real AAC users on various topics. However, such a corpus has been unavailable to date. We addressed this problem by creating a large corpus of fictional AAC data via crowdsourcing [22]. This crowdsourced corpus was then used to intelligently select a much larger set of AAC-like sentences from Twitter, Blog, and Usenet datasets. To generate the phoneme language model, we converted our fictional AAC word corpus to a phoneme-based corpus using a pronunciation dictionary. We trained a 6-gram phoneme language model for each of the phoneme-based Twitter, Blog and Usenet datasets. We then created a 6-gram mixture model using linear interpolation with mixture weights optimized on our crowdsourced development set. The same approach was applied to construct the 3-gram word mixture language model. In previous work, we demonstrated that word language models generated using this approach outperformed other models trained on telephone transcripts, which were often used in previous research on AAC prediction [22].

## 3.2 iSCAN Design

We incorporated our prediction model into the design of iSCAN. iSCAN uses the 6-gram phoneme language model to dynamically optimize the phoneme layout after each phoneme selection to allow for easier access to highly probable next phonemes. Each time the user has selected a phoneme, the system also attempts to automatically complete the word currently being entered using the word language model. This allows users to complete a word without entering every single phoneme. Our design was motivated by the following principles:

*Small number of selection targets.* Many individuals with SSI also have severe motor impairments and hence often experience difficulties in accessing interfaces containing a large number of selection targets (e.g. buttons or keys). For example, many motor-impaired users are unable to use a direct selection method to access physical or on-screen keyboards, due to the lack of fine motor skills required to precisely point over small targets among a large set of keys. Although alternative access methods, such as scanning, can be employed to facilitate target acquisition, the efficiency of these methods tends to decrease as the number of selection targets increases. Therefore, in our interface design, we aimed to enable the users to access the phoneme set using a minimal number of selection targets.

*Support various input devices.* Individuals with SSI and motor impairments utilize a wide range of input devices, such as touch screen, joysticks, trackballs, eye tracking devices, or switches, to control their AAC systems Our design, therefore, should be easily adapted to effectively support the use of those devices.

*Avoid using separate lists to present prediction results.* Predictive systems that display vertical or horizontal lists can create perceptual and cognitive demands that may outweigh the keystroke savings offered by prediction [11]. We therefore aimed to find an appropriate method of displaying the prediction results without using a list-based presentation.

### *3.2.1 Interface Design*

In iSCAN, the Jolly Phonics's phoneme set is arranged onto an eight-slice two-layer pie menu adapted from the PhonicStick™ joystick interface [1]. This design provides the users with access to the 42 phonemes by using only 9 selection targets, including the 8 slices and the center circle of the pie menu. In addition, this 8-direction gestural interface design can also be easily adapted for various input devices, such as joysticks, touch-screens, or eye-tracking systems.

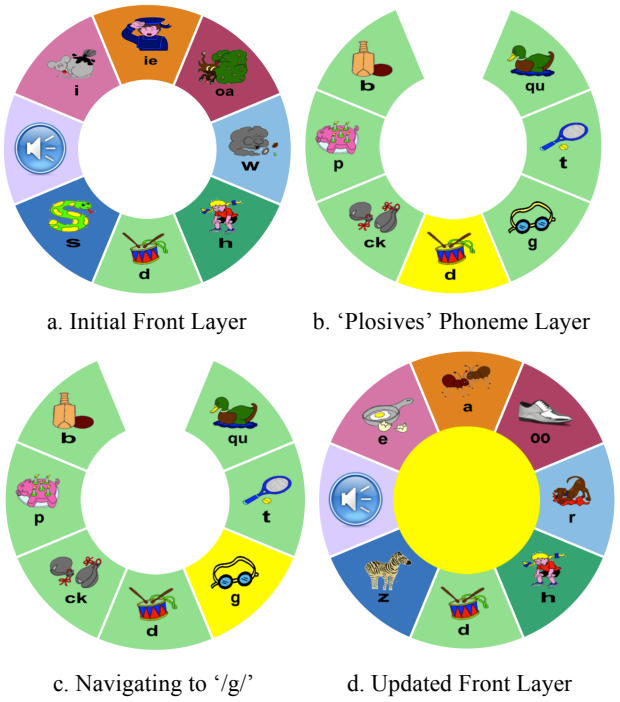
The 42 spoken phonemes are classified into 7 groups and mapped onto 7 directions on the front layer of the pie menu (Figure 1a). The phoneme groups consist of 3 vowel groups and 4 consonant groups, each of which contain from 5 to 7 phonemes. The groups are formed according to the manner of articulation and are color coded, with warm colors for vowels and cool colors for consonants (Figure 1a). Each phoneme is represented by a picture selected from the Jolly Phonics' resources and optionally a letter or digraph. The remaining direction of the front layer of the menu (i.e. West direction) is reserved for the functional group, which contains 5 functions, including 'Speak current word', 'Delete current word', 'Delete last phoneme', 'Speak current sentence', and 'Delete current sentence'. Selecting a phoneme or functional group on the front layer switches the pie menu to the secondary layer, which displays all items within the group. Each secondary layer contains a maximum of 7 item slices and at least one empty slice, which is treated as an 'escape' route to allow the user to leave the layer without selecting any items (Figure 1b).

#### *3.2.1.1 Phoneme Entry Method*

The phoneme entry consists of three steps, including: (1) selecting the correct phoneme group from the front layer; (2) navigating the secondary layer to search for the intended phoneme; (3) moving back to the center circle of the menu to confirm the selection and redisplay the front layer. Our design supports phoneme entry via either tapping or continuous gestures, thereby accommodating both novice and expert users.

Figure 1 depicts an example of the transition of the pie menu through four stages in the process of selecting phoneme */g/*, the initial phoneme of *'good'*, assuming that *'good'* is the first word

of the user's new sentence. Before the user starts creating a new sentence, the phonemes within each group are initially ordered based on their probabilities of being the first phoneme in a sentence (calculated from the phoneme language model). The phoneme with the highest probability of entry within each group is chosen as the representative of the group and displayed on the corresponding slice on the front layer (Figure 1a).




a. Initial Front Layer b. 'Plosives' Phoneme Layer

c. Navigating to '/g/' d. Updated Front Layer

**Figure 1.** Four stages of the pie menu in the process of selecting the phoneme *'/g/'* in *'good'*

To enter phoneme *'/g/'*, the user first selects the *'Plosives'* group located at the South direction of the front layer of the pie menu. This can be done either by tapping the corresponding slice or by sliding from the center circle towards the South direction. Once the user has entered the *'Plosives'* group, the menu switches to the phoneme layer displaying all phonemes within the group (Figure 1b). The user navigates counterclockwise to access phoneme *'/g/'* using tapping or sliding gestures (see Figure 1c). Auditory and visual feedback is provided to facilitate the navigation process. Once the target phoneme has been found, the user navigates back to the center circle to confirm the selection and switches back to the front layer (Figure 1d).

#### *3.2.1.2 Dynamic Phoneme Layout*

After each phoneme selection, the system recalculates the probability of entry of each phoneme based on the previously entered phonemes and rearranges the phonemes within each group accordingly. The phoneme with the highest probability of entry in each group is chosen as the new representative of the group and placed on the front layer of the pie menu. The remaining phonemes in the group are relocated in such a way that phonemes with higher probabilities are closer to the representative phoneme and hence require fewer movements to navigate to from the representative phoneme. The location of the groups on the front layer, however, remains unchanged, thereby reducing the cognitive overhead associated with dynamic layouts. Figure 1d shows the updated front layer displayed after the user has selected phoneme *'/g/'*. Phoneme *'/oo/'*, the next phoneme in the target word *'good'*, has appeared on the front layer as the new representative of the 'Rounded back vowels' group and can be selected by simply tapping or sliding to the slice at NE direction then moving back to the center circle.

#### *3.2.1.3 Phoneme-based Word Auto-completion*

After each phoneme selection, the system inputs the current phoneme prefix into a basic auto-correction function to generate alternative phoneme prefixes. The auto-correction function employs a limited set of phoneme insertion and replacement rules to deal with the *schwa* phoneme [19] and some common mistakes in phonetic spelling. For example, in our previous study on PA intervention for nonspeaking adults [18], we observed that our participants often had difficulties in distinguishing between phonemes *'/s/'* and *'/z/'* in word-ending position. Thus, we added a rule to generate alternative prefixes by replacing *'/s/'* with *'/z/'* in this position. The alternative prefixes and the original prefix are then used to look up a list of matching words in our pronunciation dictionary. If there is no matching word, the system simply blends the selected phonemes together using a speech synthesizer to generate speech output. Otherwise, the matching words are input into the 3-gram word language model to calculate their probabilities based on up to two prior words. The word with the highest probability is spoken out to the user for selection. If prediction is correct, the user can add the predicted word and a following whitespace to the current sentence by selecting the 'Speak word' function located at the west direction on the front layer (see Figure 1d) and then moving back to the center circle. The phoneme layout is updated thereafter based on the newly added word. If the prediction is incorrect, the user can simply ignore the predicted word and continue entering the next phoneme of the intended word. By offering only one spoken prediction, we eliminated the cognitive and perceptual load imposed on the user to scan a multi-word prediction list to search for the desired word.

### *3.2.2 Computational Experiments*

We evaluated the accuracy of our dynamic phoneme arrangement using hit rate. Hit rate (HR) is defined as the percentage of times that the desired phoneme appears *within* a specified distance D in a group, wherein D is defined as the distance between a phoneme in the group and the group's representative phoneme. The representative phoneme is located at distance D=0 while the two phonemes next to the representative phonemes are located at distance D=1. A phoneme is said to be *within* a distance D if its distance to the representative phoneme is smaller than or equal to D. We calculated the hit rates on three AAC-like test sets:

- *Specialists*: A collection of context specific conversational phrases recommended by AAC professionals[1]. 966 sentences, 3814 words.
- *Communication:* A collection of sentences written by students in response to 10 hypothetical communication situations [21]. 251 sentences, 1789 words.
- *SwitchTest:* Three telephone transcripts taken from the Switchboard corpus, used in Trnka et al. [20]. 59 sentences, 508 words.

For each sentence in the test sets, we generated its pronunciation string using our pronunciation dictionary. During this generation, any time we encountered a word with multiple

[1] http://aac.unl.edu/vocabulary.html, accessed 4 September 2011

pronunciations, we chose a pronunciation at random. We manually added pronunciations for out-of-vocabulary words.

Figure 2 shows the hit rates of the dynamic phoneme layout for distances D=0 and D=1 on the three test sets. At D=0, the average hit rate for the three test sets was 70.8%, which means that the user has a 70.8% chance of seeing the desired phoneme on the front layer of the pie menu. At D=1, the average hit rate increased to 93.2%. This shows that in most cases the intended phoneme is either the first phoneme or next to the first phoneme that the user encounters after entering the phoneme group.

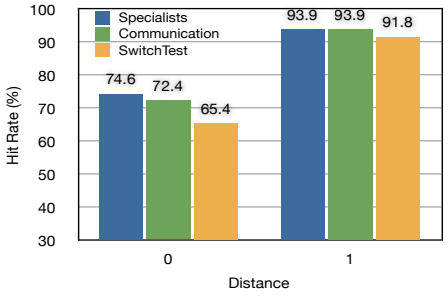



**Figure 2.** Hit rates of the dynamic phoneme layout for distances D=0 and D=1 on the three test sets.

#### *3.2.2.1 Word Auto-completion*

We estimated the accuracy of our word auto-completion feature using hit rate for 1-4 phoneme prefixes. Figure 3 shows the hit rates of word auto-completion on the *Specialists*, *Communication*, and *SwitchTest* test sets. The average hit rate for the three test sets for 1-phoneme prefix was 55.6%, which means that the user has a 55.6% chance of having the intended word autocompleted after entering just the initial phoneme of the word. The average hit rate substantially increased to 80.4% after the first two phonemes are entered, and rose to 90.0% and 96.3% for 3-phoneme and 4-phoneme prefixes respectively.

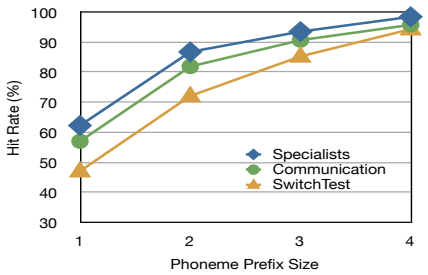



**Figure 3.** Hit rates of the word auto-completion for 1-4 phoneme prefixes on the three test sets.

# 4. FORMATIVE STUDY

To evaluate the usability of our predictive features, we conducted a lab study with able-bodied participants.

## 4.1 Participants and Apparatus

16 university students (9 male, 7 female, aged from 19-42, M=24, SD=5.3) participated study. All participants were native speakers of English with no severe speech, physical, perceptual, or intellectual impairments. Participants were compensated £15.

We developed a prototype of iSCAN on the Apple's iPad 2 providing the users with touch screen access method. Speech output is generated using the CereProc's speech synthesizer. The prototype supports two settings, namely predictive and non-predictive settings. In the predictive setting, the dynamic phoneme layout and word auto-completion features are switched on. The participants select phonemes from the dynamic phoneme layout and are offered one predicted spoken word per entered phoneme. In the non-predictive setting, these two features are switched off. The participants enter phonemes using a static phoneme layout and can hear the blending of all the selected phonemes after each phoneme selection.

To aid the transition between the non-predictive and predictive settings, we used the same starting phoneme layout in both settings, i.e. the phoneme layout is initially optimized based on their probabilities of being the first phoneme in a new sentence. This layout remains unchanged in the non-predictive setting while it is dynamically updated in the predictive setting. As all participants are literate, we decided *not* to associate phonemes with letters in both settings to minimize potential confusion between phonetic spelling and orthographic spelling.

The prototype was instrumented to present a randomly generated set of spoken test phrases to each participant, one at a time, during the experiment. These phrases were short conversational phrases derived from the *Specialists* test set and were pre-recorded using a Scottish English voice. Each phrase consists of 3-5 words (10-12 phonemes). The prototype also contained a logging function to record all user input, including all time-stamped phoneme input and touch information.

## 4.2 Procedure

The study was a within-subjects design and consisted of three sessions. Each session lasted between 30-45 minutes, with at least 2 hours and at most 2 days between sessions, and was videotaped. Sessions 1 and 2 were training sessions and session 3 was the testing session:

**Session 1:** Participants were given instructions on the phoneme layout and the key functionality of the prototype using the non-predictive setting. The phoneme groups were given more 'user-friendly' names when introduced to the participants (e.g. plosive consonants were called 'poppy' sounds). At the end of the session, participants were instructed to create two spoken phrases using the non-predictive setting.

**Session 2:** At the beginning of the session, the participants were instructed to create three spoken phrases using the non-predictive setting. Thereafter, they were introduced to the predictive setting and were instructed to create five spoken phrases spoken by the prototype using this setting.

**Session 3:** Participants were asked to create a set of spoken phrases in both non-predictive and predictive settings. For each setting, they were given one practice phrase and five test phrases. They were instructed to create the phrases as quickly and accurately as possible. After the prototype spoke a phrase, the participants could repeatedly listen to the phrase by tapping a button on the screen. The order of settings was counterbalanced. At the end of the session, the participants took part in a brief interview about the two settings.

## 4.3 Entry Speeds

We measured entry speed in both words per minute (WPM) and phonemes per minute (PPM). Results of the entry speeds (see Table 1) showed that the use of predictive features led to a 108.4% increase in average PPM and a 109.2% increase in average WPM. Data analysis using the repeated measures ANOVA test showed that there was a significant difference between the entry speeds of the non-predictive and the predictive settings, both in terms of PPM ($F_{1,15} = 79.35$, $p <$

.0001, partial $\eta^2$=0.84), and WPM ($F_{1,15}$ = 90.10, p < .0001, partial $\eta^2$=0.86).

**Table 1.** Average phoneme entry rate and average word entry rate for the non-predictive and predictive settings.

| Setting | PPM | | WPM | |
|---|---|---|---|---|
| | Mean | SD | Mean | SD |
| Non-predictive | 11.07 | 2.74 | 3.0 | 0.82 |
| Predictive | 23.07 | 6.07 | 6.29 | 1.67 |

## 4.4 Error Rates

We measured error rate using the phoneme error rate (PER) and word error rate (WER). Results of the error rates (see Table 2) show that the predictive features led to a 79.0% reduction in average PER and a 78.8% reduction in average WER. Data analysis using the repeated measures ANOVA test showed that the use of the predictive features had a significant effect on both PER ($F_{1,15}$ = 16.12, p = .001, partial $\eta^2$=0.52), and WER ($F_{1,15}$ = 12.07, p = .003, partial $\eta^2$=0.45).

**Table 2.** Average phoneme error rate and average word error rate for the non-predictive and predictive settings.

| Setting | PER | | WER | |
|---|---|---|---|---|
| | Mean | SD | Mean | SD |
| Non-predictive | 9.19% | 6.76 | 17.15% | 13.83 |
| Predictive | 1.93% | 1.91 | 3.63% | 4.63 |

## 4.5 Subjective Preferences

At the end of session 3, the participants were asked to provide feedback on the system and specify their preferences for the two settings. Overall, all participants preferred the predictive setting to the non-predictive setting. 13 of 16 participants reported that the dynamic phoneme layout was useful, stating that in most cases they found the intended phonemes on the front layer. 3 of 16 participants, however, commented that this dynamic layout was distracting. 15 of 16 participants stated that the word auto-completion was useful; many of them highly praised the accuracy of this feature. Only one participant found this feature frustrating, stating that it frequently gave her incorrect suggestions. This happened because there were a few instances that the participant chose incorrect phonemes and thus the system repeatedly offered her incorrect predictions without detecting her mistakes. The two features well complemented each other, as all participants liked at least one of the two features. This explains their overall preference for prediction.

In summary, this study demonstrated the usability of iSCAN with a group of able-bodied individuals whose cognitive, physical, and PA abilities might be different from those of our target users. The next step is to test with representative users.

# 5. CASE STUDY

We evaluated the usability of iSCAN in a longitudinal case study with a nonspeaking adult. We had two goals: (1) to compare the usability of the non-predictive and predictive settings, and (2) to investigate whether our system could provide effective communication support for nonspeaking people with limited literacy.

## 5.1 Participant

Our participant, who we will call 'Alex', was a 41-year-old male with severe speech and motor impairments due to cerebral palsy. Results of a literacy and phonological awareness test conducted four months prior to the study confirmed that he has significant spelling difficulties, as he only scored 30% for the spelling real words tasks. Alex demonstrated relatively good phoneme blending and phoneme analysis skills. However, he performed poorly on the phoneme-counting task, which requires him to count the number of phonemes in a spoken word. Prior to the study, he indicated that he had difficulty saying sounds in his head, which suggests that he might have problems with subvocal rehearsal. Alex's cognitive ability was assessed using the Raven's Colored Progressive Matrices test [15] and an adapted version of the Digit Span test from the Wechsler Adult Intelligence Scale-III [23]. Results of these tests revealed that he possibly has working memory deficits.

Alex has been using a 400-word paperboard for more than 30 years as his primary means of communication. He reported infrequent use of voice output communication aids (VOCAs), primarily for telephone conversation and occasionally for group discussion. He has used two VOCAs, including the Say-It! Sam™ communicator and the Assistive Chat application on Apple's iPad. Alex is an experienced prediction user, as he heavily relies on word prediction to generate messages in both of those systems. Prior to this study, Alex had no experience using a phoneme-based system for communication.

## 5.2 Study 1: Predictive vs. Non-predictive

The aim of this study was to compare the usability of the non-predictive and the predictive settings. The study used the same procedure described in our formative study and included two training sessions and one testing session. Each session lasted 45-70 minutes and was videotaped. These sessions were conducted at Alex's home over two consecutive days. On day one, we conducted the first training session. On day two, we conducted the second training session and the final test session, separated by about 2 hours. In the testing session, Alex used the predictive setting first followed by the non-predictive setting. We used the prototype from the formative study, but changed three phoneme pictures based on the user feedback from the formative study. Letters were *not* included in phoneme representations.

Alex completed the transcription task using the predictive setting but not in the non-predictive setting. After creating one practice and one test phrase in the non-predictive setting, Alex expressed a strong preference for the predictive setting and stated that he did not want to proceed with the non-predictive setting. Using the predictive setting, Alex achieved an average entry speed of 6.04 PPM (3.35% PER) and 1.72 WPM (0.0% WER) (note that the difference between PER and WER was due to our auto-correction mechanism described in Section 3.2.1.3). For comparison, Alex's entry speed for the one completed test phrase in the non-predictive setting was 2.35 PPM (21.43% PER) and 0.74 WPM (75.0% WER).

## 5.3 Study 2: Extended Training

To determine whether Alex's entry rate could be improved, we conducted additional sessions with him using the predictive setting to assess his performance after extended hours of practice. Previous research has also reported that long-term use is critical for accurate evaluations of prediction [14].

We conducted 13 additional sessions over an 11-day period using the predictive setting. Each session lasted 20-40 minutes, with at least two hours and at most two days between sessions and no more than two sessions per day. Each session began with

a 5-minute warm-up during which Alex was asked to create his own words and sentences. Thereafter, Alex was asked to transcribe ten test phrases as quickly and accurately as possible. We used the same prototype from Study 1.

***Entry speeds***. Figure 4 shows Alex's improvement in phoneme entry speed over the 13 sessions. His word entry speed followed a similar trend. His speeds for session 1 were 4.45 PPM and 1.22 WPM. On the 13th session, his speeds increased to 18.53 PPM and 4.82 WPM. A speed of 4.82 WPM is not an improvement compared to the frequently cited communication rate of 2-10 words per minute of AAC users. However, it is noteworthy considering the small number of practice hours Alex required to reach this speed.

***Error rates***. Figure 5 shows Alex's phoneme error rates over the 13 sessions. His word error rates followed a similar trend. Overall, his error rates were extremely low as he corrected almost all errors. The average PER over the sessions was 0.66% (SD=0.65) and the average WER was 1.31% (SD=2.23). The effect of our auto-correction mechanism was clearly shown in sessions 2, 3, 7, 8 in which incorrect phonemes were auto-corrected, resulting in 0.0% WER.

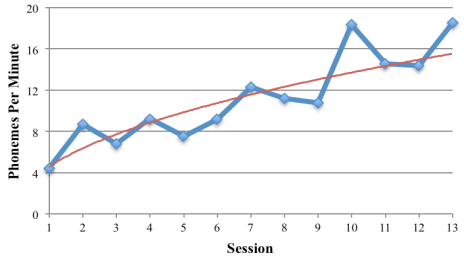



**Figure 4.** Alex's entry speeds as PPM over sessions 1-13.

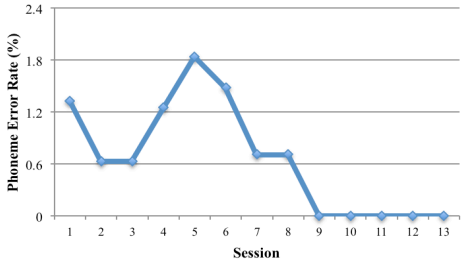



**Figure 5.** Alex's phoneme error rates over sessions 1-13.

## 5.4 Study 3: Comparative Evaluation

In this study we aimed to compare the usability of our phoneme-based predictive system with the two orthographic-based predictive systems that Alex has been using for communication. We evaluated Alex's performance using the Say-It!Sam™ and Assistive Chat systems in a transcription task. Alex has used the Say-It!Sam™ communication device for over 4 years and started using Assistive Chat at around the same time of the commencement of Study 1. Say-It!Sam™ provides 8 word predictions, organized into 2 columns of 4 predictions each, after each character entry. Assistive Chat provides 4 word predictions after each character entry. Both systems offer word predictions even before the first character of a new word is entered. The study consisted of two sessions, each of which lasted 40-60 minutes and was videotaped:

**Session 1:** The session started with a 5-minute warm-up during which Alex was asked to create his own words and sentences using Say-It! Sam™. He was then asked to transcribe a set of 10 test phrases as quickly and accurately as possible. These test phrases were the same test phrases used in the 13th session of Study 2 and were spoken by the iSCAN prototype. This session was conducted approximately two hours after the completion of the 13th session of Study 2.

**Session 2:** The session was conducted 6 days after session 1. Following a 5-minute warm-up, Alex was asked to transcribe 10 test phrases used in session 1 as quickly and accurately as possible using Assistive Chat. At the end of this session, Alex took part in a brief interview in which he ranked Say-It!Sam™, Assistive Chat, and iSCAN in his order of preferences.

The time taken by Alex to enter each test phrase was recorded using a stopwatch and was verified using the video recordings. As this study commenced only two hours after the 13th session of Study 2, we did not conduct a separate session to re-evaluate iSCAN. Instead, we compared the entry speeds and error rates of the Say-It!Sam™ and Assistive Chat with those reported in the 13th session of Study 2.

***Entry speeds.*** We only calculated Alex's average entry speeds as WPM over the 10 test phrases as it was not suitable to measure PPM for Say-It!Sam™ and Assistive Chat. In this task, Alex achieved the following entry speeds: Assistive Chat (M=5.44, SD=3.37), iSCAN (M=4.82, SD=2.63), and Say-It!Sam™ (M=2.78, SD=1.78).

***Error rates.*** Average WER over the 10 test phrases was as follows: iSCAN (M=0.0%, SD=0.0), Assistive Chat (M=11.67%, SD=21.94), Say-It!Sam™ (M=19.17%, SD=20.81).

As Alex has learned orthographic spelling through memorization, he struggled to derive the spellings of unfamiliar words. Therefore, whenever he encountered an unfamiliar word in the test phrases, he either skipped it by choosing a random word from the prediction list or attempted to replace it with a familiar word of similar meanings. This explains his high error rates for the two orthographic-based systems. With iSCAN, however, he has developed a strategy of listening to target words and sounding the words out using his dysarthric speech to identify the target phonemes, rather than relying on memorization. He was also able to confirm whether his phoneme selection was correct by listening to the blending of all selected phonemes. As a result, he showed greater confidence dealing with unfamiliar words using our system and thus attempted to complete all the target words.

***User preference.*** At the end of the study, Alex was asked to rank the three evaluated communication systems in his order of preference. He placed the Say-It!Sam™ last, which was not surprising considering its low entry speed and high error rate. Alex ultimately chose iSCAN over Assistive Chat, stating that he would like to use it for learning new words. This decision can partly be explained by his positive experience using iSCAN to produce many novel words during his 16 sessions. He also reported that difficulties in selecting the intended words from the prediction list in the two orthographic-based systems resulted from his reading problems and thus he preferred our word auto-completion feature.

# 6. CONCLUSIONS & FUTURE WORK

In this paper we have described how prediction techniques can be employed to improve the usability of phoneme-based communication systems. To our knowledge, this is the first empirical research on phoneme-based prediction. We developed

a novel phoneme-based predictive system employing robust statistical language modeling techniques to provide users with single phoneme prediction and phoneme-based word prediction. Results of the evaluations demonstrated that our predictive methods led to significant improvements in user performance, both in terms of entry rate and accuracy. Through a series of studies with a nonspeaking adult, we showed that our phoneme-based predictive communication system had the potential to provide an effective means of generating novel words and messages for a large proportion of AAC users who have literacy difficulties.

We are in the process of analyzing usage data of this participant who has been using our system in the field for four months. Results of this data analysis will allow us to evaluate our predictive methods in real-time spontaneous conversational settings. Our second case study with a nonspeaking female participant who has very limited literacy skills is also underway.

We outline a number of further studies based on this work. First, we plan to investigate whether the use of such a phoneme-based predictive system like iSCAN could have any positive effects on the phonological awareness and literacy development of nonspeaking individuals. Second, results of our longitudinal case study showed that our participant had significant difficulties in identifying vowels in spoken words. Therefore, we aim to explore a more robust auto-correction mechanism to accommodate this issue and facilitate users in vowel selection. Finally, we plan to conduct further empirical studies on how our prediction system can be incorporated into different interfaces and input devices, such as joysticks and eye-tracking systems.

## 7. ACKNOWLEDGMENTS

We thank the Scottish Informatics and Computer Science Alliance and the University of Dundee's School of Computing for funding this project. This work was supported by a Royal Society Wolfson Merit Award, by RCUK EP/G066091/1 "RCUK Hub: Social Inclusion in the Digital Economy", and by EPSRC grant number EP/H027408/1.

## 8. REFERENCES

[1] Black, R., Waller, A., Pullin, G., and Abel, E., 2008. Introducing the PhonicStick: Preliminary evaluation with seven children. In *13th Biennial Conference of the International Society for Augmentative and Alternative Communication*, Montreal, Canada.

[2] Brady, S.A. and Shankweiler, D.P., 1991. *Phonological processes in literacy*. Erlbaum, Hillsdale, NJ.

[3] Brault, M.W., 2008. *Americans with disabilities: 2005 Household Economic Studies.* U.S. Census Bureau Ed., Washington, DC, USA.

[4] Creech, R., 2004. Rick Creech, 2004 Edwin and Esther Prentke AAC Distinguished Lecturer, ASHA Convention Ed., Philadelphia, USA.

[5] Foley, B.E. and Pollatsek, A., 1999. Phonological processing and reading abilities in adolescents and adults with severe congenital speech impairments. *Augmentative and Alternative Communication 15*, 156-173.

[6] Garay-Victoria, N. and Abascal, J., 2005. Text prediction systems: a survey. *Universal Access in the Information Society 4*, 188-203.

[7] Glennen, S.L. and DeCoste, D.C., 1997. *The Handbook of Augmentative and Alternative Communication*. Thomson Delmar Learning.

[8] Goodenough-Trepagnier, C. and Prather, P., 1981. Communication systems for the nonvocal based on frequent phoneme sequences. *Journal of Speech and Hearing Research 24*, 322-329.

[9] Goodenough-Trepagnier, C., Tarry, E., and Prather, P., 1982. Derivation of an efficient nonvocal communication system. *Human Factors: The Journal of the Human Factors and Ergonomics Society 24*, 2, 163-172.

[10] Hanson, V.L., Goodell, E., and Perfetti, C., 1991. Tongue-twister effects in the silent reading of hearing and deaf college students. *Journal of Memory and Language 30*, 319-330.

[11] Koester, H.H. and Levine, S.P., 1996. Effect of a word prediction feature on user performance. *Augmentative and Alternative Communication 14*, 25-35.

[12] Light, J. and McNaughton, D., 2009. Addressing the literacy demands of the curriculum for conventional and more advanced readers and writers who require AAC. In *Practically Speaking: Language, Literacy, and Academic Development for Students with AAC Needs*, G.S.C. ZANGARI Ed. Paul H. Brookes Publishing Co, Baltimore, MD, 217-245.

[13] Lloyd, S.M., 1998. *The Phonics Handbook*. Jolly Learning Ltd., Chigwell.

[14] Magnuson, T. and Hunnicutt, S., 2002. Measuring the effectiveness of word prediction: The advantage of long-term use. *Speech, Music, and Hearing 43*, 57-67.

[15] Raven, J. and Court, J.H., 1998. *Manual for Raven's progressive matrices and vocabulary scales*. Oxford Psychologists Press Ltd., Oxford, UK.

[16] Schroeder, J.E., 2005. Improved spelling for persons with learning disabilities. In *The 20th Annual International Conference on Technology and Persons with Disabilities*, California, USA.

[17] Smith, M., 2005. *Literacy and augmentative and alternative communication*. Elsevier Academic Press.

[18] Trinh, H., 2011. Using a computer intervention to support phonological awareness development of nonspeaking adults. In *The 13th International ACM SIGACCESS Conference on Computers and Accessibility*, Dundee, UK.

[19] Trinh, H., Waller, A., Vertanen, K., Kristensson, P.O., and Hanson, V.L., 2012. Applying prediction techniques to phoneme-based AAC systems. In *NAACL-HLT 2012 Workshop on Speech and Language Processing for Assistive Technologies (SLPAT)*, Montreal, Canada.

[20] Trnka, K., McCaw, J., Yarrington, D., McCoy, K.F., and Pennington, C., 2009. User interaction with word prediction: The effects of prediction quality. *ACM Transactions on Accessible Computing 1*, 3, 1-34.

[21] Venkatagiri, H.S., 1999. Efficient keyboard layouts for sequential access in augmentative and alternative communication. *Augmentative and Alternative Communication 15*, 2, 126-134.

[22] Vertanen, K. and Kristensson, P.O., 2011. The imagination of Crowds: Conversational AAC language modelling using crowdsourcing and large data sources. In *International Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Edinburgh, UK, 700-711.

[23] Wechsler, D., 1997. *WAIS-III administration and scoring manual*. Psychological Corp.

[24] Williams, M.B., 1995. Transitions and transformations. In *9th Annual Minspeak Conference* Prentke Romich Company, Wooster, OH.