

RESTRICTING VOCABULARY SIZE IN PEDIATRIC AUGMENTATIVE AND ALTERNATIVE COMMUNICATION

Alexander Moreno¹, Christopher Rozell², Ayanna Howard²
Schools of Computer Science¹ and Electrical Engineering², Georgia Institute of Technology, Atlanta, GA

ABSTRACT

We address improving communication rate for non-verbal children using icon-based Augmentative and Alternative Communication (AAC). Non-verbal children with motor disabilities, when acquiring language skills, may have slow input using traditional single-key (switch) coded input methods and may be deficient in basic spelling skills, increasing the difficulty in using keyboard text entry. Reducing keystrokes in an icon setting could thus improve a non-verbal child's ability to communicate more effectively and improve their basic vocabulary and grammar skills. Since young children have a developing vocabulary, using a restricted vocabulary appropriate to oral communication should reduce the keystrokes required for the sentences that can be completed using an AAC device. While a balance between the selection of the vocabulary and providing ease-of-access is important, for younger children with motor disabilities, there should be a bias toward access in order to maximize potential for language skill acquisition. In this paper, we discuss a method to restrict words in pediatric-based AAC and evaluate the average keystrokes saved for completion of sentences using those words on several AAC test sets. We show a keystroke savings of 80% vs a full vocabulary on one AAC test set and a savings of 21% on another using a similar method, with the percent of completable sentences being 41% and 32%, respectively.

INTRODUCTION

Augmentative and Alternative Communication (AAC) describes communication methods to help individuals who are unable to use verbal speech. Non-verbal children with motor disabilities typically use AAC stand-alone devices or tablet apps for alternative language

development. However, many of these children, at younger ages, lack basic spelling skills and have difficulties using keyboard text entry. They generally use a switch with linear scanning as their device input, allowing them to linearly move from one icon (pictorial symbol) to the next until they reach the icon representing the desired word. Speeding up their communication rate by saving on the number of switch keystrokes, or icons moved to between selections, used in order to make sentences could help children with motor disabilities communicate more effectively and allow their parents or care-givers to help them develop a wider vocabulary and grammar.

AAC research involving young children is difficult because, due to ethical concerns, there is no large training corpus. Additionally, the interface is typically icon-based, which is traditionally slower than keyboard text entry. For icon-based AAC, there are two naive approaches. One is to list all words by an ordered probability of occurrences of each word. As the vocabulary size increases, this becomes inefficient, correlating with decreased efficiency for sentence completion using switch-based input methods. The second approach, used by most icon-based AAC apps, uses a hierarchy where there are a few common words and category symbols on the main screen that, when selected, bring up more words corresponding to those parts of speech. There is generally some kind of probability-based ordering of the words (typically derived by speech-language pathologists) within the hierarchies.

Most academic approaches to AAC focus on improving word prediction for keyboard text entry for adults, as this has been shown to increase communication rates (Trnka, Yarrington, McCaw, McCoy, & Pennington, 2007). Keyboard prediction has the advantage over an icon interface that one can restrict the

vocabulary based on the first few letters when predicting the next word. For instance, typing "I am h," "I am ok" gets eliminated and "I am here" goes up in probability. AAC research papers have adjusted to the lack of training corpus by using alternative, but similar, data sets. Since training on data with a very different distribution from the test set will lead to poor test set predictions, researchers usually extract a subset from non-AAC corpuses to train and then use small AAC test sets to evaluate. Example techniques using this practice include topic modeling (Trnka, Yarrington, McCoy, & Pennington, 2006) and using Mechanical Turk workers to create hypothetical AAC sentences and finding sentences from the training data that are similar to those hypothetical sentences (Vertanen & Kristensson, 2011).

These papers generally look to lower keystrokes or perplexity on full adult AAC test sets. Perplexity measures how well, on average, the model predicts test sentences, with smaller values corresponding to better predictions. However, young children have not acquired a vocabulary size equivalent to an adult, so measuring the average keystrokes or perplexity on a full test set isn't appropriate. In a pediatric setting, what matters is minimizing the keystrokes required for simple sentences using only the words the children know.

APPROACH

In (Vertanen & Kristensson, 2011), an AAC corpus was constructed by having Amazon Mechanical Turk workers create hypothetical AAC sentences, and dividing these sentences into a training set (TurkTrain), a development set (TurkDev), and a test set (TurkTest). Word prediction performance improves with corpus size, but creating a large corpus using mechanical turk workers is financially infeasible. Three training sets: twitter, usenet, and blogs, were then used to select sentences similar to the TurkTrain sentences by varying the cross-entropy (measures distance between distributions) difference threshold to optimize the perplexity (how well the model predicts test sentences) on TurkDev. They trained on the remaining twitter, usenet, and blog sentences separately, and then created a mixture model,

which combines multiple models together, with weights chosen again to optimize perplexity on TurkDev. This mixture model, which they called the trigram mixture model data set, was made publicly available and included all vocabulary along with unigram, bigram, and trigram probabilities for the model that optimized the weights and thresholds for perplexity, and they showed how this model performed in terms of perplexity on both AAC test sets and TurkTest. Unigram probabilities are single word probabilities, while bigram and trigram probabilities are conditional probabilities of a word given the previous one or two words, respectively. To apply this model to icon-based AAC, we restrict the publicly available test set to X words, where X is varied as discussed in the experiment section below. We then take the most frequent words found in the training set by unigram probability from the training data and delete all unigrams, bigrams, and trigrams that contain words other than these top X words.

We assume the initial layout of the AAC interface is constructed as icons associated with words ordered by their bigram probabilities given a symbol that indicates the start of a sentence. After this, word ordering shifts based on the trigram probabilities described above. This involves calculating the trigram probabilities for every possible word in the vocabulary, sorting them, and then displaying the words. For example, figure 1 shows potential words displayed after starting a sentence with "I."

```
i  
['am', 'need', 'don't', 'love', 'think', 'want', 'have', 'will',  
'hope', 'can't', 'wish', 'just', 'feel', 'can', 'like', 'would',  
'hate', 'miss', 'really', 'know', 'was', 'had', 'do']
```

Figure 1: words displayed after selecting "I"

METHODOLOGY

To evaluate the benefits and tradeoffs associated with using a restricted vocabulary set for icon-based AAC, and the use of the trigram versus unigram model, we derive two sample restricted vocabulary sets from the full test sets and compare results.

Test Sets

We used two test sets compiled in

(Vertanen & Kristensson, 2011) as described below.

- **COMM:** Sentences and vocabulary words were collected that were associated with possible communication scenarios. A total of 251 sentences and 1789 words are found in the COMM AAC data set.
- **SWITCHTEST:** Sentences and vocabulary words were compiled from three different switchboard conversation transcripts. A total of 59 sentences and 508 words are found in the SWITCHTEST AAC data set.

The original trained adult model had a 63K word vocabulary. We restricted this to between 1K words and 46K words at intervals of 5K. For each test set (COMM and SWITCHBOARD), we calculated the percent of sentences that could be completed using the restricted vocabulary, and extracted those sentences. For these sentences, we looked at the perplexity and average keystrokes for a unigram model, the restricted vocabulary trigram model, and the full vocabulary trigram model. Keystrokes are counted as one keystroke to select a word, and one keystroke to move to the next word.

RESULTS

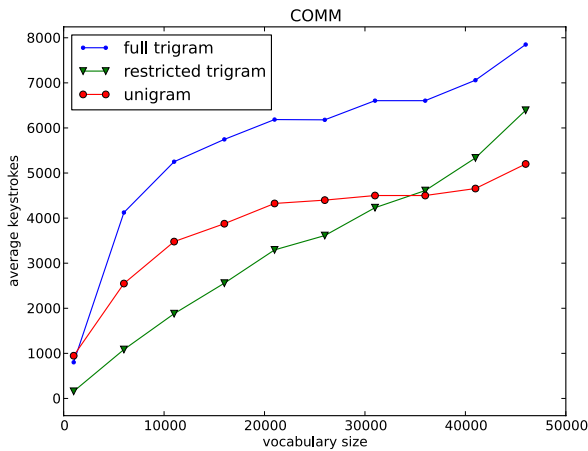


Figure 2: COMM sentences. How number of keystrokes for each model changes as we scale vocabulary size.

Figure 2 shows how the average keystrokes to complete sentences vary as we increase the vocabulary size for the COMM test set. For 1K words, the restricted trigram model requires an average of 158 keystrokes, the full model 802,

and the unigram model 947. Defining keystroke savings as

$$\left(1 - \frac{K_{new}}{K_{old}}\right) \times 100\%$$

where K_{old} is the keystrokes of the original model and K_{new} the keystrokes of the new model, we get 80% keystroke savings over using the full vocabulary and 83% keystroke savings over a unigram model for sentences completable using 1K words. For 46K words, the restricted model averages 6392 keystrokes, the full model averages 7850, and the unigram model averages 5201.

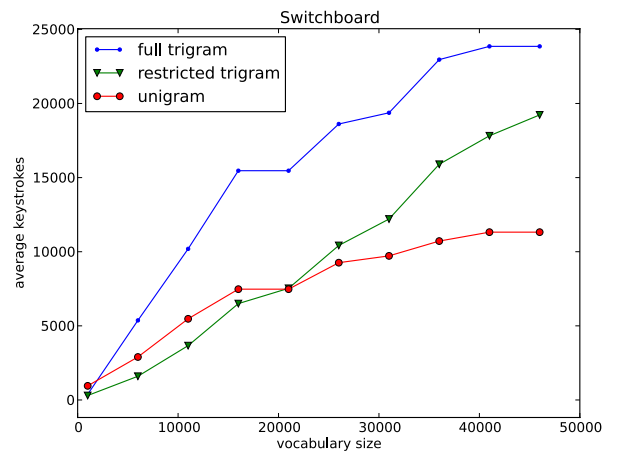


Figure 3: Switchboard Conversations

Figure 3 shows the same for the switchboard conversations. For 1K words, the restricted trigram model averages 300 keystrokes, the full model 382, and the unigram model 952. Thus, the restricted model has 21% keystroke savings over the full model and 68.5% savings over the unigram model. For 46K words, the restricted model averages 19228, the full model 23852, and the unigram model 11320.

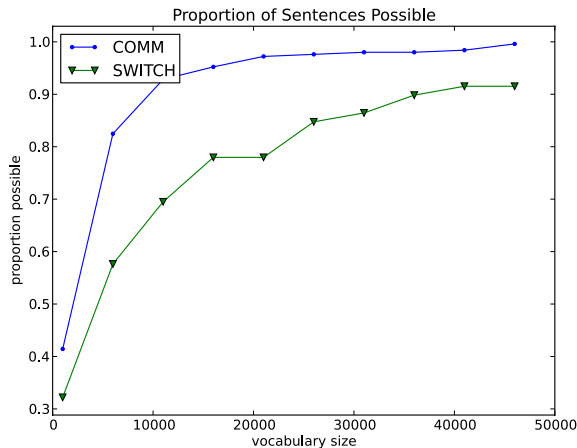


Figure 4: Sentences Completable

Figure 4 shows how the proportion of sentences that can be completed grows as we increase the vocabulary size for each test set. Both plots show a sharp initial increase that greatly slows down from 6000 to 11000 words and continues to do so. The vocabulary for the training data can complete a greater proportion for COMM than for SWITCHTEST for every vocabulary size tested.

DISCUSSION

Using a trained trigram model, we attempted to mimic AAC communications for children and examined the effects of word restriction in an icon-based AAC setting. Specifically, what percent of sentences could we complete in AAC test sets, and how does restriction affect average keystrokes required to complete these sentences? Restriction leads to large keystroke savings, which increased as we tightened it down to 1K words. This makes sense, as the user does not have to cycle through as many words to complete simple sentences. Less obviously, for large vocabulary sizes, unigrams outperformed both other models. From inspection, there were pathological cases for the trigram model, such as having words repeat themselves in the test sets, like "an an."

Icon-based AAC interfaces often have a vocabulary of thousands of words. For example, two commonly used icon sets, the Widgit Symbol Set and the Mayer-Johnson Picture Communication Symbol collection, each contain approximately 11,000 icons. At 11,000

icons, this work suggests that the keystroke savings would be 45% and 33% (vs naïve unigram layout) with a sentence completion rate of 93% and 69% on COMM and SWITCHTEST, respectively. This represents an upper bound on the vocabulary set and may not be an ideal representation for young children with motor disabilities who are still early in acquiring language. As such, thought should be given to the balance between the percent of sentences that can be completed versus the number of keystrokes required to complete those sentences.

Future Work

Many icon-based AAC apps use variations on a speech hierarchy to organize icons and words. There may be 5-10 common words on the first screen such as "I" and "am," along with "verbs," "adjectives," "pronouns," and other parts of speech leading to subsequent screens. It's important to evaluate whether the benefits of word restriction generalize to this hierarchical setting, and if so, how different restriction levels affect the average keystrokes versus sentences that can be completed in a hierarchical interface. A hierarchy along with prediction and a tight word restriction would likely lead to low enough average keystrokes to be useful for children with motor disabilities.

Finally, to be sure of the benefits of prediction in this context, a user study would be helpful. In particular, constantly shifting the ordering of words may be more confusing to the user in an icon based setting than in a keyboard based one.

REFERENCES

Beukelman, D. R., Yorkston, K. M., Pobleto, M., & Naranjo, C. (1984). Frequency of Word Occurrence in Communication Samples Produced by Adult Communication Aid Users. *Journal of Speech and Hearing Disorders*, 49(4), 360-367.

Garay-Vitoria, N., & Gonzalez-Abascal,

J. (1997, January). Intelligent word-prediction to enhance text input rate (a syntactic analysis-based word-prediction aid for people with severe motor and speech disability). In *Proceedings of the 2nd international conference on Intelligent user interfaces* (pp. 241-244). ACM.

Glennen, S., & DeCoste, D. C. (Eds.). (1997). *The handbook of augmentative and alternative communication*. Cengage Learning.

Heath, A. P., Powell, R., Suarez, R. D., Wells, W., White, K., Atkinson, M., ... & Mambretti, J. J. (2012, November). The Design of a Community Science Cloud: The Open Science Data Cloud Perspective. In *High Performance Computing, Networking, Storage and Analysis (SCC), 2012 SC Companion:* (pp. 1051-1057). IEEE.

Reichle, J., York, J., & Sigafos, J. (1991). *Implementing augmentative and alternative communication: Strategies for learners with severe disabilities*. PH Brookes Publishing Company.

Trnka, K. (2010). *Word prediction techniques for user adaptation and sparse data mitigation* (Doctoral dissertation, University of Delaware).

Trnka, K., Yarrington, D., McCaw, J., McCoy, K. F., & Pennington, C. (2007, April). The effects of word prediction on communication rate for AAC. In *Human*

Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Companion Volume, Short Papers (pp. 173-176). Association for Computational Linguistics.

Trnka, K., Yarrington, D., McCoy, K., & Pennington, C. (2006, January). Topic modeling in fringe word prediction for AAC. In *Proceedings of the 11th international conference on Intelligent user interfaces* (pp. 276-278). ACM.

Vertanen, K., & Kristensson, P. O. (2011, July). The imagination of crowds: conversational AAC language modeling using crowdsourcing and large data sources. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (pp. 700-711). Association for Computational Linguistics.

Mayer-Johnson. Picture communication symbols collections. Retrieved January 19, 2015, from <http://www.mayer-johnson.com>

Widgit. About symbols. Retrieved January 19, 2015, from <http://www.widgit.com>