RSVP-iconCHAT:

A Single-Switch, Icon-Based AAC Interface

Karl Wiegand, Rupal Patel

Northeastern University

wiegand@ccs.neu.edu, r.patel@neu.edu

Abstract

Augmentative and alternative communication (AAC) systems are often used by people with speech impairments severe enough to preclude the use of spoken communication. Due to concomitant mobility and language constraints, many of these individuals use icon-based systems controlled by switches. Single-switch AAC systems are typically simplifications of multi-array AAC systems that share elements of the same interface layout, but support some form of scanning, such as linear or row-column. From a development standpoint, this conversion technique makes it easy to transform almost any AAC system into a single-switch system; however, it means that most of these systems were originally designed for users with much greater mobility. The purpose of this work was to design an icon-based AAC interface specifically for use with binary signals, such as switches. A usability study was conducted with both non-disabled adults as well as adults with speech and mobility impairments to determine performance bounds and observe individual use cases. Results indicated similar learning curves for both groups and promising performance characteristics for the target population. These results have immediate applications to the design of icon-based AAC and implications for mobile, icon-mediated communication platforms.

Keywords

AAC, RSVP, assistive, binary, communication, icons, semantic frames, switch

Introduction

Many individuals with speech impairments severe enough to preclude spoken communication also have concomitant physical impairments that limit the use of sign language or written forms of communication (Beukelman and Ansel). Augmentative and alternative communication (AAC) systems, which range from letter boards to dedicated speech output devices, are the primary way for these individuals to convey their thoughts, needs, and desires to those around them. Although letter-based systems allow users to construct fully generative utterances, they can be slow and fatiguing (2 – 5 words per minute) because of the number of selections that must be made to complete a message (Todman). Additionally, many individuals who have sustained impairments since birth have poor or limited literacy skills (Beukelman and Mirenda). Icon-based AAC systems are advantageous because they have the potential to increase message construction speed by allowing whole words and concepts to be accessed in a single keystroke (Todman, Alm, and Elder). They can also be used by children and other users with limited or emerging literacy skills.

Depending upon their mobility impairments and language constraints, many AAC users require icon-based systems controlled by switches (Wolpaw et al). Current single-switch AAC systems are typically simplifications of multi-array AAC systems and display a complex array of vocabulary on the screen, organized into a navigation hierarchy based on categories. To increase the size of the vocabulary on these systems, the screen size must be increased, the button sizes must decrease, or the navigation hierarchy must become more complex. Each of these approaches has its limitations. An increased screen size reduces mobility of the system, smaller buttons are more difficult to view, and complicated navigation hierarchies require more time and effort to find the target button and increase the likelihood of confusing the user. Additionally, when these interfaces are used with scanning, users must visually locate their target button from among the many options on the screen. People who use single-switch AAC systems often have extremely limited physical mobility or control, making it difficult to repeatedly perform the necessary head, neck, or eye movements when attempting to locate target items (Muller and Blankertz).

Approach

Our single-switch AAC interface, called RSVP-iconCHAT, aims to minimize the amount of head, neck, and eye movements required to efficiently control the system. RSVP-iconCHAT was designed to be robust enough to function with a brain-computer interface (BCI), as well as conventional access methods, such as sip-and-puff devices, eye-blink detectors, surface electromyography (EMG), or physical switches. To that end, we leverage a technique called Rapid Serial Visual Presentation (RSVP), in which the user fixates on a relatively stable location while different images are displayed in that location, one at a time. RSVP originates from the field of psychology and has been used successfully to control letter-based AAC systems (Orhan et al).

To leverage RSVP, our interface focuses the user's attention on the message being constructed instead of displaying all of the available vocabulary. To demarcate different visual fixation areas, messages are represented as semantic frames. Semantic frames are a product of case grammar theory, which asserts that the main action, or verb, is the central component of a message (Fillmore). Each message can be expressed as a formulaic frame for which certain semantic roles are understood and expected. For example, the frame for the verb "to give" might require, at a minimum, an actor that does the giving, a participant that receives the gift, and an object that can be given or received. Semantic frames can be used to constrain relevant roles for a given action, and these roles can then be populated with appropriate concepts to generate complete utterances.

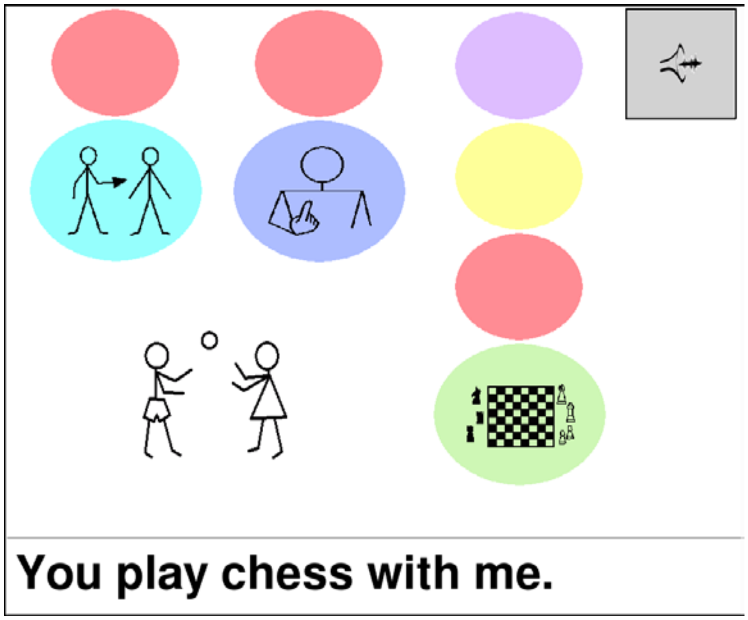

Fig. 1. Functional elements of the RSVP-iconCHAT interface, including: command field (upper right), message in standard grammatical form (bottom), and schematic layout of semantic roles.

In RSVP-iconCHAT, each message is subdivided into semantic roles (e.g. actor, action, participant, and object) and applicable vocabulary options are displayed using RSVP within each semantic role (Fig. 1). This design uncouples required screen real estate and physical movement from the vocabulary size and instead ties them to the length of the desired message. For more advanced or more mobile users, the number of available semantic roles can be increased, enabling users to create longer and more complex messages; for beginning users, or those with severely reduced mobility, the number of roles can be decreased to enable the creation of simpler messages with the same vocabulary.

Fig. 2. Subset of single-action picture cards.

To construct a message using the RSVP-iconCHAT approach, users first select the desired verb or action. Once an action has been selected, the corresponding semantic frame is displayed with semantic roles such as actor, actor modifier, participant, participant modifier, object, object modifier, quantity, and possessives. These roles are displayed as a set of fillable slots that are spatially organized around the verb. Each semantic role is then highlighted sequentially. Once a role has been selected, icons that can fulfill that role are displayed via RSVP and users can select a desired icon to populate the role. After an icon has been selected for a given semantic role, other roles are highlighted sequentially to allow users to populate as many roles as desired, and in any order. Articles (e.g. “a,” “an,” “the”) and prepositions (e.g. “in,” “of,” “to”) are automatically inserted to efficiently generate grammatically complete messages. At any point during message construction, users can select the “command field” to perform conversational actions (e.g. “speak” or “clear” the current message). Selecting a “speak” command, for example, might send the message to an integrated Text-to-Speech (TTS) system, clear the current message, and prompt the user to begin constructing a new message.

Method

We conducted a usability study involving a constrained message elicitation task for the purpose of exploring how potential users would interact with and respond to the interface. After a brief demonstration and training period, participants were shown a series of 30 picture scenes depicting simple actions (e.g. a boy drinking milk, a man combing his hair, and a woman reading a book) and asked to use our prototype RSVP-iconCHAT interface to create a sentence describing each scene (Fig. 2). The order of the picture scenes was randomized across participants in order to observe behavior as users became more familiar with the system. Participants were directed to construct sentences that were as detailed as necessary such that, if the picture cards were shown to another person, that person would be able to match the appropriate description with the scene.

Each experimental session was conducted in one 60 – 90 minute block per participant, and all sessions were conducted in a sound-treated acoustic booth. Each participant was seated in a chair, or personal wheelchair, facing a computer screen. The space bar of a standard QWERTY keyboard was designated as the switch mechanism, and the RSVP process was configured to show images in alphabetical order using a timing mechanism with a starting speed of 700 milliseconds per image; however, participants could increase or decrease the speed in increments of 100 milliseconds per image. Participants were encouraged to change the RSVP speed whenever and however they preferred, either by pressing the up and down arrows on the keyboard or by requesting it verbally. The icon set, or vocabulary size, consisted of 106 items preselected for their relevance to the picture scenes and tagged within each of 8 possible semantic roles. After each session, participants were asked to provide qualitative feedback via an informal interview.

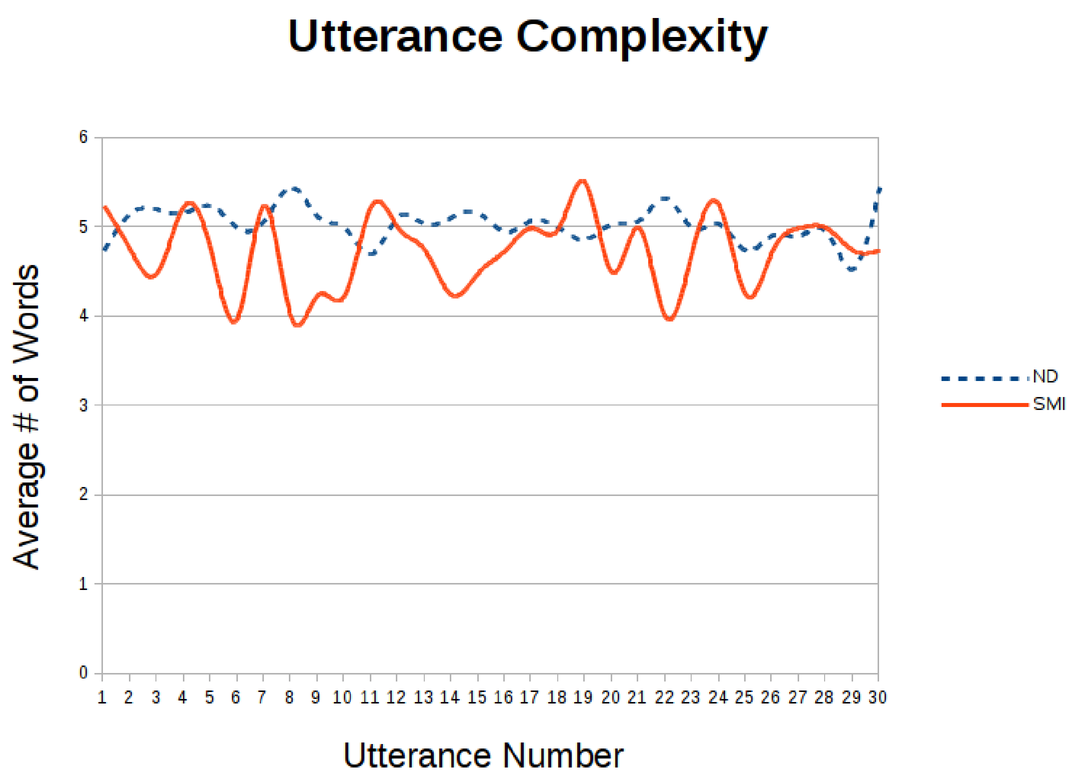

Fig. 3. Average sentence length, excluding the verb and automatically added function words.

Two groups of users were recruited: non-disabled (ND) users to provide a theoretical upper bound on performance, and users with speech and motor impairments (SMI) to provide a realistic evaluation from the target population. For the group of ND users, we recruited 24 English-speaking adults from the greater Boston area, with no declared speech, language, hearing, or cognitive impairments (10 males and 14 females; mean age 24 years; age range 19 – 43). On average, each of these participants had approximately 3 years of formal education following high school and spent approximately 11 hours per week using a computer. The ND users did not have prior exposure or experience with AAC devices.

For the group of users with SMI, we recruited 4 additional English-speaking adults from the greater Boston area (2 males and 2 females; mean age 41; age range 33 – 56). On average, each participant had approximately 4 years of formal education following high school and spent approximately 15 hours per week using a computer. Two of these participants (P1 and P2) had mild motor impairments; two (P3 and P4) had moderate-to-severe motor impairments. All of these participants used wheelchairs, except for P1 who used a walker. P1 and P2 had experience with AAC devices, but used unaided communication on a normal basis. P3 used both unaided communication and switch-based AAC. P4 was unable to use existing AAC systems and required the assistance of a caregiver to communicate.

Results

Theoretically, the open-ended design of the task allowed for the possibility of users creating nonsensical sentences; however, in practice, there were no such instances. Because our prototype implementation required that every sentence contain at least a verb, the short possible sentence was one word in length. On average, both the ND participants and the participants with SMI created sentences consisting of 5 words, excluding articles and prepositions that were automatically inserted by the system (Fig. 3). Thus, users selected a verb and an average of 4 additional icons to construct descriptions of each picture scene. In fact, the participants with SMI created slightly more complex sentences, up to 6 additional words, on at least 2 occasions throughout the study.

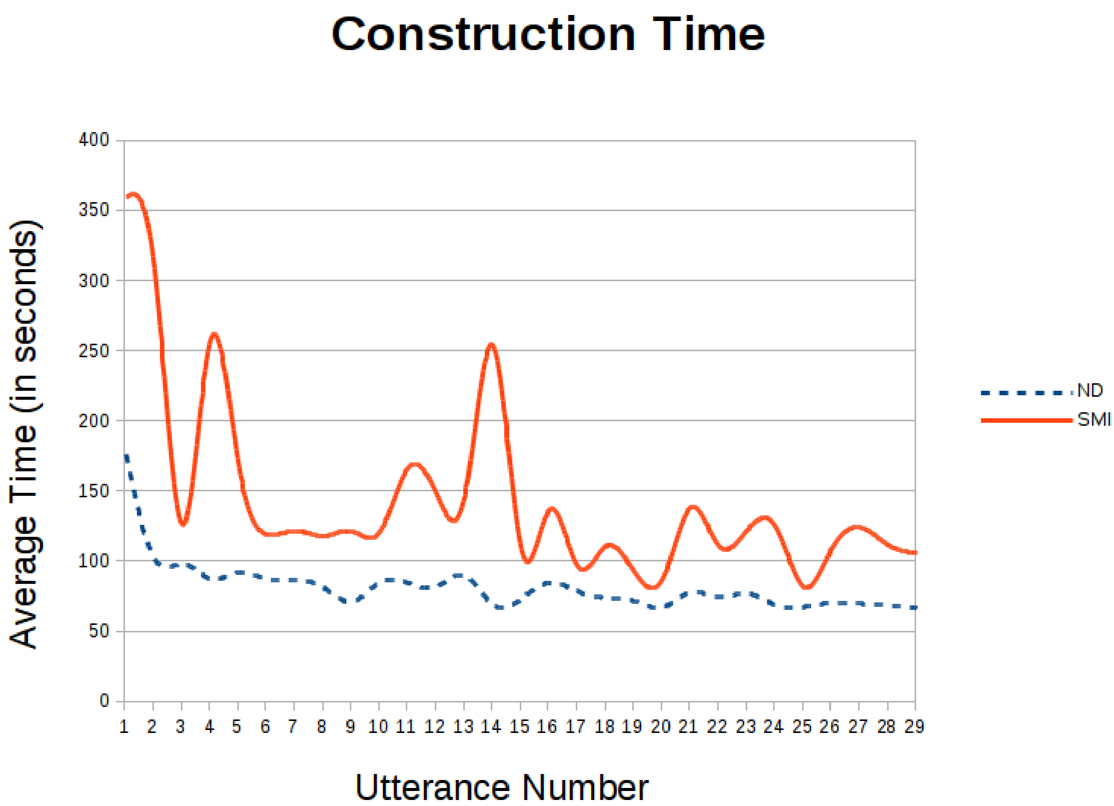

Fig. 4. Average message construction time per sentence.

In terms of message construction speed, both groups of users showed similar learning curves, with the ND group achieving a final speed approximately 1.5 times faster than the group with SMI (Fig. 4). The average time for constructing each of the last five sentences was 70 seconds for the ND users and 107 seconds for the users with SMI.

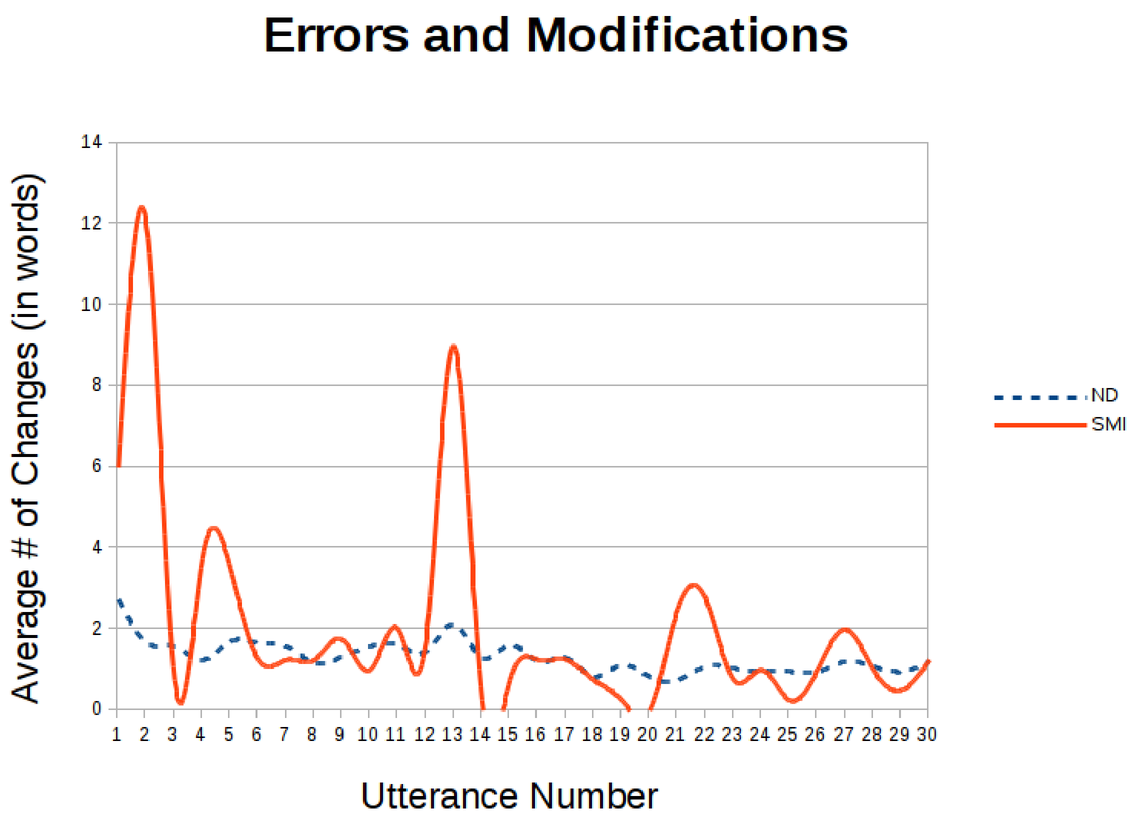

Fig. 5. Average number of self-corrections per sentence.

If users populated a semantic role more than once, even if they selected the same icon or cleared the role of any value, it was considered a self-correction. This metric was used to probe fatigue and learnability of the system. On average, ND users changed or deleted 1 word per sentence before submission, compared to an average of 2 word changes or deletions per sentence for the participants with SMI (Fig. 5).

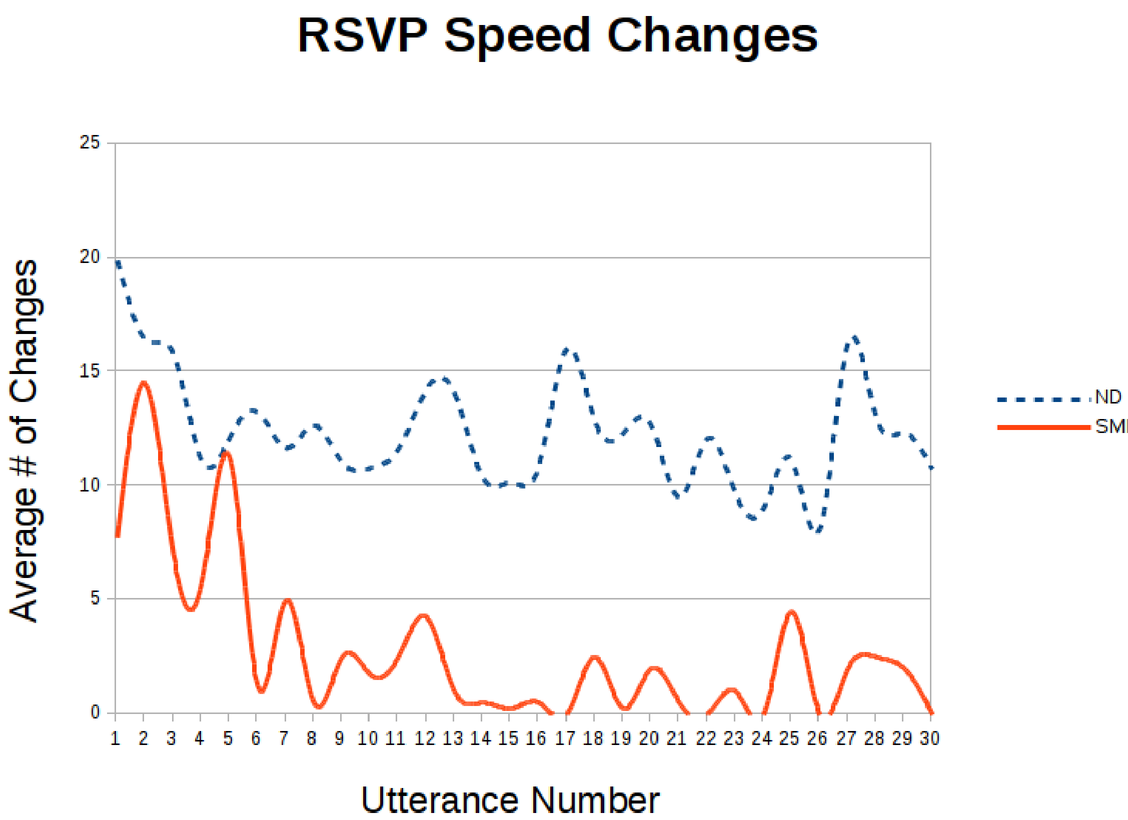

Fig. 6. Average number of RSVP adjustments per sentence.

During the study, ND users adjusted the RSVP speed an average of 10 times per sentence (Fig. 6), returning to an average ending speed of approximately 700 milliseconds per image (Fig. 7). In contrast, users with SMI adjusted the RSVP speed an average of 9 times per sentence for the first 5 sentences and an average of once per sentence for the remaining 25 sentences, returning to an average ending speed of 1200 milliseconds per image.

Discussion

This study examined user behavior while composing messages with the RSVP-iconCHAT interface and a single switch mechanism. The aim was to assess the learnability and ease-of-use of the system. Icon-based message construction via RSVP proved to be learnable within less than 30 minutes for both user groups. Users were able to construct messages of 4 – 7 words in approximately 1 minute, which is faster than some traditional letter-based systems (Wolpaw et al), but users were unable to surpass the performance of conventional icon-based systems (Muller and Blankertz).

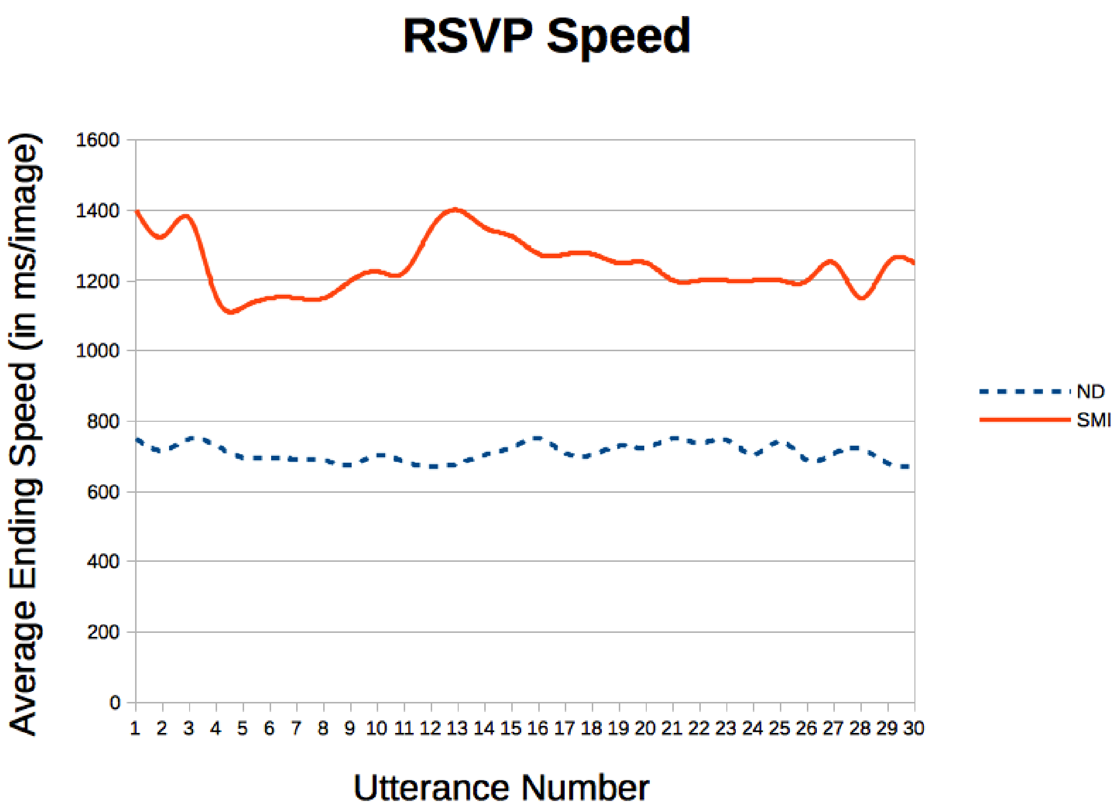

Fig. 7. Average ending RSVP speed per sentence.

The results of our study suggest that expressiveness and generativity are not necessarily compromised by limiting selection tasks to a single key. In fact, both user groups constructed relevant sentences that were an average of 5 selected words in length. Examples of constructed messages included: “an old woman knitting a sweater,” “a small child drawing a house,” and “a man talking on a blue telephone with his friend.” Although this study did not replicate the social pressures of realistic conversation rates, these sentences are longer and more complete than the simple 2 – 3 word sequences documented using some traditional icon-based systems (Van Balkom and Welle Donker-Gimbrere).

Frequently changed RSVP speed throughout the course of the study suggests that users may have been unsatisfied with a constant presentation speed and may have wanted to skip ahead to specific roles or icons. For ND users, the average of 10 changes per sentence suggests that participants increased the speed 5 times and then decreased the speed 5 times, possibly to skip through a large number of undesirable words; however, this behavior was not displayed by the users with SMI (Fig. 6). Although it is possible that the users with SMI found a comfortable speed within the first few sentences, it may have also required too much effort to change the RSVP speed more often, especially for those with moderate-to-severe motor impairments.

Users with SMI appeared to prefer an RSVP speed approximately 1.7 times slower than ND users, yet it is possible that they may be comfortable with faster RSVP speeds for other input modalities. For example, two users (P3 and P4) indicated they could have constructed messages more quickly if the interface were integrated with a sip-and-puff device. Given that both ND participants and participants with SMI converged to consistent ending RSVP speeds, their respective presentation rates (Fig. 7) may be appropriate defaults for physical input modalities, such as button presses.

Two of the participants with SMI (P1 and P2) explored almost the entire vocabulary approximately halfway through the experiment. Additionally, a spike in self-correction for these users, at approximately sentence 13 (Fig. 5), may indicate that they were exploring more expressive possibilities and testing the boundaries for sentence complexity. This phenomenon was not observed with ND users, possibly indicating different preferences between the two groups when familiarizing themselves with new communication interfaces. Self-corrections may also be explained by mistaken selection of a word due to slow motor movement, which would have been a sustained problem for the users with SMI, even as familiarity with the system increased.

While quantitative measurements of fatigue or cognitive load were not collected, qualitative feedback indicated that ND users felt “fidgety” and “impatient” at having to wait for a desired icon to be displayed, but almost all users commented that the interface was “simple” and “easy to use.” One user with SMI (P1) also expressed impatience at having to wait for the target icon; however, the other three users with SMI did not indicate any similar frustration. All 28 participants noticed and favorably commented on the fact that the RSVP-iconCHAT approach did not require them to capitalize words, conjugate verbs, or provide articles and prepositions. Two of the users with SMI (P1 and P3) remarked that they had not seen an existing AAC system with similar functionality, and several ND users asked if there were a way to enable this feature in their current mobile devices.

Conclusion

Many individuals with severe speech and motor impairments use icon-based AAC systems with switches; however, these systems often require larger screens, use complex navigation hierarchies, or necessitate repetitive head, neck, and eye movements. We aimed to design an alternative to conventional icon-based AAC systems that would require less screen real estate, yet still be easy to navigate and allow for sufficiently large vocabularies. RSVP was leveraged to display icons and reduce the required motor control to a single action. Furthermore, RSVP was combined with semantic frames to segment the screen into multiple fields and place the burden of search on the system rather than the user. By organizing vocabulary into semantic roles, rather than lexical categories, the display requirements of this approach are not tied to vocabulary size, but to the number of semantic roles necessary to construct a desired message.

The usability study suggests that an RSVP approach to icon-based message construction is a viable option for users with severe speech and motor impairments. Given that both cohorts of study participants were unfamiliar with the RSVP-iconCHAT approach, their performance should be considered as a reasonable lower bound that can be expected to improve with practice.

The RSVP-iconCHAT design has important implications for mobile devices that have small screens and a limited number of buttons. Depending on the complexity of the desired message, the number of semantic roles can be chosen to match the available display space of a given mobile device. All search and prediction tasks can be delegated to the system, requiring only a single reliable selection mechanism for control. While the minimal control requirements are a single binary signal, as in the conducted usability study, control over the RSVP process can be expanded to include directional control of the display sequence or even the ability to modify RSVP speed.

Our prototype implementation of RSVP-iconCHAT accepts keyboard entry or mouse clicks, but the design can be configured to work with eye blinks, muscle twitches, brain waves, or any other input that can be discretized into binary form. This interface is potentially beneficial for users with profound impairments, such as those with locked-in syndrome who require electromyography (EMG) or brain-computer interface (BCI) solutions. Because many EMG and BCI systems provide a single output signal, and RSVP-iconCHAT requires only a single input signal, integrating such signaling methods is feasible and likely to be successful. Once the need for a voluntary motor response is removed, natural language processing and machine learning could be used to dynamically reorder the sequences of suggested semantic roles and associated icons, further increasing communication speed.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. 0914808. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Works Cited

Beukelman, D. and B. Ansel. “Research priorities in augmentative and alternative communication.” Augmentative and Alternative Communication 11.2 (Jan. 1995): 131–134. Web.

Beukelman, D. and P. Mirenda. Augmentative and Alternative Communication: Supporting Children and Adults With Complex Communication Needs. Paul H. Brookes Publishing Co., 2006. Print.

Fillmore, C. J. “Frame Semantics and the Nature of Language.” Annals of the New York Academy of Sciences 280. Origins and Evolution of Language and Speech (Oct. 1976): 20–32. Web.

Muller, K. R. and B. Blankertz. “Toward noninvasive brain-computer interfaces.” Signal Processing Magazine, IEEE 23.5 (2006). Web.

Orhan, U., et al. “RSVP keyboard: An EEG based typing interface.” Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on. IEEE, 2012. 645–648. Print.

Todman, J. “Rate and quality of conversations using a text-storage AAC system: Single-case training study.” Augmentative and Alternative Communication (Sept. 2000): 164–179. Web.

Todman, J., N. Alm, and L. Elder. “Computer-aided conversation: A prototype system for nonspeaking people with physical disabilities.” Applied Psycholinguistics 15.01 (1994): 45–73. Web.

Van Balkom, H. and M. Welle Donker-Gimbrere. “A psycholinguistic approach to graphic language use.” Augmentative and Alternative Communication: European Perspectives

(1996): 153–170. Print.

Wolpaw, J., et al. "Brain-computer interfaces for communication and control." Clinical Neurophysiology 113.6 (June 2002): 767–791. Web.