33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Message Formulation, Organization, and Navigation Schemes for Icon-Based Communication Aids

Rupal Patel

Abstract—**Individuals with severe speech and motor impairments rely on assistive communication devices to convey their needs and desires in social, educational, and vocational situations. Users with limited motor control or literacy often choose to use icon-based devices that afford increased speed of message formulation at the cost of fully generative language formulation on letter-based devices. A major challenge with large vocabulary icon-based systems is rate of communication. Message formulation, vocabulary organization, and navigation schemes can be used to mitigate the trade-off between vocabulary size and communication rate. This paper summarizes our research efforts to leverage semantic frame theory, situational context, and rapid serial visual presentation to improve message formulation speed and completeness in our iconCHAT and RSVP iconCHAT systems. Usability data and persisting challenges are discussed.**

I. Introduction

SPOKEN communication is not always a viable modality for individuals with severe speech motor impairment. Such profound communication impairments may result from congenital conditions, such as cerebral palsy, or may be acquired through traumatic injury, neurological disorders, or stroke. These individuals typically rely on augmentative and alternative communication (AAC) to express their needs and desires. AAC techniques range from simple signs (auditory, visual or gestural) that convey intentions/desires to complex messages constructed by concatenating a series of graphical symbols (letters and/or icons) that are then converted into spoken utterances using text-to-speech synthesis (see [35] for a taxonomy of AAC systems).

For speech-generation devices, there are three primary methods of message formulation: (1) by selecting letters, (2) by selecting words or phrases, and (3) by selecting icons/symbols. Letter-based AAC systems are the most generative, allowing users to spell out any word, but are also slow (approximately 2-5 words per minute) [4][20][38]. In comparison, word-based and icon-based systems are constrained by the vocabulary options available, yet message formulation speed can be faster (up to 15 words per minute) [13][28][29].

Users with limited literacy and/or cognitive, physical, or visual impairment often prefer, and may require, icon-based systems. For each user, vocabulary selection is an individualized process typically coordinated by a speech and language clinician in consultation with users, family members, and other regular conversation partners. Current commercially available icon-based AAC systems run on laptop-sized computers with touch-screens and can be mounted on wheelchairs or easily carried by ambulatory users. For devices with this type of form factor, accommodating larger vocabularies requires the size of each icon to be reduced, making visual search and physical access more difficult, or the screen size to be increased, thereby limiting mobility. To balance vocabulary size, icon size, and screen real estate, icons are typically organized in nested hierarchies and arranged in a prototypical linguistic structure (e.g. subjects are grouped on the left, verbs in the middle, and objects on the right to facilitate the subject, verb, object ordering of English) [16]. Such formulation and organizational schemes impose constraints on requisite user skills and ease of learning. Navigating multiple layers of icon arrays also requires more complex motor control as well as increased cognitive and memory demands, which may surpass the abilities of the target user population. Thus, an important design challenge with icon-based systems is to optimize message formulation, organization, and navigation for the end user's complex needs and abilities while maintaining sufficiently large vocabularies that enable expressive communication.

II. Approach to Design of Icon-Based Aids

A. Syntactic vs. Semantic Frame-Based Message Formulation

Virtually all icon-based AAC devices use a similar strategy of message formulation, which is based on the syntactical ordering of English (see for example Dynavox systems, Prentke-Romich Company, etc; exceptions include systems that use multi-meaning icons [1][2][5]). For example, to generate the utterance "I want chocolate ice cream", the user must select symbols corresponding to "I", "want", "chocolate," and "ice cream" in this precise linear sequence; however, many AAC users have difficulty with this type of message construction. Their utterances are often limited to simple sequences of two or three words and many users employ unusual syntax [31][32]. For example, AAC users may select the icons for girl + house + go (subject, object, verb) or house + go + girl (object, verb, subject) when trying to formulate the message "The girl is going home" (i.e., girl + go + home). These inventive constructions may reflect differences in linguistic

Manuscript received April 15, 2011. This work was supported in part by the National Science Foundation under Grants 0083032 and 0914808.

Rupal Patel is an Associate Professor in the College of Computer and Information Science and the Bouvé College of Health Sciences at Northeastern University, Boston, MA, 02115, USA (phone: 617-373-5842; fax: 617-373-2239; email: r.patel@neu.edu).

competence or compensatory strategies used to accelerate communication in light of poor interface design.

Semantic frame-based message formulation may offer an alternative to linear ordering of icons. The use of semantic frames as the central element of sentence formulation in our iconCHAT system [10][21] is inspired by the ideas underlying case grammars [11]. Case grammars focus on the functional relations between the predicate (verb) of a sentence and other sentence elements. Therefore, in the previous example "I want chocolate ice cream" the main predicate is "want," which takes an agent ("I") and an object ("ice cream"). These objects can be associated with modifiers, including "chocolate" for the object "ice cream." Empirical evidence supports this notion of predicate-central focus during sentence planning and execution [14][15]. This concept of case grammars is leveraged in iconCHAT allowing users to construct messages by first selecting the predicate, which defines a semantic schema indicating fillable slots for the agent, object, and various other predicate-dependent message components (**Fig. 1**). Note that the predicate forms the central graphical component of the visual schema (top left, **Fig. 1**). All other roles are arranged spatially to convey their relationship to the predicate and to each other, and the interface has been designed for flexibility in the ordering of icon selection. Semantic roles can be filled by navigating through the lexical categories (bottom middle, **Fig. 1**) and selecting lexical items (bottom right, **Fig. 1**). The selected icons are displayed in syntactical order (top right, **Fig. 1**) and are used to generate the spoken message using a Text-to-Speech (TTS) synthesizer. An important advantage of the iconCHAT system is that the semantic schema is not language dependent. Thus our message formulation approach could be used seamlessly across users of different language communities. Language-dependent aspects are incorporated into the case grammar and only realized in the generated output.

The use of an order-free semantic schema in iconCHAT aims to reduce linguistic and cognitive demands for utterance formulation imposed by syntactically ordered message construction systems. The goal was to move away from linear ordering and toward meaningfully structured visual images. Semantic frames in this context provide the scaffolding for users to compose complete sentences [11] [17][18][19][33]. We believe this kind of representation may be more accessible to non-literate and pre-literate communicators, but can also effectively serve linguistically skilled users. User studies with typically developing children have demonstrated that frame-based message formulation speed is equivalent to that using syntactic ordering systems that resemble written communication and that semantic formulation can be easily learned [23].

B. Hierarchical Category vs. Context-Sensitive Vocabulary Organization

Users of letter-based systems can generate an arbitrarily large variety of messages using only a few characters. These characters are always accessible on letter-based interfaces [38]. In contrast, messages formulated using icon-based systems are limited by the vocabulary size and available screen real estate needed to display each vocabulary item. Several hundred icons are necessary to formulate messages generative enough to cover topics in face-face conversation. This means that the full icon set cannot be viewed at once

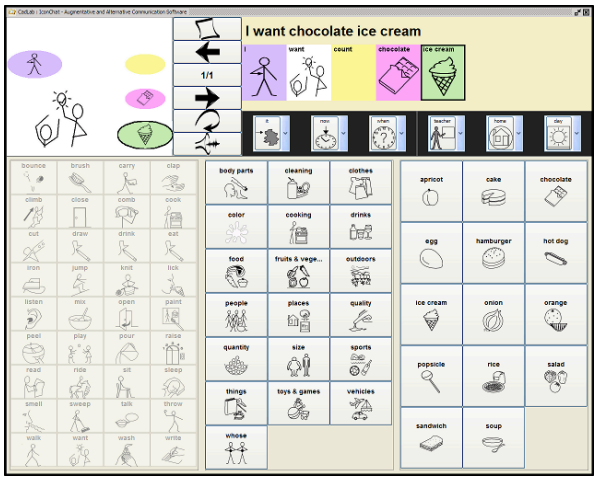

Fig. 1. Screenshot of iconCHAT interface illustrating the spatially organized semantic schema (upper left corner), predicate panel (bottom left), lexical categories (bottom middle), lexical items (bottom right).

requiring categorization and organizational schemes. To date, most commercially available devices place the burden on the user or his/her clinician/caregiver to categorize and organize icon layouts. In many cases, the categories chosen may not reflect the user's own conceptual organization leading to increased search time and frustration.

Many AAC devices employ methods to enhance access to frequently used vocabulary through dynamic layering or thematic organization of vocabulary (i.e. school, bedtime, mealtime "pages") [9][37]. Thematic "pages" put the onus on the user to locate the appropriate theme and then formulate messages appropriate to that theme. While "paging" permits access to large numbers of vocabulary items, the flexibility of out-of-context conversations is compromised. Another strategy is to impose a hierarchical categorization scheme in which icons are nested in categories, which are themselves nested in subcategories, etc. The cognitive demands needed to operate such a device quickly escalate, as does the difficulty in accessing icons in a timely manner. With sufficiently large vocabularies, a key challenge is locating the desired item in the nested hierarchy. Moreover, AAC users also require concepts that are infrequent or context-specific. One study reported a core vocabulary of 500 words based on conversational samples across five non-speaking adults [3]. Across the five participants, only 33% of any given user's communicative utterances could be generated using only those 500 core vocabulary items.

One potentially effective strategy for improving vocabulary access for fluent AAC users may be to facilitate

access to situationally dependent words. Constraining the word prediction to offer situationally appropriate vocabulary items may decrease the cognitive dissonance associated with searching lists of unrelated and unnecessary words [26][34]. Because situational vocabulary can potentially be enormous, it requires new ways to thinking about how to make it available to AAC users such that their interactions can be timely and relevant to the topic and setting of conversation. As a first step toward this goal, we demonstrated that vocabulary usage patterns tend to cluster by geographic location. We further showed that location cues can be used to automatically filter the available vocabulary choices on an icon-based interface to enhance communication rate [20]. Because contextual cues were only used to alter the icons displayed in the lexical item panel, users could choose to formulate situationally specific messages or out-of-context messages. In addition to location cues, we have explored the use of time of day, message history, and conversational topic (using topic spotting based on the conversational partner's speech) as independent cues and used in conjunction with location.

C. *Array-Based vs. Single-Key Navigation*

Although the iconCHAT vocabulary is organized using shallow hierarchies and grouped by semantic roles, navigating the interface requires free movement of a finger, mouse, or joystick, which alienates users with severe motor impairments such as cerebral palsy, amyotrophic lateral sclerosis (ALS), multiple sclerosis, or locked-in syndrome (LIS). The iconCHAT schema, however, provides a promising interface for designing a small-footprint, one-key message formulation system.

For users with severe motor impairment who have only one-key control, a viable navigation scheme must allow for: (1) browsing multiple icons, and (2) selecting a desired icon. One approach that has been used to simultaneously address both browsing and selection issues in navigation is rapid serial visual presentation (RSVP). RSVP is commonly used in experimental psychology [6][23] to display words or image content sequentially [12]. RSVP has also been shown to be an effective method for searching and browsing information on small screens, such as those on cellular phones and PDAs [8][25], making it ideal for use on small, portable AAC screens. Additionally, RSVP has shown promise for image retrieval [7], video selection [30], and letter-based AAC systems [27].

In our new system, called RSVP iconCHAT (**Fig. 2**), the hierarchical arrays of icons were removed to conserve screen real estate and enable access with one "switch." Vocabulary is organized according to relevance within semantic roles and displayed serially. Available input methods include any single-action input signal, such as a button, a blink, a muscle movement, or brain-wave activity.

To construct a message using RSVP iconCHAT, users first select the predicate from a series of predicates that are displayed in rapid succession. The selected predicate instantiates a spatially organized semantic frame much like in iconCHAT. Each semantic role within the frame is then highlighted sequentially to select the next role to be filled. In this way, users can select and populate semantic roles in any order. Once a semantic role has been selected, icons that can fulfill that role are serially displayed. At any point during message construction, users can select the "command field" that provides access to control functions, including "Speak" and "Clear." Selecting the "Speak" command sends the message in standard grammatical form to the integrated TTS system.

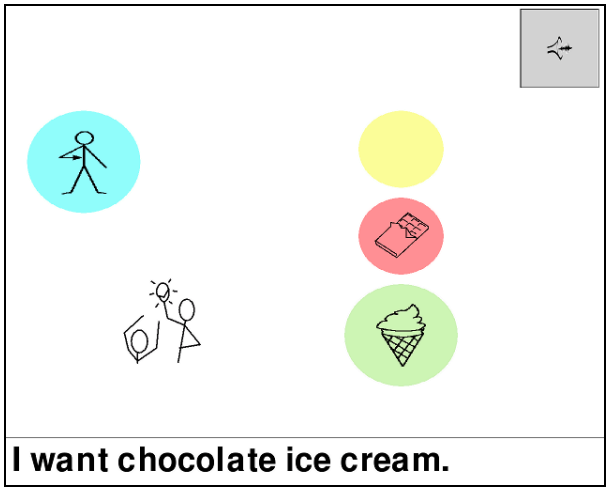

Fig. 2. Screenshot of the RSVP iconCHAT interface during formulation of the sentence "I want chocolate ice cream."

A usage study comparing the iconCHAT and RSVP iconCHAT systems with able-bodied users revealed that message formulation via RSVP was only twice as slow as conventional navigation methods, but equally expressive [36]. Errors in RSVP iconCHAT were largely due to slow motor reaction speeds, while errors in iconCHAT could be attributed to memory and cognitive load issues given that errors gradually diminished as users learned the vocabulary organization structure.

III. Conclusion and Future Directions

This paper summarizes some of the challenges in designing icon-based communication interfaces and describes several efforts aimed at enhancing communication rate through the use of novel message formulation, vocabulary organization, and navigation schemes. Much work still remains in validating the proposed prototypes for individuals with speech and motor impairment in terms of communication efficiency, ease, and satisfaction.

Some outstanding challenges include ways to: (1) design icon-based interfaces that grow with users to avoid requiring them to repeatedly learn new systems; (2) identify information-bearing and reliable contextual cues for real-time vocabulary filtering and reorganization; and (3) display large vocabularies on small-footprint, perhaps even mobile, platforms. While there are numerous competing demands, icon-based communication interfaces have the potential to enhance communication not only for those with severe

speech and motor limitations, but also for able-bodied users who could use them for language translation or language learning purposes.

ACKNOWLEDGMENT

We are grateful to individual users who have participated in usability studies and provided feedback at various stages in the interface design process. This material is based upon work supported by the National Science Foundation under Grants 0083032 and 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.

REFERENCES

[1] Baker, B. (1982). Minspeak: A semantic compaction system that makes self-expression easier for communicatively disabled individuals. *Byte*, *7*,186–202.

[2] Baker, B. (1986). Using images to generate speech. *Byte*, *11*,160–168.

[3] Beukelman, D., Yorkston, K., Pobleto, M., and Naranjo, C. (1984). Frequency of word occurrence in communication samples produced by adult communication aid users. *Journal of Speech and Hearing Disorders*, *49*, 360-367.

[4] Beukelman, D. and Mirenda, P. (1992). *Augmentative and Alternative Communication: Management of Severe Communication Disorders in Children and Adults*. Baltimore, Maryland: Paul H. Brookes Publishing.

[5] Bliss, C. K. (1965). *Semantography.* Sidney, Australia: Semantography Press.

[6] Chun, M. and Potter, M.C. (1995). A two-stage model for multiple target detection in rapid serial visual presentation. Journal of Experimental Psychology: Human Perception and Performance, 21, 109– 127.

[7] Combs, T.T. and Bederson, B.B. (1999). Does zooming improve image browsing. *Proc. JCDL1999, ACM Press*, 130–137.

[8] De Bruijn, O. and Spence, R. (2000). Rapid serial visual presentation: a space-time trade-off in information presentation. *Proc. AVI 2000, ACM Press*, 189- 192.

[9] Demasco, P. (1994). Human factors considerations in the design of language interfaces in AAC. *Assistive Technology, 6*(1), 10-25.

[10] Dominowska, E., Roy, D., and Patel, R. (2002). An Adaptive Context-Sensitive Communication Aid. In *Proceedings of the 17th Annual International Conference Technology and Persons with Disabilities, Northridge, CA*.

[11] Fillmore, C. (1968). The Case for Case. In E. Bach & R. Harms (Eds.), *Universals in Linguistic Theory*. New York: Holt, Rinehart, and Winston.

[12] Forster, K.L. (1970). Visual perception of rapidly presented word sequences of varying complexity. Perception & Psychophysics, 215–221.

[13] Foulds, R. (1980). Communication rates for non-speech expression as a function of manual tasks and linguistic constraints. *Proc. RESNA*, 83-87.

[14] Griffin, Z. M. (1998). *What the eye says about sentence planning.* Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign, IL.

[15] Griffin, Z. M. and Bock, K. (2000). What the eyes say about speaking. *Psychological Science, 11*(4), p. 274-279.

[16] Goodenough-Trepagnier, C. (1994). Design goals for augmentative communication. *Assistive Technology*, *6*(1), 3–9.

[17] Kingsbury, P., Palmer, M., and Marcus, M. (2002). Adding Semantic Annotation to the Penn TreeBank. In *Proceedings of the Human Language Technology Conference, San Diego, CA*.

[18] Levin, B. (1977). *Mapping Sentences to Case Frames* (Working Paper 143). Cambridge: Massachusetts Institute of Technology, Artificial Intelligence Laboratory.

[19] Levin, B. (1993). *English Verb Classes and Alternations: A Preliminary Investigation.* Chicago, IL: University of Chicago Press.

[20] Muller, K.R. and Blankertz, B. (2006). Toward Noninvasive Brain-Computer Interfaces. *IEEE Signal Processing Magazine, 23*, 126-128.

[21] Patel, R., Pilato, S., and Roy, D. (2004). Beyond Linear Syntax: An Image-Oriented Communication Aid. *Journal of Assistive Technology Outcomes and Benefits*, 1, 57-66.

[22] Patel, R. and Radhakrishnan, R. (2007). Enhancing Access to Situational Vocabulary by Leveraging Geographic Context. *Assistive Technology Outcomes and Benefits,* 4, 99-114.

[23] Patel, R., Schooley, K., Radhakrishnan, R. (2006). Comparison of semantic versus syntactic message formulation: A pilot study. *Assistive Technology Outcomes and Benefits*, 3, 96-110.

[24] Potter, M.C., and Levy, E.I. (1969). Recognition memory for a rapid sequence of pictures. *Journal of Experimental Psychology, 81*, 10–15.

[25] Rahman, T. and Muter, P. (1999). Designing an interface to optimize reading with small display windows. In Human Factors, Human Factors and Ergonomics Society.

[26] Renaud, A. I. (2002) *Diagnostic evaluation measures for improving performance of word prediction systems*. Unpublished master's thesis, University of Waterloo, Canada.

[27] Roark, B., de Villiers, J., Gibbons, C., and Fried- Oken, M. (2010). Scanning methods and language modeling for binary switch typing. *Proc. SLPAT*, 28-36.

[28] Szeto, A., Allen, E., Littrell, M. Comparison of Speed and Accuracy for Selected Electronic Communication Devices and Input Methods. (1993). *Augmentative and Alternative Communication, 9*, 229-242.

[29] Todman, J. (2000). Rate and quality of conversations using a text-storage AAC system: Single-case training study. *Augmentative and Alternative Communication*, *16*, 164-179.

[30] Tse, T., Marchionini, G., Ding, G., Slaughter, L., and Komlodi, A. (1998). Dynamic key frame presentation techniques for augmented video browsing. Proc. AVI 1998, ACM Press, 185–194.

[31] Udwin, O. and Yule, W. (1990). Augmentative communication systems taught to cerebral palsied children: A longitudinal study: I. The acquisition of signs and symbols, and syntactic aspects of their use over time. *British Journal of Disorders of Communication, 25*, 295-309.

[32] Van Balkom, H. and Welle Donker-Gimbrere, M. (1996). A psycholinguistic approach to graphic language use. In S. von Tetzchner & M. Jensen (Eds.), *Augmentative and alternative communication: European Perspectives* (pp. 153-170). London: Whurr Publishers.

[33] Van Valin, R. D. Jr. (2004). *Lexical representation, co-composition, and linking syntax and semantics.* Retrieved February 24, 2004, from <http://wings.buffalo.edu/linguistics/rrg/>.

[34] Venkatagiri, H. S. (1994). Effect of window size on rate of communication in a lexical prediction AAC system, *Augmentative and Alternative Communication, 10*, 105-112.

[35] Venkatagiri, H. S. (2002). Clinical Implications of an Augmentative and Alternative Communication Taxonomy. *Augmentative and Alternative Communication*, 18, 45-57.

[36] Wiegand, K., Patel, R., and Erdogmus, D. (2010). Leveraging Semantic Frames and Serial Icon Presentation for Message Construction. *ISAAC Conference for Augmentative and Alternative Communication,* Barcelona, Spain, July 2010.

[37] Wilson, M. S. and Fox, B. J. (1993). Computer technology: Emerging technology meets universal grammar. *American Speech and Hearing Association, 35*(9), 37-38.

[38] Wolpaw, J.R., Birbaumer, N., McFarland, D.J., Pfurtscheller, G., and Vaughan, T.M. (2002). Brain Computer Interfaces for Communication and Control. *Clinical Neurophysiology 113*(6), 767-791.