Proceedings of RESNA 1993 Annual Conference, Las Vegas, NV.

Methodological Decisions in the Study of Augmentative Communication Systems

Heidi Horstmann Koester and Simon P. Levine
Rehabilitation Engineering Program, Department of Physical Medicine and Rehabilitation
University of Michigan

Abstract
The design of research in augmentative communication (AAC) requires the researcher to make many difficult decisions and to understand the consequences of those decisions as fully as possible. This paper discusses issues in AAC research design, using our particular research goals and resulting methodological approach as a framework. The intent is not to recommend one approach at the exclusion of others, but to encourage the communication of research design decisions and to stimulate further discussion on the trade-offs involved.

Introduction
There are many reasons why it is desirable to investigate the performance achieved during use of augmentative communication and computer access systems. Often, a primary goal is to provide a means of comparison between systems, or between aspects of the same system (e.g., 1,5,6). While there remains a strong need for more data on how different users perform with different AAC systems, purely empirical studies can be limited in their generality. One way to address these limitations is through the development of user performance models that integrate system and user factors and support the simulation of a wide range of user-system combinations (3,9). Several theoretical studies have been performed, but the validity of these will remain controversial until the accuracy of the predictions has been tested empirically (4,7). Theoretical model development depends on empirical data to both test and refine model accuracy.

Important questions exist regarding how to best design research that enables valid empirical comparisons and the development of performance models. This paper discusses some of the issues in AAC research design, based on experience in our laboratory. It is not intended to be a detailed account of specific methods, but rather to encourage dialogue on the trade-offs involved in research design decisions. We hope this will be valuable to researchers and non-researchers alike, as it will provide some insight into the research design process in general as well as the compromises and benefits involved in our particular approach.

Background
To provide a foundation for the discussion, the goals of our research are outlined briefly. The overall aim is to progress toward general principles underlying the way people use AAC systems and the way in which the system design affects behavior and performance. In pursuit of this goal, our current focus is on the validation of user performance models for particular AAC systems and user groups. Principles of human-computer interaction provide the basis for developing these quantitative models. User performance can then be simulated with the model, using parameter values that represent the system and the user. The resulting predictions are compared to actual performance observed over a range of user characteristics and system configurations.

The data collected in the pursuit of the model development goals will also support the evaluation of the specific interfaces under study. This evaluation includes statistical comparisons of performance across subjects, across strategies of use, and across time. While these comparisons fill an important gap in current knowledge, the model validation goals remain our priority because of their potential applicability to a much broader range of AAC system-user combinations.

Approach
The decision areas discussed below are: the type of system to be studied, subject characteristics, system input method, subjects' strategy of use, experimental protocol, and data collection. The main consideration for each decision is how well the different options support the research goals, followed by practical considerations (e.g., research budget). We outline several options in each area, our reasons for making a particular choice, and the compromises involved in that choice. Interdependencies between decisions are discussed where most relevant.

<u>System</u>. Modeling techniques can be applied to many different types of AAC systems. In our research, we have chosen to focus on word prediction, a general technique intended to enhance rate by reducing keystrokes (11). Word prediction is widely used clinically, and there are several well-implemented commercial systems available. However, word prediction may not always provide a significant enhancement in rate, which has been attributed to the trade-off between reduced motor actions and increased cognitive and perceptual activities (3,10). Application of modeling to word prediction provides an ideal opportunity to understand the net effect of this trade-off more rigorously.

Rather than use a commercially available word prediction system, we have chosen to develop our own system for research purposes, to gain sufficient control over the system configuration as well as the means of data collection. A main consideration is that the research system be highly representative of commercial word prediction systems. Almost all systems use the same basic method of finding words in an internal dictionary that match the initial letters entered by the user, then presenting the set of matching words that are most frequently used in English. Some systems augment this basic algorithm by one or more of the following techniques: using syntactic knowledge to remove inappropriate predictions; using recency-of-use as well as frequency-of-use to choose predictions; updating the frequency of words based on user entries; and using information about the previous word to refine predictions.

These variations on the basic word prediction algorithm share two main effects. First, they improve the success of the system's predictions. This effect is simulated in our research system by changes in the text to be entered or the dictionary used. A second effect is that there are many possible prediction lists for a given word-initial sequence. While the software is capable of simulating these changing word lists, the constant novelty limits the user's ability to learn when words will appear in the lists, which at least partially counteracts the increase in predictive success provided by more sophisticated algorithms. For this reason, the basic algorithm is often used alone in actual systems, and it has been chosen for study in our current research.

The main implication of this choice is that the results will be most directly applicable to performance with the basic algorithm. However, models that will result from this work should be general enough to simulate aspects of performance with other word prediction algorithms, by changing model parameter values (e.g., the percent of words selected from the list). And while the results will not be directly applicable to systems that do not use word prediction, any success in applying the models to word prediction should provide confidence for future applications to other systems.

Subjects. While the focus of this work is to develop models for the performance of users with disabilities, our current research employs both able-bodied and disabled subjects. The use of able-bodied subjects has three purposes: First, it provides a feasible way to gain sufficient statistical power, since able-bodied subjects are more readily available. Second, the performance measured from able-bodied subjects allows a "best-case" test of model accuracy, which will help determine if inaccuracies in modeling disabled users are due to flaws in the model itself or to increased variance in disabled user performance. Third, using both types of subjects provides a means of assessing the extent to which able-bodied performance is similar to that of users with disabilities. Our hypothesis is that the results from able-bodied subjects will prove to be generalizable at least to a large segment of the target population, specifically those individuals who have no cognitive impairments and whose physical disability does not prevent them from using the tested input method with consistency.

An important implication of using able-bodied subjects is that they generally have no prior experience with word prediction. For consistency, then, this restricts our choice of disabled subjects to those who have no experience with word prediction. This decision to employ novice subjects has affected several other decision areas, as discussed below. An alternative is to avoid the use of able-bodied subjects and employ only disabled subjects who are experienced word prediction users, following a common method in studies of human-computer interaction (e.g., 8). The resources necessary to pursue that approach are currently unavailable, which adds a practical reason to the theoretical grounds for studying novice subjects.

System Input Method. Although a variety of input methods could be modeled, early theoretical studies focused on single switch scanning (2), so our first pilot study used scanning as the input method. This initial choice has been revised in our current work. Direct selection using a mouthstick or other typing aid is now the basic input method for several reasons. First, it is frequently used by actual users of word prediction. Second, user performance with direct selection is driven primarily by user ability, since the user is free to make a selection at any time. This is in contrast to single-switch scanning, in which user skill is confounded with the system timing parameters as an influence on user strategy and performance, since these parameters determine when a particular selection is possible. Third, existing data on the performance of able-bodied subjects using a standard keyboard with a mouthstick (5) is a source of valuable pilot information. Finally, direct selection is easier to learn than a more complex system such as Morse code, which is important due to the use of novice subjects.

Strategy of Use. Strategy refers to the plan of action that guides the user's behavior while interacting with the system. In the case of word prediction, strategy involves the way in which the user employs the word list. For example, at one extreme is a strategy in which the list is searched carefully before each and every selection. Alternative strategies include postponing list search until after one or two letters have been chosen, or deciding when to search the list based on the perceived likelihood of the word's presence.

While there are important questions regarding the strategies that users develop in the absence of specific instruction, several considerations have led to our decision to teach subjects particular strategies for using the word prediction system. The main reason is our commitment to performance modeling goals. Since the model equations are based on the component actions that the user executes during use of the system, model accuracy is improved by the ability to know or predict user behavior. Because subjects are word prediction novices, strategy instruction provides an important means of reducing variation in their behavior. Without some sort of instruction, each subject would employ a different strategy or mix of strategies, requiring the development of a different model structure for each subject. By teaching subjects a particular strategy, and enforcing its use, performance models can be built prior to the collection of data, and their predictions compared with the performance of all subjects who used that strategy. As confidence in modeling techniques is established, models can be created to match any particular strategy.

The primary reason for strategy instruction stems from the modeling goals, but the approach has interest and validity from a clinical perspective as well. First, teaching explicit strategies points to the potential clinical application of the modeling tools as a means of determining optimal, or at least good, strategies for using a system. Second, it is at least as realistic as letting subjects evolve their own strategy, since actual users frequently receive at least minimal and occasionally extensive training in a particular strategy.

Experimental Protocol. Because all subjects are word prediction novices and some subjects are mouthstick novices, the protocol must allow sufficient time for subjects to develop skill. Since true asymptotic expertise would likely take a large number of sessions, the goal is for subjects to become "skilled novices" and for their learning curves to begin to level off by the last session. This approach provides data on skilled performance, as well as data on unskilled novice use and principles of learning, which can be analyzed in future studies. It cannot, however, directly address issues of true expert performance.

A text transcription task has been chosen for this research, in which subjects transcribe unique blocks of text in each session. This approach does not allow us to address important questions regarding the quality of text that an actual word prediction user might compose. However, it has the advantages of ensuring that subjects' performance can be validly compared to each other and that the same model simulations (which strongly depend on the text characteristics) can be used for all subjects who use a common strategy.

A great deal of care has been taken in the creation of transcription text samples, since text characteristics can have a significant effect on performance with a word prediction system. In particular, sessions that are intended to have the same system configuration must use texts that match with respect to the following characteristics: average word length, percent of words that can be selected from the word list, average number of letters generated per word list selection, and percent of keystrokes saved. Creation of these matched texts was facilitated by the development of software that simulates the entry of a text sample using word prediction with a given configuration and strategy, then calculates the resulting text characteristics.

Data Collection. The emphasis on performance modeling in our research places the focus of data collection on performance time, i.e., text entry rate, as measured by seconds per character. Each keystroke is recorded with its associated time of entry and encoded to reveal what parts of the text were selected letter-by-letter and which resulted from a word list selection. The number of errors committed in a session is recorded as well, primarily as a means of ensuring that all subjects are at a similar point on the speed-accuracy trade-off. Sessions are also videotaped to provide a visual record of subject behavior. This allows verification of adherence to the assigned strategy and provides a means for revising the behavioral basis of the performance models if necessary.

The focus on quantitative data is consistent with the decisions to constrain qualitative variables such as subject behavior and the text to be entered, as discussed above. This does not mean, however, that qualitative data must be ignored completely. Subject motivation and affect are observed during sessions, and subject comments are solicited after each session regarding their impressions as well as their rating of task difficulty.

Discussion

The design decisions outlined above reflect a consistent focus on model development and quantitative performance in augmentative communication. The goal is to develop a modeling framework that provides an understanding of the factors that determine text entry speed and predicts the speed that may be accomplished under a range of conditions. Commitment to this goal provides the rationale for our research design decisions, but as in any study, it also restricts the range of questions that can be validly addressed. Research on other important issues such as word prediction's effect on user spelling, motivation, or fatigue would require a different set of methodological decisions. Indeed, gaining multiple perspectives through diverse approaches is the key to advancing knowledge in the field. Our point, therefore, is not to recommend one particular approach at the exclusion of others, but to encourage the clear communication of research design decisions and the open discussion of their consequences.

References

1. Gardner-Bonneau, D.J. and Schwartz, P.J. (1989). A comparison of Words Strategy and traditional orthography. *Proc. of 12th Annual RESNA Conf.*, Washington, D.C.: RESNA, 286-287.
2. Horstmann, H.M. and Levine, S.P. (1990). Modeling of user performance with computer access and augmentative communication systems for handicapped people. *AAC, 6:4*, 231-241.
3. Horstmann, H.M. and Levine, S.P. (1991). The effectiveness of word prediction. *Proc. of 14th Annual RESNA Conf.*, Washington, D.C.: RESNA, 100-102.
4. Horstmann, H.M. and Levine, S.P. (1992). Modeling AAC user performance: response to Newell, Arnott, and Waller. *AAC, 8*, 92-97.
5. Levine, S.P., Gauger, J.R.D., Bowers, L.D., and Khan, K.J. (1986). A comparison of mouthstick and Morse code text inputs. *AAC, 2:2*, 51-55.
6. Light, J., Lindsay, P., Siegel, L., Parnes, P. (1990). The effects of message encoding techniques on recall by literate adults using AAC systems. *AAC, 6*, 184-201.
7. Newell, A.F., Arnott, J.L., and Waller, A. (1992). On the validity of user-modeling in AAC: Comments on Horstmann and Levine (1990). *AAC, 8*, 89-92.
8. Olson, J. and Nilsen, E. (1988). Analysis of the cognition involved in spreadsheet software interaction. *Human Computer Interaction, 3:4*, 309-350.
9. Rosen, M. (1989). Editorial. *Assist. Technol. 1*, 79-80.
10. Soede, M. and Foulds, R.A. (1986). Dilemma of prediction in communication aids. *Proc. of 9th Annual RESNA Conf.*, Washington, D.C.: RESNA, 357-359.
11. Swiffin, A.L., Arnott, J.L., Pickering, J.A., and Newell, A.F. (1987). Adaptive and predictive techniques in a communication prosthesis. *AAC, 3:4*, 181-191.

Address

Heidi H. Koester
Rehabilitation Engineering Program
1C335 University of Michigan Hospital
Ann Arbor, MI 48109-0032, USA