The Development of Behavior-Based User Models for a Computer System

By Robert Bushey[1], Jennifer Mitchell Mauney[2], and Thomas Deelman[1]

[1] SBC Technology Resources, Inc., Austin, TX, USA
[2] Mauney Consulting, Austin, TX, USA

Abstract. This paper examines the development of user models for the graphical user interface of a telecommunication computer system used during service and sales negotiations. User models enhance the requirements gathering phase of system design by capturing the diversity of the user population and capitalizing on the variety of distinguishable and categorizable strategies that affect performance. The CDM method (Categorizing, Describing, and Modeling method) was developed as a technique to generate user models. In the CDM method, the user population is first categorized into a reasonable number of groups. The behaviors for each group are described and then qualitatively and quantitatively modeled. These models are subsequently used during the system design and operational processes to optimize performance of the entire user population.

1 Introduction

In gathering requirements for the design phase of system development, it is common practice to interview and observe the behaviors of a random sample of users from the intended user population. System design requirements typically characterize the user as one entity with a single set of behaviors, namely expert, novice, or a composite of all the users. For example, if the system development team is emphasizing the high-end performance, the behaviors and characteristics that emerge are items related to the expert user. Therefore, the designers tune their comments and suggestions towards the expert users. However, in some design projects, ease of learning, training, or novice aspects are emphasized to a greater extent. In this case, interviewers focus on the comments and suggestions by novices. If there is no overwhelming performance issue or training issue that directs the team, then anecdotal behavioral information is obtained on a variety of users. Typically, these approaches limit the overall performance when the system is implemented because not all of the users' behaviors are accounted for during the design phase.

The development of a system that accommodates the diversity of the user population and improves the users' performance is optimal. One method to improve the users' performance is to categorize the system users into groups, describe and model each group's behaviors, and then incorporate this information in the design and operational processes. Currently, there is no technique that will do all three phases of the above, rather tools and techniques exist that accomplish only small portions of the desired process.

Due to this lack of an encompassing technique that incorporates categorizing, describing, and modeling users and then applying this information to the system design and process operation, the CDM (Categorize, Describe, and Model) method was developed. The purpose of the CDM method is to build a set of precise and accurate models that represent the interaction of diverse user behaviors with the system. The results of these models are then applied during the system design and operational processes to optimize performance for all users.

Since the CDM method focuses on modeling different users' behaviors, it is best implemented on systems where users' behaviors are measurably different. One illustration of users' behaviors being measurably different is shown in Figure 1.

Figure 1 presents a wide range in variability on a metric used to measure users' performance. Note that there is more than a "500%" difference from the lowest to the highest value. In the current and previous projects, a broad range of user performance has been a key indicator that user behaviors are indeed different.

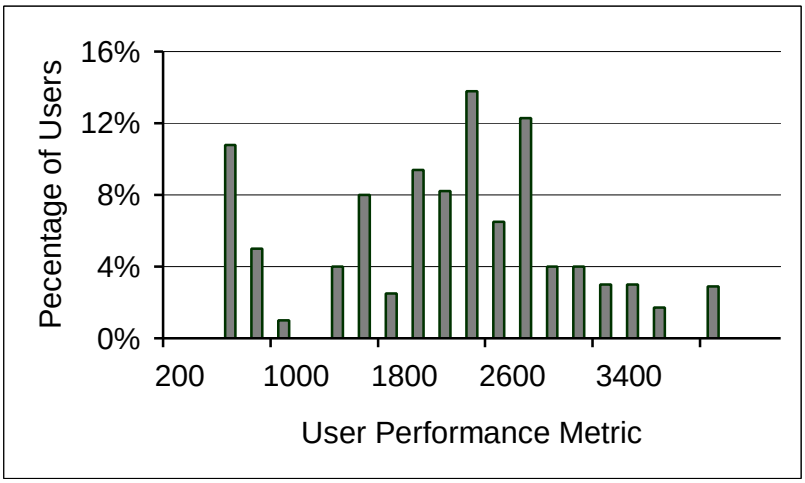

Figure 1. Histogram of user's performance.

When it has been determined that users' behaviors differ, the first step in the CDM method is to categorize the users. Both cognitive and performance measures are calculated for each user and the users are categorized into groups with similar results.

Next, a subset of users from each group are selected and their behaviors observed and documented. Behaviors are examined and should demonstrate similarities within a group and differences between groups of users. The emphasis of these descriptions focuses on behaviors that affect performance and are incorporated into the user models (Eberts, 1994; John and Kieras, 1996; and Kieras, 1996).

When the users' behaviors are well understood within a given user group, a user model is constructed. There are two types or levels of models: qualitative and quantitative. The qualitative models include statements of a user's behavior from the specific user group in certain situations or performing certain functions. The quantitative models also represent the behavior of a specific user group, but are more formal by incorporating the capability to make numerical performance predictions (Harrell and Tumay, 1995; Pegden, 1990).

2 Categorization of Users

As previously mentioned, users are categorized based upon similar behavioral characteristics that are important to system interface design and use. Users are categorized using mental workload measures, user characteristic measures, and performance measures.

Since it is expected that differences in mental workload relate to types of strategies (Shingledecker, 1980; Welford, 1978; Wei, 1997), measuring mental workload is an indirect way to measure types of strategies. Furthermore, measuring mental workload can be quick, easy to perform, and inexpensive (Eggemeier, 1988). Conversely, determining each user's strategy that is evidenced by their behaviors for each task would take a considerable amount of time, requires a lot of effort, and be very expensive.

The NASA Task Load Index (NASA-TLX) (Hart and Staveland, 1988) was chosen as a mental workload measure to facilitate the categorization of users. However, it was modified for this system. The original NASA-TLX contains six measures: mental demand, physical demand, temporal demand, performance, effort, and frustration. Physical demand was removed due to its absence in the users' duties. Effort was confused with mental demand in pre-study testing; therefore effort was removed from the TLX. Furthermore, the performance measure was removed. Users may have viewed the performance measure as a scale relating to their performance reviews. Lastly, the label "temporal demand" was changed to "time pressure". The modified TLX contained the three measures of mental demand, time pressure, and frustration.

Another method for facilitating the categorizing of users is to assess user characteristics. User characteristics refer to qualities or traits that are measurable and differ between users. Since the CDM method is based on modeling different behaviors and because user characteristics should, to some degree, reflect user behaviors, it was hypothesized that user characteristics may facilitate the categorization of users.

An example of a user characteristic is the users' ability and preferred method to navigate through the system. For the system examined, the ability to navigate is a user characteristic that was measurable, differed between users, could be modeled, and reflected users' behaviors of interacting with the system. This characteristic was also important to system design. For example, users who rated themselves as having difficulty navigating to specific areas in the system were expected to benefit from a menu-based interface. Menu-based interfaces require less mental demand (memorizing) than other types of interfaces. Users who rated themselves as having no difficulty navigating through the system and wanted to quickly navigate were

expected to benefit from an interface where menus could be skipped and shortcut keys could be used.

Using preliminary surveys on a small subset of users, the general characteristics regarding navigation through the system and recall of packages/services and prices were found to be important differentiators in the system design and thus, for establishing user groups. In addition, the user characteristic of cross-selling was also found to be an important differentiator for establishing user groups. Cross-selling is the act of selling additional products or services to a customer. In the system examined, it was found that some users almost always tried to cross-sell, others never cross-sold, and still others only cross-sold in certain situations.

In applying the CDM method to this system, the users were given a survey containing the modified NASA-TLX and the user characteristic questions. The users rated the three measures (mental demand, time pressure, and frustration) of the modified NASA-TLX on a twenty-point scale. The users then performed paired comparisons between these three measures. The paired comparisons were used to weight the measures and facilitate the development of a score representing the mental workload experienced.

The users also rated themselves on the user characteristic questions regarding navigation, recall, and cross-selling. The questions were rated on a nine-point scale. An example of a user characteristic question can be seen in Table 1.

Table 1. Example of a user characteristic recall question.

How easy was it to recall the various products and services while you were on the phone with a customer?

Very Easy								Almost Impossible

Since system users from more than one center were examined, the centers were examined for consistency of user responses. Table 2 presents the modified NASA-TLX data and the normalized, collapsed user characteristic data for 262 users. Bushey, Mauney, and Deelman (1998) present this data in more detail along with a more detailed account of the categorizing section of the CDM method.

The mental workload scores display a high degree of consistency between the different centers. In addition, the normalized recall, purchase, and navigation data show consistency between centers.

The normalized data from both the modified NASA-TLX and the user characteristic questions were weighted and incorporated into an equation called the cognitive metric. The values were weighted based on their importance to system operation. The cognitive metric, along with the performance measures, facilitated the categorization of users.

Table 2. Modified NASA-TLX scores and user characteristic scores for each center.

Centers	Mental Workload Score*	Normalized Navigation Questions **	Normalized Recall Questions **	Normalized Cross-Selling Questions **
Center A	64	0.7	0.7	1.6
Center B	61	0.8	0.9	1.3
Center C	65	0.7	1.0	1.3
Center D	68	0.7	0.8	1.5
Center E	66	0.7	0.8	1.6
Center F	55	0.8	0.8	1.4
Center G	63	0.6	0.8	1.6
Average	62.7	0.7	0.8	1.5

* Denotes a range of 1-100, 1-low, 100-high.
** Normalized to individual average and denotes a 9 point Likert scale, 1-very easy, 9-impossible.

There may be multiple performance measures for a telecommunication system. These measures may include, net sales, gross sales, cross-selling sales, sales per call type, retention of sales, customer satisfaction, number of calls per time period, etc. The performance measures used for the CDM method were the primary performance measures used by the systems operation management. Currently, the users are monetarily compensated based on their performance based on these specific measures. It was hypothesized that these performance measures would also facilitate the categorization of users since the users' behaviors should be reflected in these measures.

Performance measures were acquired on the same subset of users who performed the mental workload and user characteristic survey. The performance metric is an equation that incorporates the performance measures for the specific system. Each normalized performance measure in the equation is weighted based on its importance to system operation.

In order to categorize the users into specific groups, data from the cognitive metric and performance metric were plotted. Next, Ward's minimum variance method (Milligan, 1980) was used to cluster similar data points based on both the cognitive metric and performance metric. The users were initially categorized into four groups. Observational data was then collected in order to validate and refine the original groupings.

3 Description of Users

In the describing section of the CDM method, a subset of individuals in each categorized group was observed for behaviors used to perform their tasks. In this stage of the CDM methodology, an observer unobtrusively recorded a variety of activities including the manner in which users navigated through the current interface and details of the sales negotiation performed between the user and actual customers.

Two behaviors that made a predominant impact on users' job performance were the number of cross-selling attempts made to the customer and the length of the call. For example, users who did not make any cross-selling attempts and quickly performed the service requested by the customer, rapidly completed a large number of customer sales. This behavior typically resulted in a large number of low revenue calls. Other users talked longer to the customer to determine the most likely types of products or services that they could successfully cross-sell to the customer. This behavior typically resulted in a smaller number of high revenue calls.

It was also determined that cross-selling behavior was dependent on the task of the customer. For example, customers may have the task of "I need to get a product/service". Therefore, some tasks allow for more successful cross-selling attempts by a user.

Since users from more than one call center were examined, the call centers were examined for consistency on a variety of measures. One measure was the consistency of customer tasks. Overall, customer tasks could be divided into three basic types. Table 3 presents the data for the customer tasks by center.

Table 3. Percent of customer task types for each center.

Centers	**Type A Customer Tasks**	**Type B Customer Tasks**	**Type C Customer Tasks**
Center A	75%	17%	8%
Center D	74%	15%	11%
Center E	79%	19%	2%
Center F	73%	19%	8%
Center H	78%	15%	7%
Center I*	74%	9%	17%
Center J*	76%	12%	13%

* After hours centers.

Centers A through H are relatively consistent for customers' tasks. Although Center E has a lower percentage of Type C customer tasks, these C tasks were not used in any of the analyses. Centers I and J are after hours centers, meaning they received calls between the hours of 5:00pm and 8:00am. These centers tended to have a relatively smaller number of Type B customer tasks and a relatively larger number of Type C customer tasks. For this reason, and because there were some job differences at these two centers, the users from Centers F and G were not included in the rest of the analyses or results. Over 1000 customer tasks (calls) were used to calculate the percentages in the table above.

Based on the number of cross-selling attempts per call and average length of call, the users were grouped by similar behaviors. The results of this grouping technique are shown with a subset of users in Figure 2.

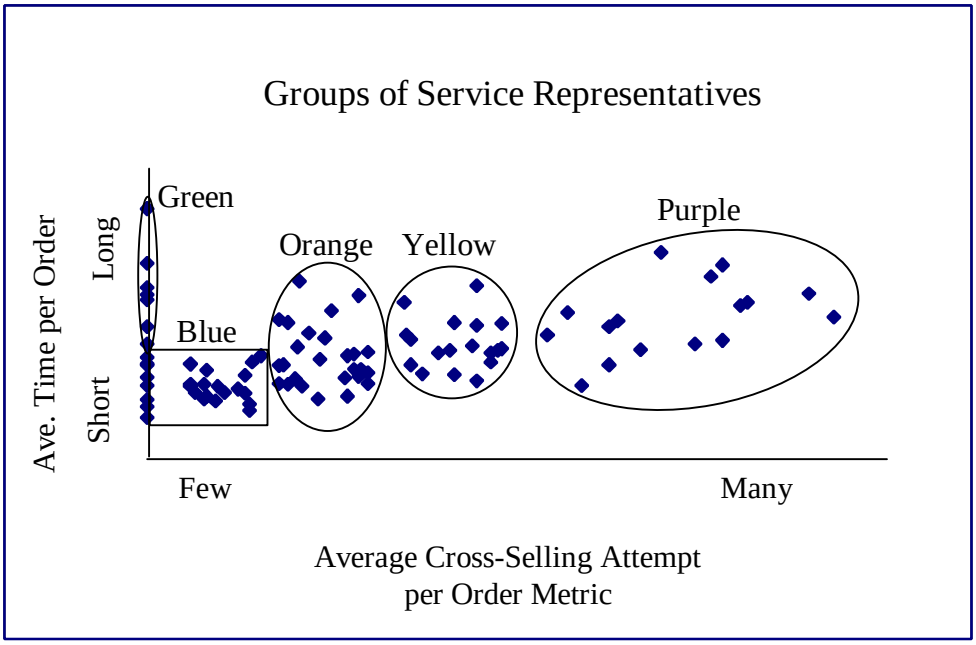

Figure 2. Groupings based on observational data.

Based on the observed, objective data, five groupings of users emerged. In the figure above, members of the Blue group are quick call takers who rarely or never make cross-selling attempts. In contrast, members of the Green group average spending over twice as much time on a call and never make sales attempts. Members of the Purple group also average spending more time on a call, but make several sales attempts.

As discussed previously, users were originally categorized into four groups. This divergence in the number of user groups is due to the detail in the data collection procedures. The users were originally categorized into four groups based on a short, subjective survey. The observed, behavioral data not only were collected using objective methods, but also contained much more detail. Due to the greater level of detail in the objective measures, the users were more distinctly categorized, resulting in five user groups.

The categorization of users based on survey and performance data was intended to be at a validity level such that the labor and cost intensive task of grouping users based on observable behaviors was not needed. For this system development, categorizing users based on the cognitive and performance data compared to categorizing users based on observable behaviors was not optimal. However, the quality of information gained from the cognitive and performance metrics greatly facilitated the system development and notably contributed to the describing and modeling sections of the CDM method. In addition, this information also

contributed to the design of the user interface. For these reasons, the initial categorizing step of the CDM process was highly beneficial and necessary.

The groupings that emerged through the survey and performance data collections and from the observed behaviors provide a rich combination of information that is currently being applied to interface design. The observational data is particularly useful when the types and order of interface screens that users visit are linked with the user's group membership. The current set of user behaviors and characteristics enable detailed descriptions of each group to be made. These descriptions take the form of qualitative and quantitative models. These models represent the diversity of the user population and show the particular needs required of each group that can be incorporated into interface design.

4 Modeling the Users

The qualitative models include statements of how users within a specific user group behave in certain situations or perform certain functions. These models allow the design team to represent each of the various user groups in the design process. As the various design decisions are addressed, these models are extremely valuable by ensuring that the various user group needs are not lost in the development process.

Since there are five groups of users, five qualitative models were developed. A section of the Blue Qualitative Model included the following. The Blue group tends to be efficient call takers who most likely do only what the customer requests. They will want a system that allows them to process calls very quickly. They tend not to cross-sell. To facilitate members of the Blue group to make cross-selling attempts, incorporating sales incentives would be appropriate. In addition, some members may need to be trained on successful cross-selling techniques.

Another a section from a qualitative model included the following. The Yellow group consists of average call takers who try to engage customers with a specific task into buying at least one additional product or service. They will optimally perform using a system that allows them to easily make cross-selling attempts without having to decrease the time they spend conversing with the customer since it is the conversation with the customer that facilitates the successful cross-selling attempt.

Quantitative models represent the behaviors of the user groups in more detail than those of the qualitative models. Specifically, the quantitative models incorporate the capability to make numerical performance predictions. Items such as arrival patterns, resources allocations, and task duration times are also included and combined with the flow of incoming work. These quantitative models are developed only to the degree of detail necessary to adequately represent the user groups for the design team during the system development process.

These quantitative models show how certain groups of service representatives could perform more in accordance with business goals. The quantitative models could also be used to make predictions about changes to the system and in operational processes. For example, if

operational management wanted to stress customer accessibility, these models could be used to predict how long customers would have to wait in the queue if no more than one cross-selling attempt were made per customer. It would also predict the decrease in sales if no more than one cross-selling attempt were made per customer.

5 Conclusion

Given that the users of this system differed, the CDM method facilitated the categorization of similar groups of users. It also facilitated the interface designers' understanding of the requirements of different types of users through applying qualitative and quantitative models. For example, qualitative models of the most distinctive user groups that positively contribute to operational efficiency were applied during the design stage of interface design. Specifically, each design implementation was evaluated to ensure that the needs of these user groups were accommodated. This way, the system is designed to accommodate the behavioral diversity of the user groups that most strongly contribute to meeting business goals. This method essentially allows the system to be customized to facilitate desired behavior and optimize preexisting behavior that has proven to be successful in the competitive sales and service environment.

The CDM method has only been applied to the sales negotiation telecommunication computer system described in this paper. Future work will focus on a variety of areas including the application of the CDM method to other systems including the documentation of the successes and failures of those applications, the development of a heuristic categorization technique that captures the type of data collected in the describing section of the CDM method, and the optimization of the user interface design and the system operation.

A patent has been filed regarding the work contained in this paper.

6 References

Bushey, R., Mauney, J. M., and Deelman, T. (1998). The application of a behavioral categorization technique in the development of user models. *The Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting*, 434-438.

Eberts, Ray E. (1994). *User interface design.* Englewood Cliffs, NJ: Prentice Hall.

Eggemeier, F. T. (1988). Properties of workload assessment techniques. In P. A. Hancock and N. Meshkati (Eds.), *Human mental workload*. (pp. 139-183) Amsterdam: North-Holland

Harrell, C. R. and Tumay, K. (1995). *Simulation Made Easy*. Industrial Engineering and Management Press.

Hart, S. G., and Staveland, L. E. (1988). Development of the NASA-TLX (task load index): Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati (Eds.), *Human mental workload*. (pp. 139-183) Amsterdam: North-Holland.

John, Bonnie E. and Kieras, David E. (1996). Using GOMS for user interface design and evaluation: Which technique? *ACM Transactions on Computer-Human Interaction, 3* (4), 287-319.

Kieras, David E. (1996). Task analysis and the design of functionality. In T. Allen (Ed,), *Handbook of computer science and engineering*. Boca Raton, FL: CRC Press.

Milligan, G. W. (1980). An examination of the effects of six types of error perturbation on fifteen clustering algorithms. *Psychometrika, 45*, 325–342.

Pegden, C. D., Shannon, R. E., and Sadowski, R. P. (1990*). Introduction to Simulation Using SIMAN*. McGraw – Hill, Inc.

Shingledecker, C. A. (1980). Operator strategy: A neglected variable in workload assessment. *The American Psychological Association, Division 21 Symposium on "Mental Workload Measurement: The Theory Application Interface"*. Montreal, Quebec, Canada: American Psychological Association.

Wei, Z. G. (1997). *Mental load and performance at different automation levels*. Delft, Netherlands: WEI, Z. G.

Welford, A. T. (1978). Mental work-load as a function of demand, capacity, strategy, and skill. *Ergonomics, 21* (3), 151-167.